Nonlinear Dynamics

Nonlinear Dynamics

Edited by

Todd Evans

Intech

Published by Intech

Intech
Olajnica 19/2, 32000 Vukovar, Croatia

Free online edition of this book you can find under www.sciyo.com
Additional copies can be obtained from:
publication@sciyo.com

First published January 2010
Printed in India

Technical Editor: Teodora Smiljanic

Nonlinear Dynamics, Edited by Todd Evans
p. cm.
ISBN 978-953-7619-61-9

Preface

This volume covers a diverse collection of topics dealing with some of the fundamental concepts and applications embodied in the study of nonlinear dynamics. Each of the 15 chapters contained in this compendium generally fit into one of five topical areas: physics applications, nonlinear oscillators, electrical and mechanical systems, biological and behavioral applications or random processes. The authors of these chapters have contributed a stimulating cross section of new results, which provide a fertile spectrum of ideas that will inspire both seasoned researches and students.

Editor

Todd Evans
General Atomics
United States

Contents

1

Nonlinear Absorption of Light in Materials with Long-lived Excited States

Francesca Serra and Eugene M. Terentjev
University of Cambridge
United Kingdom

1. Introduction

The absorption of light is an important phenomenon which has many applications in all the natural sciences. One can say that all the chemical elements, molecules, complex substances, and even galaxies, have their own “fingerprint” in the light absorption spectrum, as a consequence of the allowed transitions between all electronic and vibronic levels.

The UV-Visible (UV-Vis) light (200-800 nm) has an energy comparable to that typical of the transitions between the electrons in the outer shells or in molecular orbitals. Each atom has a fixed number of atomic levels, and therefore those spectra are composed of narrow lines, corresponding to the transitions between these levels. When molecules and macromolecules are considered, the absorption spectrum is no longer characterised by thin lines but by wide absorption bands. This is due to the fact that the electronic levels are split in many vibrational and rotational sub-levels, which increase in number with the increasing complexity of the molecules. IR spectroscopy is often used to investigate these lower energy modes, but for very complex biological molecules not even this technique can resolve each line precisely because the energy split between the various levels is too small. One possible way to obtain higher resolution spectra is to lower the sample temperature, in order to suppress many of the vibrational and rotational modes. For biological molecules, though, lowering the temperature can be a problem if one wants to study, for example, the activity of enzimes, which only work at physiological temperatures. One of the advantages of absorption spectroscopy (IR and UV-Vis) is to be a non-disruptive technique, also for “delicate” molecules like polymers and biomolecules.

In the process of light absorption by molecules, once a photon with the right energy is absorbed, the molecule goes into an excited state at higher energy [Born and Wolf 1999, Dunning & Hulet 1996]. Eventually, it spontaneously returns to the ground state, but it can relax following several mechanisms. When excited, the molecule reaches, in general, one of the sub-levels of a higher electronic state. The first process is then, generally, a relaxation to the lower energy state of that electronic level (schematised in figure 1). This process is usually very fast (in the femtosecond scale) and not radiative. From this level, there are several pathways to dissipate the energy: a radiative transition from the lower level of the excited state to the ground state (fluorescence), accompanied by the emission of a photon at lower energy than the absorbed one; a flip of the electronic spin, which leads to a transition between singlet and triplet state (intersystem crossing), often associated with another

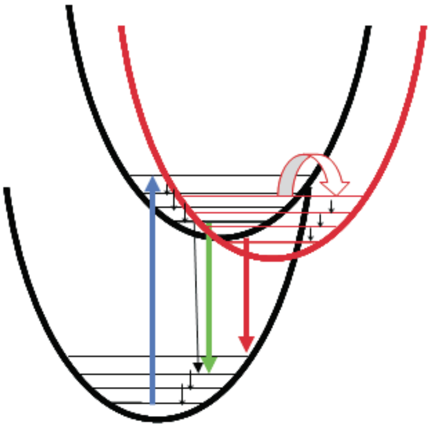


Fig. 1. A scheme representing some possibility of excitation/disexcitation of a molecule. Each electronic level is split into many vibrational and rotational sub-levels. The blue arrow describes the absorption of a photon, the green arrow the emission of a photon from the lower energy level of the excited state (fluorescence), while the black arrows indicate all the nonradiative energy dissipation mechanisms, which can be alternative to fluorescence. The intersystem crossing is another mechanism of disexcitation: the triplet state is represented with the red curve, and the transition with the thick arrow. The molecule can relax over long time to the ground state either with a nonradiative process or via phosphorescence (red arrow).

radiative process (phosphorescence); a non radiative decay where the energy is released by heat dissipation. In some molecules the relaxation pathway following the excitation is more complex, and it can involve interaction with other molecules. In such cases the energy can be transferred to other molecules via radiative or non radiative processes: azobenzene, for example, is a photosensitive molecule which, after excitation, undergoes a conformational change; a more common molecule, like chlorophyll in plant cell chloroplasts, transfers the excitation to the neighbouring molecules until the energy reaches the photosynthetic complex where the photosynthesis takes place.

The common characteristic shared by fluorescent molecules, molecules with a triplet state and photosensitive molecules like azobenzene, is that the lifetime of the excited state is long compared to the time it takes for the excitation to occur. This brings us to the subject of this chapter, which deals with a phenomenon, closely associated with the lifetime of the excited state, which we called “dynamic photobleaching”. In general usage, the term “photobleaching” has been taken to refer to permanent damaging of a chemical, generally due to prolongued exposure to light. Here, we will not consider this, but rather a reversible phenomenon whereby the number of molecules in the ground state is depleted as a consequence of the long lifetime of the excited state.

This effect has important consequences for UV-Visible spectroscopy measurements. In practical use, UVVis light absorption experiments are simple and straightforward: a collimated beam of light is sent onto a sample, the transmitted light is collected by a spectrometer and the ratio between the incident and the transmitted light is measured. Its simplicity means that this technique is widely used in many areas of science. The information one can get from these measurements concerns the allowed electronic transitions. On the other hand, once the electronic structure of a substance is known, computer simulations are able to reproduce absorption spectra.

A very common use of UV-vis spectroscopy is to measure the concentration of substances, and this requires the celebrated Lambert-Beer (LB) law. This semi-empirical law states that the light propagating in a thick absorbing sample is attenuated at a constant rate, that is, every layer absorbs the same proportion of light [Jaffe & Orchin 1962]. This can be expressed simply as the remaining light intensity at a depth x into the sample is: $I(x) = I_0 \exp(-x/D)$ where I_0 is the incident intensity and D is a characteristic length which is called the "penetration depth" of a given material. If an absorbing dye is dispersed in a solution (or in an isotropic solid matrix) this penetration depth is inversely proportional to the dye concentration. In this way it is possible to determine a dye concentration c by experimentally measuring the absorbance, defined as the logarithm of intensity ratio

$$A = \log\left(\frac{I_0}{I}\right) = \ln 10 \ln\left(\frac{I_0}{I}\right) = \frac{x}{D} = x\frac{c}{\delta} \tag{1}$$

where x is the thickness of the sample (the light path length), D is the penetration depth, c the concentration of the chromophore, and δ the universal length scale characteristic of a specific molecule/solvent. One should note that in chemistry and biology one often uses base-10 logarithm in defining the Absorbance, rather than the more intuitive natural logarithm. If c is in molar units, the constant of proportionality ε is the "molar absorption coefficient" and it is inversely proportional to the characteristic length δ defined above.

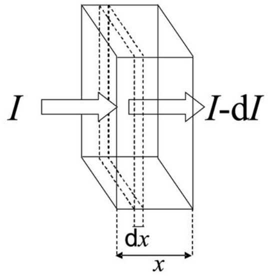


Fig. 2. Schematic diagram of a typical measurement of light absorption. The amount of absorbed light dI across the layer dx is proportional to the number of chromophores in that volume.

The derivation of this empirical law is straightforward. It assumes that the fraction of light absorbed by a thin layer of sample (thickness dx) is proportional to the number of molecules it contains (see figure 2), expressed as the volume fraction n times the volume of the thin layer ($Area \cdot dx$)

$$\frac{dI}{I} \propto n \cdot Area \cdot dx$$

where I is the intensity of the incident light. Introducing the cross section σ, which is a measure of the probability of a photon being absorbed by a chromophore, the differential equation becomes

$$\frac{dI}{I} = \sigma n dx$$

Solving the equation from 0 to x (total thickness of the sample), with a light I_0 incident on the front of the sample, one has

$$\ln(I/I_0) = -A \ln 10 = -\sigma n x$$

and we obtain equation 1 (rearranging the units opportunely).

Thanks to the Lambert-Beer law, UV-visible absorption spectroscopy is a useful and practical tool in many areas of science [Serdyuk et al. 2007]. The technique is widely used in organic chemistry and biology, as macromolecules often have a characteristic absorption in the UV and, more rarely, in the visible region of the EM spectrum. For example, all proteins have a characteristic absorption band around 190nm, due to the molecular orbital formed by the peptide bond, and another band around 280nm due to the aromatic side chains of aminoacids. Usually, this band is used to determine the concentration of proteins in a compound. Nucleic acids also absorb in the UV region and have a strong absorption band at 260 nm. The ratio between the absorption peak at 260 and 280 nm can give information about the relative quantity of DNA and protein in a biological complex, like ribosome. In atmospheric sciences, absorption spectroscopy is used to identify the composition of the air [Heard 2006]. Because the concentration of the species is very low, the light path must be very big to yield a detectable signal. Because L is so large and the concentration can change over the long range, a generalised Lambert-Beer law is preferred:

$$\ln\left(\frac{I_0}{I}\right) = \int_0^L \sigma_i c_i dx$$

where σ_i is the absorption cross section of each species i. Visible absorption can even be applied as a diagnostic tool. In medicine, for example, it is used to measure microvascular hemoglobin oxygen saturation (StO2) in small, thin tissue volumes (like small capillaries in the mouth) to identify ischemia and hypoxemia [Benaron et al. 2005].

All these applications rely on the validity of the LB law. However, this empirical law has limitations, and deviations are observed due to aggregation phenomena or electrostatic interactions between particles. The simpler form of the LB law also fails to describe the two-photon absorption and the excited state absorption process, and it must be substituted by a generalised Lambert-Beer law [Nathan et al. 1985]. These phenomena are usually present only at very high incident light intensity. Also, highly scattering media, very relevant for the medical and geological applications, produce large deviations from LB law.

This chapter addresses the topic of deviations from the LB law occurring in photosensitive media due to self-induced transparency, or photobleaching [McCall & Hahn 1967, Armstrong 1965]. This effect has been reported in a number of different biological systems such as rhodopsin [Merbs & Nathans 1992], green fluorescent protein [Henderson et al. 2007] and light harvesting complexes [Bopp et al. 1997] stimulated with strong laser radiation.

In figure 1, we showed how the excitation/disexcitation of a molecule is essentially a 3-state (or more!) process. Some of the energy loss, however, occurs very quickly and only involves vibrational levels. Considering the different time scales, one can simplify this into a 2-state model: an excitation process which promotes the molecule into a long-lived metastable state and its relaxation to the ground state. The origin of the long life of the metastable state depends on the particular system under study. In the case of spin flip of the excited electron, the physical reason underlying the stability of the triplet state is to be found in the selection rules, which practically forbid the transition between two different spin states (excited triplet state- ground singlet state). This process has raised a vivid interest in the scientific community in the last few decades, because triplet state is often a big problem in organic semiconductor devices [Wohlgenannt & Vardeny 2003]. Alternatively, the molecule, excited by light, gets "trapped" in a metastable state, separated from the ground state by an energy barrier. This is the case for azobenzene, a small molecule which exists in two different forms (isomers *trans* and *cis*). The transition between the two isomers requires breaking a double bond. UV light with a certain energy induces this double-bond breakage and lets the molecule rotate around its axis; with a certain probability, the bond will reform when the molecule is in a metastable *cis* isomer. The relaxation to the ground (lower energy) state can only happen if there is enough energy to break the double bond again. This can occur if the molecule is excited with a light at a different wavelength, or if the thermal fluctuations provide the molecule with enough energy to overcome the energy barrier and return to the ground state. The thermal relaxation is very slow and the characteristic lifetime depends on the nature of the chromophore and of the surrounding environment. This is a classical Kramers problem of overcoming an energy barrier (the breakage of the double bond) between the metastable and the ground state. In the case of this simple molecule, the Lambert-Beer law is no longer accurate because of a phenomenon which we call here "dynamic photobleaching" or saturable absorption. It means that the photons which shine on a sample are absorbed by the chromophores in the first layers. If these molecules don't return to their ground state immediately, when new photons fall on the sample they can't be absorbed anymore in the initial layers and therefore propagate through the sample with a sub-exponential law. So, the effective photo-bleaching of the first layers allows a further propagation of light into the sample and this leads to nonlinear phenomena which are interesting both from the theoretical [Andorn 1971, Berglund 2004, Statman & Janossi 2003, Corbett & Warner 2007] and from the experimental point of view [Meitzner & Fischer 2002, Barrett et al. 2007, Van Oosten et al. 2005, Van Oosten et al. 2007].

The aim of this chapter is to explore the effect that this phenomenon has on the typical absorption measurements which are commonly performed on these kinds of molecules. We will propose a new theory which can mathematically describe this effect and then we will give experimental evidence of its validity both on azobenzene, a molecule with a very long-lived excited state and whose kinetics of transition can be followed, and on more common fluorescent molecules, like chlorophyll, focussing on the absorption of light at equilibrium.

2. Materials and methods

2.1 Azobenzene

The molecule 4'-hexyloxy-4-((acryloyloxyhexyloxy)azobenzene (abbreviated as AC_6AzoC_6) was synthesized in our lab by Dr. A.R. Tajbakhsh. Its molecular structure is shown in figure 3 and its synthesis is described in [Serra & Terentjev 2008a]. All azobenzene-based molecules exist in two isomers, *trans* and *cis*: the transition between the trans isomer, more stable, to the cis isomer, metastable, is stimulated by UV light, while the opposite reaction can be spontaneous. The isomers of the described molecule are shown in fig. 3

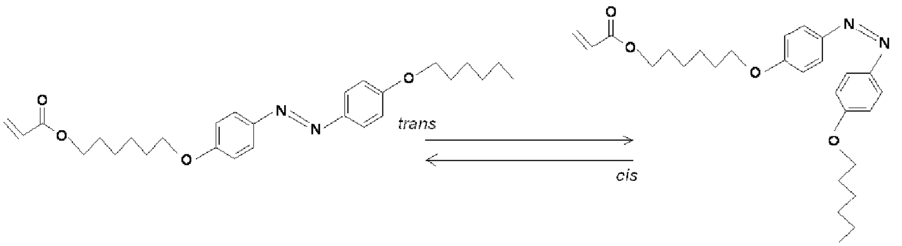


Fig. 3. The monomer used AC_6AzoC_6 has an acrylate head group followed by a carbon chain where the azo-group is attached. It is schematised here in its two isomers, *trans* and *cis*.

The two isomers of this molecule absorb light at different wavelengths: the trans isomer has a peak around 365 nm, while the cis isomer absorbs at 440 nm. It is thus possible to monitor the kinetics of *trans-cis* transition.

Monitoring a conversion process in real-time presents difficulties for a traditional spectrophotometer, because measurements over the whole spectrum of wavelengths take a long time, and moreover it is often difficult to access the sample in order to provide the UV illumination for isomerisation. For this reason, we chose a spectrometer equipped with a CCD camera, which is able to collect signal across the whole visible spectrum simultaneously. This technique works by illuminating the sample with white light; a system of gratings then splits the transmitted light into its various spectral components, whose intensity is measured by an array of photodiodes. This type of spectroscope does not require a fixed or enclosed sample holder, therefore placing another source of illumination close to the sample is easy.

For the measurements of light absorbtion a Thermo-Oriel MS260i (focal length 260mm) spectrometer was used. The apparatus consists of a quartz probe lamp with an adjustable slit, a quartz cuvette with 1cm optical path, an optical liquid lightguide to conduct the light from the cuvette to the spectrometer, a 50 μm slit at the entrance of the spectrometer, and the Andor linear-array CCD camera connected to a computer. The simultaneous measurement of all spectral frequencies allows for a response as fast as 0.021 s and the possibility of reducing noise by averaging over many measurements. Before every absorption spectrum, a background and a reference spectrum were collected: the background is the spectrum collected without the illumination from the probe lamp, and the reference was the spectrum collected with the probe lamp illuminating the cuvette filled with solvent (without the chromophore dye). The absorbance was then calculated from the counts of the detector as:

$$A = \log\left(\frac{\text{Reference} - \text{Background}}{\text{Signal} - \text{Background}}\right).$$

For all the experiments, it was important to verify that the linear relation between absorbance and concentration held for the value of absorbances considered. It was shown that the absorption-concentration relation was linear below $A \approx 1.2$ in the base-10 defined absorbance. At the intensity used for this experiment, for a concentration c expressed in moles, the penetration depth at 365nm was $\delta = c * 480$nm.

At higher concentrations, the linearity fails because of various phenomena, including aggregation of the molecules (especially with molecules like proteins or polymers), the scattering of light from big particles and stray light in the spectroscope.

We provide monochromatic illumination to stimulate the isomerisation of azobenzene using a Schott KL 1500 LCD lamp, placed at 90 degrees with respect to the incident probe light and the optical fiber that collects the light from the sample. In this way the cuvette is irradiated with UV light while the absorption spectrum is recorded along the perpendicular beam path, so the absorption can be followed in real time without interference of the illumination light. The intensity of the monochromatic light was in the order of a few tens of μW cm^{-2}.

All isomerisation reactions were followed for 90 minutes, which was a sufficient time for reaching the respective photostationary states. After this, the lamp was switched off and the absorbance was measured during thermal isomerisation in the dark. An example of spectrum measured with this technique is shown in fig. 4.

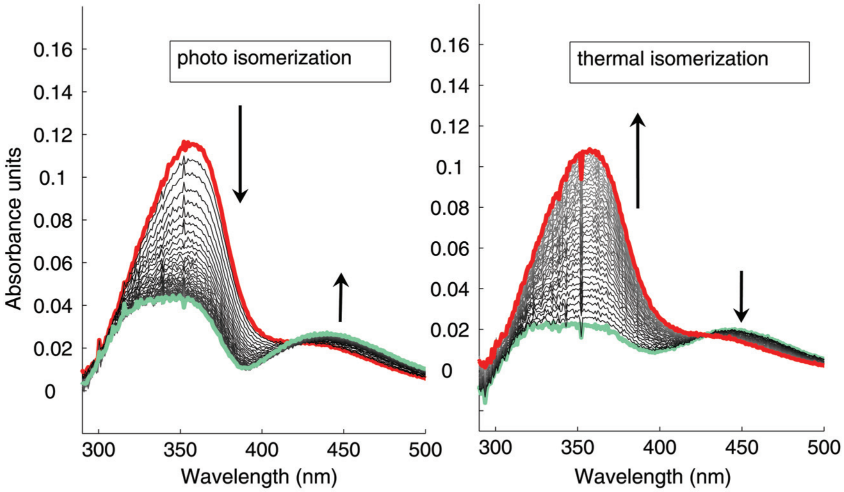


Fig. 4. Photo-induced isomerisation (a) and thermal relaxation (b) curves of AC6AzoC6 recorded as a time sequence. The arrows indicate the direction of the peak movement during the reaction.

2.2 Other chromophores

Chlorophyll was extracted from *Commelina Communis* leaves[1]. The leaves were first boiled in distilled water, in order to kill the enzymes which digest chlorophyll once the leaf is cut from the plant. The leaves were then dried and ground up with a pestle, with a few drops of

[1] Leaves were kindly provided by J. McGregor

acetone, and then left in a 50% hexane/water mixture (the hexane forms a layer on top of the water), to separate the chlorophyll from the water-soluble compounds (vitamins, etc). The extracted solution was filtered to avoid impurities, like dust particles or even intact chloroplasts which could be responsible for light scattering effects. The whole extraction process was carried out in the dark. This method of extraction does not allow the separation of chlorophyll from the carotenoids which could be present in the leaves. However, the collected spectrum shows that the stronger absorption bands are those of chlorophyll, and this means that the other compounds are present only in low concentrations. Moreover, for the purpose of our experiment, an highly purified chlorophyll is not needed, because the analysis focusses on the absorption band around 660nm, far from the absorption band of carotenoids (blue region). What is important to remark is that there are different kinds of chlorophyll, whose relative content varies from species to species. The two main chlorophyll components, called “chlorophyll a” and “chlorophyll b”, differ by a carboxylic group, attached only to the porphyrine ring in chlorophyll b. The two molecules have slightly different absorption spectra, but this is not relevant for the experiments, provided that the plant species from which the chlorophyll is extracted is always the same (and has therefore the same percentage of chlorophyll a and b).

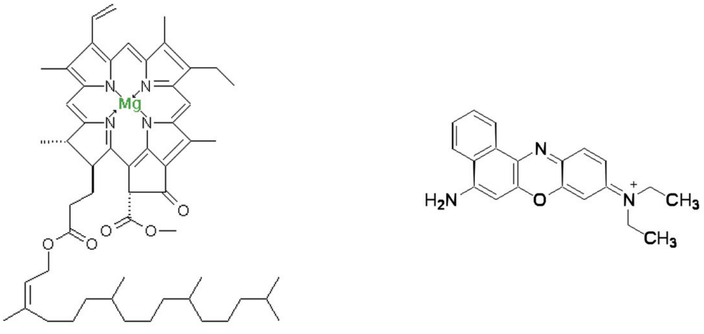
From the discussion above, it is clear that chlorophyll is a very special molecule, and has many peculiarities. In order to demonstrate that the theory is more general, a commercial dye with a strong absorption in the visible region is also investigated. Nile Blue A, a dye commonly used for staining DNA, but with spectral properties similar to chlorophyll, was selected. The chemical structure of chlorophyll and Nile Blue is shown in figure 5.

The spectroscopy experiments were conducted with an Ocean Optics USB 4000 spectrometer, equipped with optical fibers. A 25W halogen lamp with spectral range 400-880 nm, whose intensity could be tuned, was used as illumination source. The light was focussed with collimating lenses onto a 1 cm cuvette containing 3 ml of solution.

For comparison with more “conventional” spectroscopes (meaning, with a fixed sample holder and a fixed intensity of incident light), a Cary UV-Vis Spectrophotometer was used to measure spectra and absorbances of the two substances at various concentrations. The intensity of the incident light from the spectrometer was also measured with a power meter.

In order to measure the intensity of the incident light, key quantity in our experiments, the light from the source was shone onto the detector directly, in the absence of any sample or cuvette in between. Knowing the characteristic response of the spectrometer detector, it is possible to measure the intensity of light. Three different values, the number of counts at a single wavelength or the integral of the intensity over the range of wavelength which correspond to the absorption peak, and the integral over all the wavelengths could be used to quantify the incident light intensity. In all cases the outcome of the experimental results was the same. Using as a value of intensity the intensity at the single peak-wavelength made it possible to compare it to the conventional spectroscope (which produce monochromatic light). It was verified that the detector had a linear response in counts versus intensity over the range of intensities we used.

For the absorption spectra, measured as $A = \log_{10}(I_0/I)$ reference spectra for the pure solvent were taken before each measurement at every light intensity. The absorption and fluorescence spectra of these materials are shown in figure 6. We chose to refer all the absorption values to the wavelengths of the peaks in the yellow-red region.



(a) (b)

Fig. 5. Chemical structure of a) chlorophyll a and b) nile blue dye.

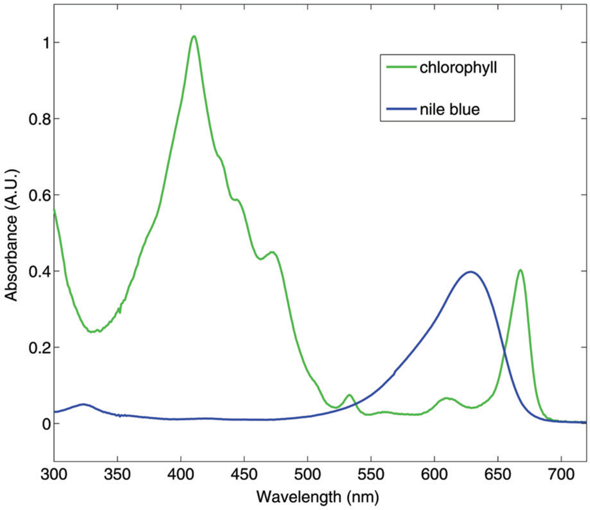


Fig. 6. Absorption spectra of chlorophyll (green curve) and nile blue (blue). The absorption peaks which were considered in this work are those at 668nm and 628nm respectively.

For each solution, the linearity of the absorption/concentration curve was verified, in order to avoid falling into a trivial nonlinear regime. The experiments were conducted in random order of light intensity, and the reproducibility was verified. The absorption of each chlorophyll solution at 660 nm was stable over a range of hours at constant incident light intensity, indicating the absence of chemical irreversible bleaching. Fluorescence from the dye was also ruled out as a possible source of disturbance, because at the light intensity we used it is not detectable with our equipment.

3. Theory

Here we present a description of the dynamical photobleaching effect in the case of azobenzene isomerisation, which was previously discussed. We will then generalise the discussion to all the molecules with a long-lived excited state, and show how this affects the measurements of light absorption.

The non-Lambertian propagation of light through a medium has important consequences for the analysis of photo-isomerization kinetics: when the photo-bleaching becomes important, the measured absorbance no longer follows a simple (traditionally used) exponential law. In photosensitive molecules like azobenzene, irradiation with light of a certain wavelength induces a conformational change (isomerization) from an equilibrium *trans* state where the benzene rings are far apart to a bent *cis* state where they are closer. The isomerization process follows first-order kinetics. Calling the fraction of molecules in the two states *trans* and *cis* n_t and n_c we have

$$\frac{dn_t}{dt} = -Ikn_t + Ik_b n_c + \gamma n_c$$

where *I* is the intensity of light, *k* is the *trans-cis* isomerisation rate, k_b the stimulated back transition rate (*cis* to *trans*), and γ the thermal relaxation rate. In the experiments on azobenzene described below we use an illuminating light monochromated at the *trans-cis* transition wavelength. In this case the stimulated *cis-trans* isomerization is negligible (that is, $Ik_b \to 0$) and, remembering that $n_c = 1 - n_t$ the kinetic equation reduces to

$$\frac{dn_t}{dt} = -\gamma\Big(\left[1 + Ik/\gamma\right] n_t - 1\Big). \tag{2}$$

In this equation the intensity $I = I(x)$ is the light intensity at a certain depth into the sample. It is convenient to define a non-dimensional parameter $\alpha = I_0 k/\gamma$, which represent the balance of photo- and thermal isomerization at a given incident intensity I_0. In this notation, the amount of molecules in the *trans* conformation in the photostationary state, when the balance between n_t and n_c is stable, and therefore $dn_t/dt = 0$ the equation reduces to simply

$$\gamma n_t + \frac{I}{I_0}\alpha n_t - \gamma = 0$$

therefore

$$n_t = \frac{1}{1 + \alpha I/I_0} \tag{3}$$

To express mathematically the reversible photobleaching phenomenon, it can be assumed that the change in light intensity across a thin layer of sample (thickness dx) is proportional to the number of molecules which are excited, i.e. the number of chromophores which absorbed light in a small volume of sample of thickness dx; neglecting the stimulated *cis-trans* isomerization (which is appropriate in our study), the model can be much simplified to give, per unit time:

$$\Delta I \propto Ikn_t(x,t) \cdot Area \cdot dx$$

Then, combining all the parameters, such as the photon cross section and the transition rates, we recover the penetration depth $D = \delta / c$ as the relevant parameter of the relation, and the final expression is:

$$\frac{dI}{dx} = -I\frac{n_t}{D} \tag{4}$$

with D the penetration depth, inversely proportional to concentration. In order to study this problem at the photostationary state, one needs to combine the equations (4) and (3).

$$\frac{dI}{dx} = -\frac{I}{D}\frac{1}{1+\alpha I/I_0} \tag{5}$$

Solving the differential equation, the stationary-state light intensity at a depth x is given by the relationship [Corbett & Warner 2007]

$$\ln\left(\frac{I(x)}{I_0}\right) + \alpha\left(\frac{I(x)}{I_0} - 1\right) = -\frac{x}{D}. \tag{6}$$

Looking at the equation above some important insights can be gained. The most important is that in the limit α = 0 the equation reduces to the Lambert-Beer law, i.e. an exponential decay in the transmittance through the medium. Therefore all the nonlinearity is included in α. The opposite limit, when α is very big, leads to a linear relation between transmittance and sample thickness. Figure 7 is a representation of equation 6 and it shows the intensity variation $I(x)$ for several values of the parameter α: from the plot of transmittance as a function of x, it is clear that, if the incident intensity, and therefore α, is low enough, the Lambert-Beer law is valid and the decrease is exponential, but if the incident intensity is high the bleaching of the first layers becomes progressively more relevant such that they become partially transparent to the radiation. The decay thus tends to become linear in the bulk of the sample, $I(x) \approx I_0(1 - x/\alpha D)$.

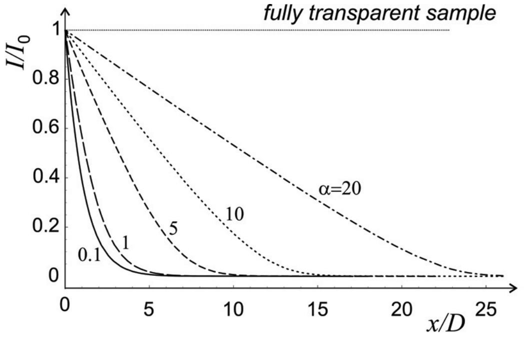


Fig. 7. Transmitted intensity ratio I/I_0 in the photostationary state as a function of the parameter x/D (proportional to sample thickness or inversely proportional to chromophore concentration) for several values of α. At small α the decay is exponential; the light penetrates deeper into the sample as α increases, as the decay tends to be linear. [Serra & Terentjev 2008b]

In order to model the dynamics of photoisomerisation, which is evidently inhomogeneous across the sample, it is not enough to model photobleaching with equation 6, but instead the equations (2) and (4) should be coupled. Calculating a time derivative of equation 4 leads to

$$\frac{d}{dt}\left(\frac{1}{I}\frac{dI}{dx}\right) = -\frac{1}{D}\frac{dn_t}{dt} \tag{7}$$

In the right hand side expression, equation 2 can be substituted, giving

$$\frac{d}{dt}\left(\frac{1}{I}\frac{dI}{dx}\right) = -\frac{\gamma}{D}\left(1 - \left[\alpha I/I_0 + 1\right] n_t\right) \tag{8}$$

This equation can be solved with the following method [Corbett et al. 2008]. Introducing the variable y = ln(I/ I_0), which is a very sensible variable, being also the inverse of the absorbance, the left-hand side is greatly simplified and one obtains

$$\frac{d}{dt}\left(\frac{dy}{dx}\right) = -\frac{\gamma}{D}\left(1 - (\alpha\exp(y) + 1)\right) n_t \tag{9}$$

Also, from equation 4

$$\frac{1}{D}\frac{dn}{dt} = -\frac{d}{dt}\left(\frac{dy}{dx}\right) \Rightarrow n_t = -D\frac{dy}{dx} \tag{10}$$

Substituting n_t in equation 9 and rearranging, one finds

$$\frac{d}{dt}\left(\frac{dy}{dx}\right) = -\frac{\gamma}{D} + \gamma(\alpha\exp(y) + 1)\frac{dy}{dx} \tag{11}$$

In the next step, one has to keep in mind that

$$\frac{d}{dx}(\gamma\alpha e^y + \gamma y) = \gamma\alpha\exp(y)\frac{dy}{dx} + \gamma\frac{dy}{dx}$$

Rearranging equation 11 and exchanging the order of derivatives on the left-hand side, the equation reduces to

$$\frac{d}{dx}\left(\frac{dy}{dt} + \gamma(\alpha\exp(y) + y)\right) = -\frac{\gamma}{D} \tag{12}$$

It is now possible to integrate this expression. Integrating between 0 and x, the integral and the derivative on the left-hand side cancel out and one finds:

$$\frac{dy}{dt} + \gamma(\alpha\exp(y) + y) - \gamma\alpha = -\frac{\gamma x}{D} \tag{13}$$

The factor $\gamma\alpha$ comes from the solution of the definite integral for x=0 (the lower integration limit). In fact, if x=0, I = I_0 and therefore y=0 by definition.

The last step is a time integration. At time zero, the absorbance A = $-y$ must be equal to the Lambert-Beer law value x/ D. Including these considerations, the integral expression for the intensity I(x, t) becomes:

$$\gamma t = -\int_{x/D}^{A} \frac{dy}{\left(x/D - y - \alpha + \alpha e^{-y}\right)}. \tag{14}$$

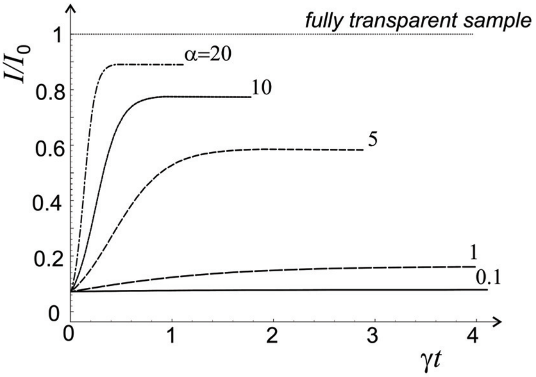


Fig. 8. Transmitted intensity ratio I/I_0 as function of time through a fixed value of x/D = 2.7 and different incident light intensities. There are several things to observe in this figure. First, as α changes, the photostationary state reaches different levels, as expected: if α is bigger, the sample becomes more transparent. The second thing is that when α increases (therefore intensity of light) the sample reaches the stationary state more quickly. The last observation, most important for our study, is that with increasing α the deviation of the kinetics from a simple exponential becomes more and more evident. [Serra & Terentjev 2008b]

The upper limit of this integral is the measurable absorbance $A = \ln[I_0/I(x, t)]$ from a sample of thickness x. Figure 8 shows a simulation predicting the time-evolution of intensity transmitted along the path x/D. Note that at $t = 0$ all curves converge to the Lambert-Beer $I/I_0 = \exp(-x/D)$, while at long times a significant portion of chromophore is bleached and the transmitted intensity increases.

We should note that the problem of non-linear photo-absorption dynamics is not only restricted to azobenzene isomerisation. Even ordinary dye molecules that do not undergo conformational changes stimulated by photon absorption, still follow the same dynamic principles, but with electronic transitions in place of *trans-cis* isomerization. Therefore, the results of this paper should be looked upon as widely applicable to other systems. In particular, the two key conclusions, that the crossover intensity into the non-linear photo-absorption regime is independent of dye concentration and that the rate of the transition is independent on solvent viscosity, are probably completely general.

The model we propose only assumes a two-configuration system, and it does not imply anything about the nature of the two states. Therefore, it is important to verify that this model has a wider and broader validity, and, in detail, that it helps to understand the behaviour of a large class of chromophores, like fluorescent molecules. Azobenzene molecule exists in two physical states, *trans* or *cis*; for chlorophyll, one could make an analogy and, considering the electronic transition, call the two states "ground" and "excited", we still find the same formula at the photostationary state

$$\ln\big(I_0/I(x)\big) + I_0(k/\gamma)\big(1 - I(x)/I_0\big) = x/D \tag{15}$$

where x is the path length of light through the sample. It is important to see here that the absorbance has a nonlinear dependence on the incident light intensity. The limits where the

LB law is recovered are either very low incident intensity (practically, it can never be achieved) or a very fast recovery to the ground state compared to the excitation.

Equation 15 can has important implications for the interpretation of absorption data. Solving the equation for the absorbance A = log(I_0/I) leads to the expression

$$A\ln(10) + I_0\frac{k}{\gamma}(1 - 10^{-A}) = \frac{xc}{\delta} \tag{16}$$

from which it is clear that the value of the absorption does not only explicitly depend on the concentration and optical path length, but also on the intensity of the incident light I_0.

4. Non-linear kinetics of photobleaching.

Azobenzene, as previously discussed, is a small molecule with two double-bonded nitrogen atoms linked to two benzyl rings. It is photosensitive, because exposure to light induces an isomerisation *trans-cis* (indicating two possible spatial arrangements of the benzyl rings) around the double bond between the nitrogens, and this results in a molecule shape change. The process is fully reversible either by stimulated backward photoconversion (with a light at a different wavelength), or by spontaneous relaxation to the equilibrium *trans* configuration.

The isomerization of azobenzene and its derivatives has been extensively studied for the last fifty years [Sudesh Kumar & Neckers 1989, Renner & Moroder 2006, Rau 1990], because this molecule has many interesting features and its applications range from electronics to biomedicine. It has been used as a model molecule for all the biological processes that involve similar reactions, like the isomerization of retinal in rhodopsin, or as a probe for measuring the free volume in polymers [Victor & Torkelson 1987]. More recently, its characteristic response to the polarization state of light made it a suitable molecule for surface patterning [Nathanson et al. 1992, El Halabieh et al. 2004]. Finally, azobenzene-containing elastomers can give rise to inhomogeneous photo-mechanical effects and their applications as photoactuators and artificial muscles are under study [Hogan et al. 2002, Finkelmann et al. 2001, Yu et al. 2004].

However, in spite of the large literature on the subject, many fundamental mechanisms and effects have not been clarified yet. It is assumed that the isomerization reaction is very sensitive to both electrical and mechanical characteristics of the environment which surrounds the molecules, but identifying and separating these effects is a difficult and often ambiguous task.

Of the two possible isomers of azobenzene, the *trans* form is the lowest energy form, since the benzyl ring electron clouds are far apart (see fig. 3), but under UV light an isomerisation occurs; once in the *cis* state, the molecules can return back to their equilibrium trans state both by stimulated isomerisation with visible light or by thermal relaxation [Rau 1990]. The rate constants of these two processes are usually different, thermal isomerisation being slower. The microscopic mechanism that leads to the isomerisation is still not clear, but there are suggestions for both rotational and inversional [Asano & Okada 1984] mechanisms may be competing.

The isomerisation of azobenzene can be monitored by UV-Vis (ultraviolet and visible light) spectroscopy, because the two *trans* and *cis* compounds have different absorption spectra in this range: the *trans* isomer absorbs around 365 nm, while the *cis* isomer at around 440 nm

[Rau 1990]. Irradiation with light at the wavelength of the *trans* peak progressively depletes the molecules in this conformation. This reaction can be "followed" by measuring the intensity of the absorbance of the spectrum peak at 365 nm, which decreases as the *trans-cis* photo-isomerisation reaction proceeds. An example of spectral evolution as isomerisation proceeds can be found in figure 4.

The models that were proposed for reaction kinetic are basically first order models, with the important exception of azobenzene in polymer matrices. The fraction of isomers in the *cis* state, n_c, varies as [Zimmerman et al. 1958, Mechau et al. 2005]:

$$\frac{dn_c}{dt} = Ikn_t - Ik_b n_c - \gamma n_c \qquad (17)$$

where I is the irradiation intensity, k_b and k are the *cis-trans* and *trans-cis* constant of photoisomerisation respectively, γ is the thermal *cis-trans* isomerisation and n_t represents the fraction of molecules in the *trans* state, and it is equal to $(1 - n_c)$. In the derivation of the formula the sensible assumption was made that the *trans-cis* thermal isomerisation constant is negligible.

A basic characteristic of the photoisomerisation problem is the rate of spontaneous thermal *cis-trans* isomerisation γ. For a given azobenzene derivative, at fixed (room) temperature and sufficiently low dye concentrations to avoid self-interaction, this rate is approximately the same for all our solutions. We measured this rate after monitoring the relaxation of the spectrum after the UV illumination is switched off (see [Serra & Terentjev 2008a] for detail) and obtained $\gamma \approx 1.25 \cdot 10^{-4} s^{-1}$ (or the corresponding relaxation time of ~ 8000s).

In order to test the predictions of the theory, dynamic absorption measurements were performed for different dye concentrations and different light intensities. Considering equation (14) this is equivalent to changing x/D (where D is inversely proportional to the dye concentration) and α, which is proportional to the incident intensity I_0. With our experimental setup it was possible to follow all the isomerisation kinetics and thus the time dependence of I/I_0 [Serra & Terentjev 2008b].

We prepared three different dye solutions with (non-dimensional) weight fractions $c = 2.5 \cdot 10^{-3}$, 0.01 and 0.025, resulting in values of penetration depth ranging from D = 36 mm, to D = 3.6 mm. We recall here the physical meaning of the penetration length, which is the distance through the sample over which the light falls across a sample to 1/10 of its original intensity. The cuvette containing the sample is 1 cm long; therefore a sample with D equal to a few millimeters is almost completely opaque.

The measurements were performed using the Thermo-Oriel spectroscope described in the materials and methods section. Illumination was provided by a Nichia chip-type UV-LED, emitting at 365nm (bandwith about 10nm wide) whose output power was accurately regulated by a power supply. The LED light was attenuated by passing through a black tube of controlled length, placed in front of a quartz cuvette with 1 cm optical path. Several values of intensity were used in reported measurements, ranging between I_0 = 4 and 60 μW/cm^2. It is important to point out that these intensity values are very low and that most experiments on azobenzene isomerization are performed with intensities which are orders of magnitude higher, making the photobleaching much more of an issue. The low values of intensity allowed us to have a kinetics slow enough to detect the features which the theory predicts at short times. Every point of the spectrum was collected as an average of 100 measurements. All isomerization reactions were followed for several hours, until a photostationary state was reached. The measurements were repeated at different illumination intensities (regulated with the power supply) and at different dye concentrations.

In all cases it was important to verify that the dye concentration remained in a range where, at time t = 0, the linear proportionality between absorbance and concentration (Lambert-Beer law) held. This is important because the concentration of molecules in the *trans* state at every instant was determined from the absorbance at 365 nm. Absorbance was measured at several concentrations. The deviation from linearity started at $A \approx 1.2$, which corresponds to the dye weight fraction of c = 0.03 (3 wt%) in our 1 cm cuvette. After this point, aggregation effects start playing a role and the basic Lambert-Beer law is no longer valid, undermining the theoretical relationship given by the equation (14). We always kept the concentrations below this value, so that the linearity at t = 0 was maintained, with $A = x/D$ where the penetration depth is inversely proportional to chromophore concentration, expressed as weight fraction, $D \approx \delta / c$ with δ = 91 μm.

For our detailed dynamic experiments, a very important issue was the viscosity of the solution. In fact, at high illumination intensity we have encountered an unexpected problem. Figure 9 shows that the transmission of light through a low-viscosity dye solution (in pure toluene) displays a characteristic oscillatory behaviour. Detailed analysis of this phenomenon is under further investigation. Whether the oscillations are linked to the local convection due to the heating of the sample spot [Nitzan & Ross 1973] or to the diffusion of the less dense *cis* molecules – or whether they are intrinsic to the non-linear photochemical process [Borderie et al. 1992] – is not clear at this stage.

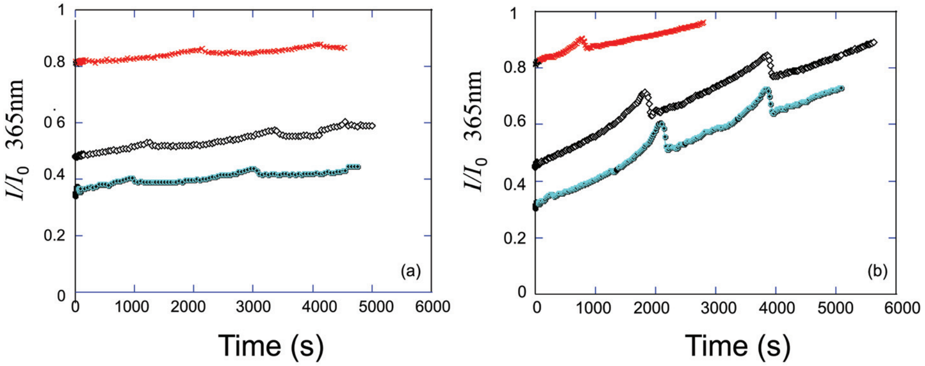


Fig. 9. Kinetics of isomerisation monitored through the observation of I/I_0 over time for 3 different values of x/D (× - 0.2, ◊ - 0.7, • - 1.1) and 2 different values of α, corresponding to: (a) $I_0 = 4\mu\text{W cm}^{-2}$, and (b) $I_0 = 20\mu\text{W cm}^{-2}$. The periodic instability was reproducible in all low-viscosity experiments. [Serra & Terentjev 2008b]

In order to avoid this difficulty, the dye solutions were prepared in a mixture of toluene and polystyrene of high molecular weight. Adding polystyrene increases the viscosity of the solution by over 2 orders of magnitude, and in this way prevents fluid motion in the cuvette on the time scales of our measurements. Polystyrenetoluene solutions were prepared at a fixed weight ratio. Adding polystyrene to toluene increases the Rayleigh scattering of the solution, but we felt that we could safely do that because on one hand the absorption dynamics is not affected (we used the same concentration for all the measurements and for the reference spectrum), and on the other we measured the transmittance of the toluene-polystyrene solution, which is almost equal to the pure toluene solution at 365nm.

In figure 10 the representative experimental results are shown for the solution with the highest chromophore concentration (D = 4.6 mm, leading to x/D = 2.2) and three values of incident intensity. One finds that all curves converge to the same initial value corresponding to the $I/I_0 = \exp[-x/D]$, which for this concentration means quite a low transmission ($I/I_0 \approx$ 0.11). If the isomerisation didn't take place, the sample absorption would be constant in time according to the classical Lambert-Beer. A traditional description of the kinetics of isomerisation would predict an exponential decay of the absorption over time, but from figure 10 we see a strong deviation. We fit the data with the theoretical model given by equation (14) where we input the values of γ(thermal relaxation) and x/D, leaving only $\alpha = I_0k/\gamma$ free. Two data sets at higher intensity show the transmitted $I(x,\ t)$ reach saturated values. In this case we are confident of the fit because we have to match both the slope and the amplitude of the curve. We obtain $\alpha \approx 60$ for $I_0 = 60\ \mu\mathrm{W/cm^2}$, $\alpha \approx 20$ for $I_0 = 20\ \mu\mathrm{W/cm^2}$, and $\alpha \approx 4$ for $I_0 = 4\ \mu\mathrm{W/cm^2}$ (the matching of values is pure coincidence). We found that one particular output of experimental recording, the absorbance plateau value (photo-bleached) at long times, was extremely sensitive to the reading of reference intensity I_0. The latter measurement could be affected by various stray factors and in a few cases we had to rescale the raw absorbance readings with a proper reference value. This issue did not have any effect on absorbance at t = 0, or the kinetics.

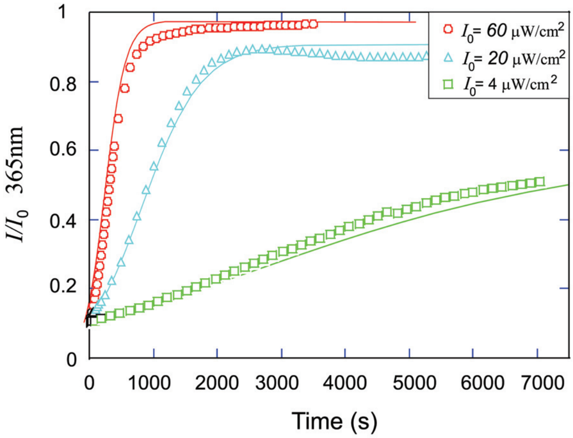


Fig. 10. The effect of photo-bleaching for samples with high dye concentration (x/D = 2.2). Three values of irradiation intensity are labelled on the plot. Solid lines are fits to the data with only one free parameter α, giving α = 60 for the highest intensity, α = 20 for the middle intensity, and α = 4 for the lowest intensity. [Serra & Terentjev 2008b]

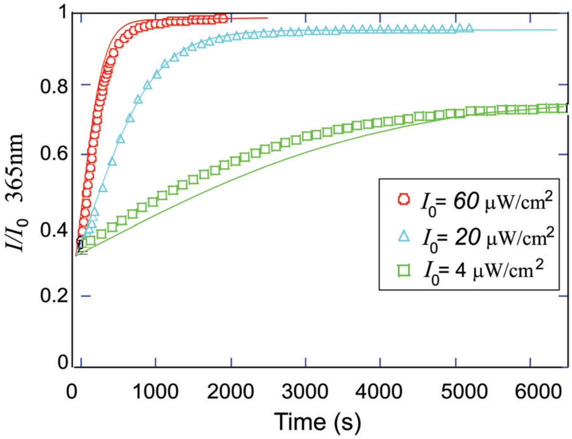


Fig. 11. The same experiment as in figure 10 but with an intermediate dye concentration (x/D = 1.1), and the same values of irradiation intensity. Here the solid lines are not fits, but theoretical plots of equation (14) for α = 60, 20 and 4 for the decreasing I_0, respectively. [Serra & Terentjev 2008b]

At a lower concentration of chromophore, corresponding to $D \approx 9.2$ mm and x/D = 1.1 (the transmitted intensity is about 1/10 of the incident intensity), figure 11 shows the similar features of the non-linearity, which are especially evident at very short times. Again all curves start at the same $I/I_0 \approx 0.33$. At higher irradiation intensities we achieve the saturation and the steady-state value $I(x)$ corresponding to the solution of equation (6). The change of curvature, notable in figures 8 and 10, is not so clear here even at the highest I_0. However, in the comparative analysis of data we now take a different approach. Assuming all the parameters for the curves are now known (γ and x/D from independent measurements, and α from the fitting in figure 10), we simply plot the theoretical equation (14) on top of the experimental data. It is clear that the theory is in excellent agreement with the data.

Finally, we study the case of low dye concentration ($D \approx 91$ mm, x/D = 0.14) in figure 12: this is also the case which is more relevant for biological spectroscopy studies, where the concentration of chromophore is usually small. In this exemplified case, the initial transmittance is very high: almost 85% of the incident light goes through the sample. Here, the complicated integral equation (14) simplifies dramatically, since at small $x/D \ll 1$ the difference between A = ln 10 ln(I_0/I) and x/D (which is the range of integration in (14), is also small. The integration can then be carried out analytically, giving

$$\ln\left(\frac{I(x,t)}{I_0}\right) \approx -\frac{x}{D}\left[1 - \frac{\alpha}{1+\alpha}\left(1 - e^{-\gamma(1+\alpha)t}\right)\right], \qquad (19)$$

which gives in the stationary state the correct solution of equation (6) approximated at small x/D:

$$\ln\left(\frac{I(x)}{I_0}\right) \approx -\frac{x/D}{1+\alpha}.$$

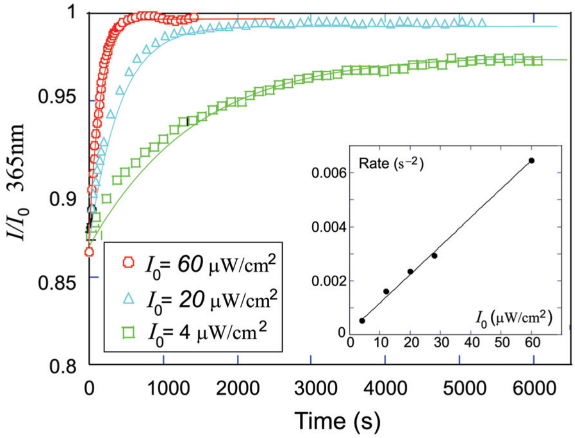


Fig. 12. At low dye concentration (x/D = 0.14) the sample is relatively transparent. The data are for the same three values of irradiation intensity as in the earlier plots (but note that the I/I_0 axis starts from 0.8). The solid lines are theoretical fits for α = 60, 20 and 5. The inset shows the plot of exponential relaxation rate τ^{-1} against I_0, with the linear fit. [Serra & Terentjev 2008b]

The fits of the data for $I(t)$ are again in good agreement with the full theory. More importantly, we also see that that the rate of the process described by the approximation (19) is given by the simple exponential, $\tau^{-1} = \gamma(1 + \alpha) = \gamma + kI_0$. This is in fact the rate originally seen in the kinetic equation (2). Therefore, if we instead fit the family of experimental curves in figure 12 (and several other data sets we measured) by the simple exponential growth of the absorbance, we can have an independent measure of the relaxation rates obtained by this fit. The inset in figure 12 plots these rates for all the I_0 values we have studied. A clear linear relation between the relaxation rate and I_0 allows us to independently determine the molecular constant:

$$k \approx 10^{-4}\mathrm{cm}^2\mathrm{s}^{-1}\mu\mathrm{W}^{-1}$$

The measurement of k, with high accuracy, gives the ratio $k/\gamma \approx 1$ cm^{2}s$^{-1}\mu$W^{-1}, which explains the fitted values of the non-dimensional parameter $\alpha = I_0k/\gamma$.

The consequences of this nonlinear behaviour have in the last year raised an interested in some research groups who studied the azobenzene-based actuators. The original work by Corbett and Warner, in fact, focused only on the steady-state behaviour, could lead to accurate prediction about the effect of dynamic photobleaching on the bending angle of elastomers [Corbett & Warner 2007]. In fact, the dynamic photobleaching is the reason why heavily doped cantilevers, where the penetration depth is very small, can still bend if irradiated with sufficiently intense beams. Because the contraction of cantilevers is due to the force generated by the differential contraction of the top and bottom layers, if the light was propagating exponentially in the medium the bending would be impossible, because the thin layer where the light propagates is too small to generate enough force. A non-exponential propagation of light due to photobleaching, instead, can explain this effect. Subsequent work by Van Oosten, Corbett et al. [Van Oosten et al. 2008, Corbett & Warner

2008] and White et al. [White et al. 2009] shown experimental evidence of this effect on the bending of cantilevers. Lee et al. also shown the nonexponential kinetics on a different azobenzene-based molecule [Lee et al. 2009].

5. Absorption of fluorescent molecules.

Because absorption spectroscopy is so widely used in biology, we want to show the effect of dynamic photobleaching on a biological molecule, and we chose chlorophyll, an important substance in biology (and in everyday life). Chlorophyll has a very recognisable absorption spectrum, which shows two clear peaks, one in the blue and the other in the red region (which procures its green colour) of the electromagnetic spectrum. It is also fluorescent in the far red and the characteristic lifetime of its excited state is about 4 ns [Hipkins 1986, Jaffe & Orchin 1962]. If it is irradiated by UV light or very strong visible light it undergoes a photo-chemical bleaching which degrades the molecule irreversibly and leads it to precipitate from solution, as many studies reported [Mirchin et al. 2003, Mirchin & Peled 2005, Carpentier et al. 1987]. We wish to observe a dynamical reversible bleaching due to the absorption of light, rather than this chemical degradation process.

In the previous section, the theoretical model was verified in the case of azobenzene, a molecule with a very long lived excited state. Because the kinetics of transition could be followed by a spectrometer, it was also possible to model it with the kinetics law (equation 14). The model, as we said, does not make any hypothesis on the nature of the transition, and can therefore be extended to all “two-state” (or more realistically, to the simplified 3-state) systems. Fluorescent molecules have an excited state with a characteristic lifetime of a few nanoseconds, which is still much slower than the typical time of excitation. These characteristic times, though, are too short to be followed with conventional spectroscopes, and the transition kinetics cannot be followed as in the previous case. The model, however, also makes predictions also about the transmittance at the photostationary state, which differs from the LB law transmittance. To clarify, in figure 10 the Beer limit would be the transmittance at time zero, and the stationary state the transmittance at long times.

It was important, for our experiments, to rule out all possible mechanisms leading to failure of LB law. As it was previously discussed, LB law has many limitations. It fails at high concentration of dyes, when they start to interact with each other and form aggregates; it fails if the stray light is high and the apparent absorption seems to reach a saturation level; it can fail at high intensity of the incident light if nonlinear effects like multiple photon absorption, or saturable absorption occur [Abitam et al. 2008, Correa et al. 2002]; it fails for highly scattering samples because the light is sent out at a non-zero angle. In order to rule out all these possible effects, we place ourselves in the most favourable experimental conditions: low concentration of dye and low illumination intensity.

According to the model, the behaviour at the stationary state is described by equation 16. The important thing to observe is that the absorbance (or, equivalently, the transmittance) also depends on the intensity of the incident light I_0. In order to experimentally verify this dependence, five different solutions of chlorophyll at known concentrations were measured at various light intensities. In this section all the absorbances will be reported in base-10 logarithmic form. Figure 13 shows the outcome of measurements of chlorophyll absorption of the same solution using different incident intensities. The result was striking: the change in the measured absorbance was very substantially affected by this parameter.

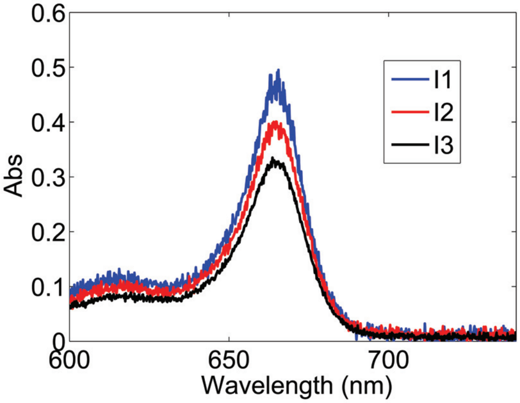


Fig. 13. Absorption of chlorophyll in ethanol at the same concentration (in fact, exactly the same solutions) measured only changing the incident illumination intensity, $I1$ = 6.5, $I2$ = 13.1, $I3$ = 27.5 $\mu Wm^{-2}s^{-1}$.

Some interesting consequences of this effect are shown in figure 14 and 15. The values correspond to the steady-state absorption at the peak wavelength. Indeed it is possible to see a strong dependence on the incident light intensity which is enhanced at high solute concentrations. A change in intensity of about 80% of the maximum value leads to a change in absorbance of about 50%. Figure 14 shows the dependence of the absorbance on the intensity at various concentrations. Equation 16 cannot be explicitly solved for A, but only for I_0

$$\frac{I_0 k}{\gamma} = \frac{x/D - A\ln(10)}{1-10^{-A}}$$

which gives the fits in the plot. Figure 15 shows the same data in the classical absorbance-concentration plot, for different intensities. It is important to remark that the experimental points can be satisfactorily fitted with a straight line in all cases (as the LB law says) but the line slopes are very different. Therefore the absorption coefficient may have different values if it is measured with a different light source. The exchangeability of results between different laboratories is thus in question.

We obtained analogous results with Nile Blue, a simpler chromophore. We decided to test this dye, described in the Material and Methods section, because it has an absorption spectrum similar to chlorophyll in the red region, but it is a simpler and well studied molecule. This also proves that the results are general, and that aggregation phenomena which may occur in chlorophyll solutions (giving rise to scattering phenomena from still intact chloroplasts) are ruled out as a possible cause for the observed behaviour.

All of the experiments were repeated several times and the behaviour was reproducible. Moreover, the intensity of light was increased and decreased alternatively to exclude the hypothesis of a chemical permanent photobleaching as a reason for absorption decrease.

Due to the phenomenon of reversible (dynamic) photobleaching, a simple absorption experiment like the one described in the introduction is in practice impossible. The values of the absorption coefficients are meaningless if they don’t carry the information about the intensity of the incident light.

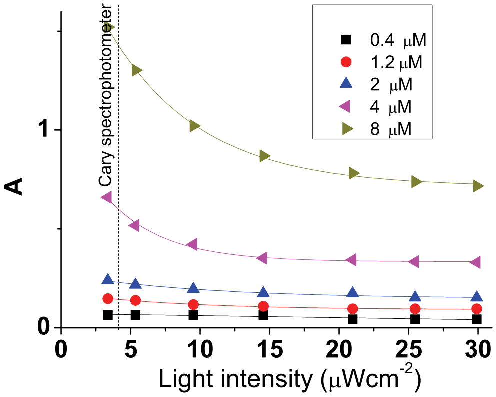


Fig. 14. Absorption of chlorophyll as a function of the intensity of incident light. One can see an increase of absorbance at low intensities. The values are reported for five different concentrations. In the figure, the black dotted line corresponds to the intensity of the incident light of a commercial "traditional" spectrophotometer (Cary-UV-Vis). This comparison is done in order to show that the range of intensity of our set-up is the same as a more conventional one.

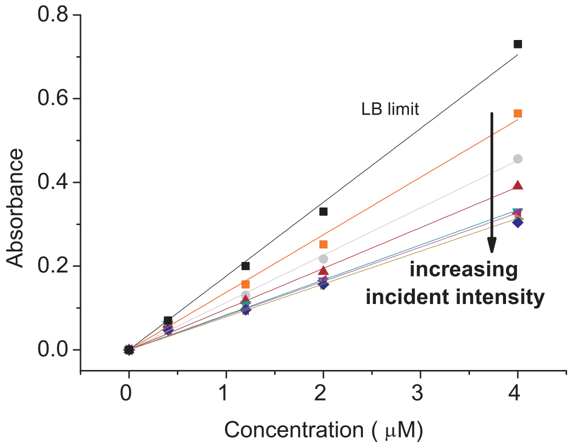


Fig. 15. Absorption as a function of concentration for the different values of incident light. All the lines have a good "Lambert-Beer" linear form but different proportionality coefficients. The LB limit was extrapolated from the ideal limit of zero intensity.

In light of this, can we use the theoretical model to find a new method to determine concentrations using absorption spectroscopy, removing this dependence on the incident light intensity? Looking at equation 15, knowing the ratio of the concentrations of two solutions makes it possible to measure the combined ratio of parameters $\alpha = (k/\gamma)I_0$. If one solution has an unknown concentration c_1 and another solution is obtained by a dilution of the first one, so that $c_1/c_2 = r$, measuring the absorbance of the two solutions 1 and 2, at the same incident light intensity and the same path length x, one obtains:

$$r = \frac{A_1 \ln(10) + \alpha(1 - 10^{-A_1})}{A_2 \ln(10) + \alpha(1 - 10^{-A_2})} \tag{20}$$

then

$$\alpha = \frac{(A_1 - rA_2)\ln(10)}{(10^{-A_1} - 1) - r(10^{-A_2} - 1)} \tag{21}$$

Knowing α, one can simply determine the unknown concentration c_1 as

$$c_1 = \frac{\delta}{x}\left[A_1 \ln(10) + \alpha - \alpha 10^{-A_1}\right] \tag{22}$$

The relation yields the ratio c/δ, and therefore the knowledge of the absorption coefficient δ^{-1} is required. On the other hand, the same method allows determination of δ once the concentration is known.

We took a series of concentrations of Nile Blue solutions and the corresponding absorbance measured at different intensities of incident light. Taking pairs of measurements of absorbance at known concentration, at the same value of light intensity, we extracted the value of the parameter r from each pair, and from that we calculated α and ϵ according to equation 21 and 22. Averaging over all of them, we obtained the value ϵ = 120000 ± 20000$M^{-1}cm^{-1}$. The literature reports ϵ = 77000$M^{-1}cm^{-1}$, but we attribute this to the fact that, at intensities greater than zero, the absorption values are always systematically smaller (see figures 14 and 15). Therefore the deviation from the literature value is still consistent with our findings.

The limitation of this method is that it very sensitively depends on the value of the concentration ratio, and therefore the errorbars are quite large. One should not, in fact, rely on the value of α measured only with one pair of measurements. Figure 16 shows the dispersion of the estimate values of α obtained using different pairs of values. The strong dependence of the parameter α on the value of r is evident from formula 21. It is possible to see that the function is divergent when $(10^{-A_1} - 1) - r(10^{-A_2} - 1) \longrightarrow 0$ but this happens when $r \approx A_1/A_2$, which is exactly the region of interest. For this reason, the values of α are very scattered (some of them are even negative, which is physically meaningless), and this formula, although it is correct is principle, is hardly applicable to real experimental data.

6. How to use absorption spectroscopy

In order to overcome this disadvantage, a more robust method is suggested to determine the value of the absorption coefficients. The strong dependence of α on r obviously remains, because it comes directly from formula 21. A method based on a linear regression can be used instead to calculate δ from a series of absorbances and known concentrations. Also, a series of measurements at known δ allow the evaluation of a substance concentration, just like in traditional absorption experiments. Rearranging equation 15, one obtains:

$$c(1 - 10^{-A})^{-1} = \frac{\alpha\delta}{x} + \frac{\delta}{x} A \ln(10)(1 - 10^{-A})^{-1} \tag{23}$$

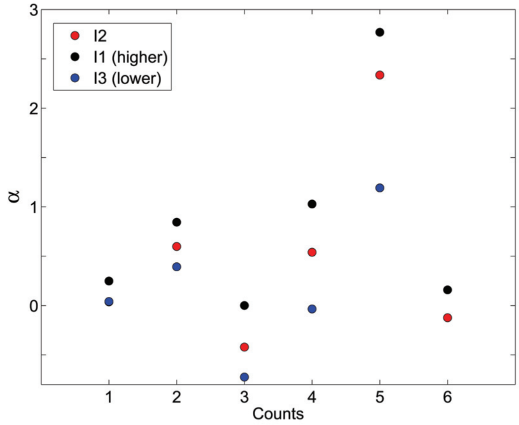


Fig. 16. Values of α obtained for three different values of light intensity, using various pairs of measurements, according to the suggested method. It is important to notice that the values of α are systematically higher at higher intensity, as expected. As one would expect from equation 21, the values of α are very scattered and therefore the calculation of ϵ is not precise: this is due to the strong dependence of α on r around the point where $r \approx A_1/A_2$.

If a set of concentrations and relative absorptions are known, one can plot the quantity $c(1-10^{-A})^{-1}$ as a function of $A\ln(10)(1-10^{-A})^{-1}$. The result is a line whose slope is δ/x and whose intercept is $\delta/x\alpha$. It is therefore possible to determine all the important parameters. Figure 17 shows this plot obtained for a set of Nile Blue dye. Promisingly, all the lines obtained with this method are well fitted with parallel lines, which indicates that they all converge to the same value of ϵ. Several lines indicate several values of incident light intensity. From the plot one can find the parameter α for all the intensities, and also δ, which

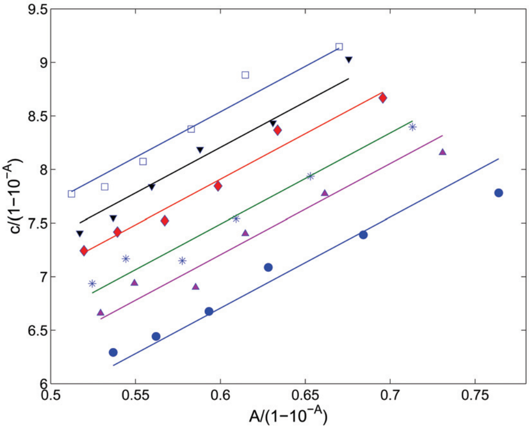


Fig. 17. Application of the suggested method to determine the relevant parameters α and δ. The slope depends on δ, which is the same for all the samples, but the intercept depends on α which changes with the intensity of light.

is simply the slope of the lines. Once δ is known, one can then determine, for x = 1, ϵ = δ^{-1}, the molar absorption coefficient. Following classical error analysis we obtain ϵ = 117700±120M^{-1} cm^{-1}. The agreement with the previous method is good and this second method has the advantage of greatly reducing the experimental errors.

Whilst this method appeared highly successful at a first glance, we discovered that plotting the values of α against the intensity of the incident light, measured from the number of counts on the spectroscope, generated a relation which is not linear. According to the theory, α is simply the product between the incident intensity and some characteristic constant of the material, therefore the nonlinearity shown in figure 18 is not acceptable.

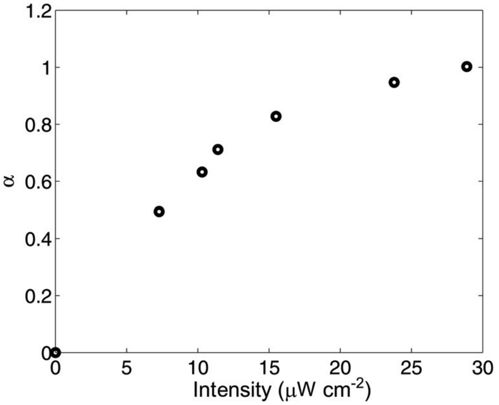


Fig. 18. The parameter α extracted from the intercepts of figure 17 as a function of the intensity of the incident light. The theoretical model predicts a linear relationship, which is not what is observed in the figure!

Considering the possible causes of this discrepancy, one can see in the model that stimulated emission is completely neglected. Neglecting the light-stimulated back-transition to the ground state was reasonable in the case of azobenzene, where the *trans* and *cis* peaks were very far apart, but for fluorescent molecules the same light excites both transitions and this factor should therefore be considered. The calculations become more complicated but the procedure is the same as that described in the introductory section of this chapter.

One should now return to the kinetics equation, which we re-write here for simplicity. We call n the fraction of molecules in the ground state and k_b the rate of the stimulated back transition.

$$\frac{dn}{dt} = -Ikn + Ik_b(1-n) + \gamma(1-n) \tag{24}$$

At the stationary state, the left hand side of the equation is zero and n becomes

$$n = \frac{Ik_b + \gamma}{I(k+k_b)+\gamma} \tag{25}$$

This is the value which should be inserted in the expression for the photobleaching 4, giving

$$\frac{dI}{dx} = \frac{I}{D}\frac{Ik_b + \gamma}{Ik + Ik_b + \gamma} \tag{26}$$

This can be simplified by dividing by k, introducing the parameter $\varphi = k_b / k$ and integrating the equation

$$\int_{I_0}^{I} \frac{dI}{I} \frac{I\varphi + I + \gamma/k}{I\varphi + \gamma/k} = -\int_0^x \frac{dx}{D}$$

Note that earlier, neglecting the stimulated back-reaction, we essentially had $\varphi = k_b / k \rightarrow 0$. While the integration on the right-hand side is trivial, the left-hand side splits into a sum

$$\int_{I_0}^{I} \frac{\varphi + 1}{I\varphi + \gamma/k} + \int_{I_0}^{I} \frac{\gamma/k}{I(I\varphi + \gamma/k)} = -\frac{x}{D}$$

The integration gives

$$\frac{\varphi + 1}{\varphi} \ln\left(\frac{1 + \varphi k/\gamma I}{1 + \varphi k/\gamma I_0}\right) - \ln\left(\frac{I_0}{I} \frac{1 + \varphi k/\gamma I}{1 + \varphi k/\gamma I_0}\right) = -\frac{x}{D}$$

Final simplification leads to:

$$-\ln\frac{I_0}{I} + \frac{1}{\varphi}\ln\left(\frac{1 + \varphi k/\gamma I}{1 + \varphi k/\gamma I_0}\right) = -\frac{x}{D} \tag{27}$$

It is convenient here to reintroduce our usual non-dimensional parameter $\alpha = I_0 k/\gamma$

$$-\ln\frac{I_0}{I} + \frac{1}{\varphi}\ln\left(\frac{1 + \varphi\alpha I/I_0}{1 + \varphi\alpha}\right) = -\frac{x}{D} \tag{28}$$

This expression is the full and general result. In many cases we expect φ to be small, so the expansion at the first order correctly recovers the usual expression 6. Expansion to the second order, instead, gives

$$-\ln\frac{I_0}{I} + \frac{1}{\varphi}\ln\left(\frac{1 + \varphi\alpha I/I_0}{1 + \varphi\alpha}\right) = -\frac{x}{D} \tag{29}$$

Using this equation, the fit to the experimental data improved. The expression was readapted to take into account the base-10 logarithm of the absorbance. The parameter space was restricted because we expected δ and α to be in the same range as previously determined. The best fit to the curves was obtained using $\delta = 8.75 \Rightarrow \epsilon = 114000 M^{-1} cm^{-1}$ and $\varphi = 0.3$. The value of φ, quite substantially greater than zero, is consistent with the need to modify the original equation. Figure 19 shows the concentration c on the y-axis and the absorbance A on the abscissa (different curves for different light intensities): this is because formula 29 can be easily inverted. The values of α, obtained by fitting, increase linearly with the incident intensity, as shown in figure 20. This is good evidence that the stimulated emission cannot be neglected: the theory, thus modified, can well reproduce the experimental data.

7. Conclusions

The most important conclusion of our work is that one has to be cautious with the classical concept of light absorption, represented by the Lambert-Beer law. Even without considering

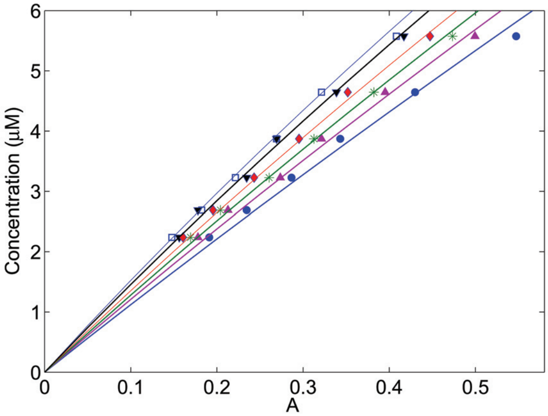


Fig. 19. Fitting of the absorbance/concentration curves at different intensities, obtained with the model which considers the stimulated back-transition.

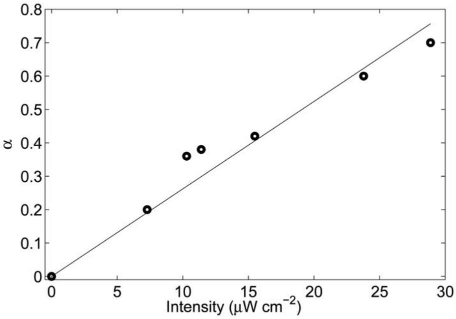


Fig. 20. The parameter α extracted from the fit in the figure above as a function of the intensity of the incident light. In this case we observe the linear relationship with the correct intercept in the origin.

multi-particle effects at high concentration or multiple photon absorption, even at very low concentrations (corresponding in our case to the low x/D ratio) the illumination intensity above a certain crossover level would always produce a non-linear dynamical effect equivalent to the dynamic photo-bleaching, which increases the effective transmittance of the sample. We emphasise that this is a totally reversible phenomenon, unrelated to the chemical bleaching, which involves irreversible damage to the material. The crossover between linear and strongly non-linear regimes is expressed by the non-dimensional parameter $\alpha = I_0 k/\gamma$ and is, therefore, an intrinsic material parameter of every chromophore molecule, but not dependent on the dye concentration. Note that the thermal *cis-trans* isomerization rate γ is strongly temperature dependent, influencing the crossover intensity.

Azobenzene is an ideal molecule for this kind of study, because it allows investigation of the transition kinetics using a simple spectrometer. The experimental data we obtained confirm the predictions of the theoretical model, which provided a satisfactory fit to the data. At high illumination intensity one finds a characteristic sigmoidal shape. Our experiments were deliberately carried out in a highly viscous solvent to eliminate the additional complexities, presented in figure 9, caused by the possible local convection flows of different isomers. Certainly, a much more in-depth study will be required to take such effects into account.

The second result is related to the extension of the model to all fluorescent molecules, or indeed any molecule which has a long-lived excited state. We showed that, even for those molecules where the conversion between the two states is too fast to be followed by the spectroscope, the nonlinearity has an important influence. The absorption values at the stationary state were sensitive to the experimental conditions, and particularly the intensity of the incident light.

It must be said that in the literature the light intensity of the spectrometer light source is very rarely mentioned, therefore it is possible that many values of absorption coefficient reported are wrong or meaningless. The reason why, to our knowledge, no one took this phenomenon into consideration before is that many commercial spectroscopes always work with the same light intensity, so the results are self-consistent. Also, weighing proteins or other materials is often a difficult task, therefore a discrepancy between the expected value and the literature values could be easily explained away. In fact, we showed that there is a much deeper reason.

Moreover, this raises also problems of interpretation of data obtained comparing, for example, the intensity of two different absorption peaks, because, even at the same incident light intensity, the nonlinearity also depends on the factor (k/γ) which is different at each wavelength. All these problems could be overcome using the method we suggested, which is simple and straightforward and it allows reproducibility of results.

8. Acknowledgements

The authors wish to thank A. Tajbakhsh for the preparation of chemicals, A. Smith for the advice in chlorophyll extraction, M. Warner and D. Corbett for discussing the theoretical aspects of the work, J. McGregor for reviewing the manuscript, J. Huppert, P. Cicuta, M. Kolle, K. Thomas and L. Payet for helpful discussion.

9. References

[Abitam et al. 2008] H. Abitam; H. Bohr; P. Buchhave. Correction to the beer-lambert-bouguer law for optical absorption. *Appl. Optics*, 47:5354, 2008.

[Andorn 1971] M. Andorn; K. H. Bar-Eli. Optical bleaching and deviation from beer's law of solutions illuminated by a ruby laser. i. cryptocyanine solutions. *J.Chem.Phys.*, 55:5008–5016, 1971.

[Armstrong 1965] J.A. Armstrong. Saturable optical absorption in phthalocyanine dyes. *J.Appl.Phys.*, 36:471, 1965.

[Asano and Okada 1984] T. Asano; T. Okada. Thermal $2-e$ isomerization of azobenzenes. the pressure, solvent, and substituent effects. *J.Org.Chem.*, 49, 4387–4391, 1984.

[Barrett et al. 2007] C. J. Barrett; J. Mamiya; K. G. Yager; T. Ikeda. Photo-mechanical effects in azobenzenecontaining soft materials. *Soft Matter*, 3, 1249, 2007.

[Benaron et al. 2005] D. A. Benaron; I. H. Parachikov; W-F. Cheong; S. Friedland; B. E. Rubinsky; D. M. Otten; F. W. Liu; C. J. Levinson; A.L. Murphy; J.W. Price; Y. Talmi; J. P. Weersing; J. L. Duckworth; U. B. Horchner; E.L. Kermit. Design of a visible-light spectroscopy clinical tissue oximeter. *J. Biomed. Opt.*, 10, 044005, 2005.

[Berglund 2004] A. J. Berglund. Nonexponential statistics of fluorescence photobleaching. *J.Chem.Phys.*, 121: 2899–2903, 2004.

[Bopp et al. 1997] M. A. Bopp; Y. W. Jia; L. Q. Li; R. J. Cogdell; R. M. Hochstrasser. Fluorescence and photobleaching dynamics of single light-harvesting complexes. *Proc.Natl.Am.Sci.USA*, 94, 10630, 1997.

[Borderie et al. 1992] B. Borderie; D. Lavabre; J.C. Micheau; J.P. Laplante. Nonlinear dynamics, multiple steady states, and oscillations in photochemlstry. *J.Phys.Chem*, 96, 2953, 1992.

[Born and Wolf 1999] M. Born; E. Wolf. *Principles of Optics- 7th edition*. Cambridge University Press, Cambridge, 1999.

[Carpentier et al. 1987] R. Carpentier; R. M. Leblanc; M. Mimeault. Photoinhibition and chlorophyll photobleaching in immobilized thylakoid membranes. *Enzyme Microb.Technol.*, 9, 489, 1987.

[Corbett and Warner 2007] D. Corbett; M. Warner. Linear and non-linear photo-induced deformations of cantilevers. *Phys.Rev.Lett.*, 99, 174302, 2007.

[Corbett and Warner 2008] D. Corbett; M. Warner. Polarization dependence of optically driven polydomain elastomer mechanics. *Phys. Rev. E*, 78, 061701, 2008.

[Corbett et al. 2008] D. Corbett; C. L. van Oosten; M. Warner. Nonlinear dynamics of optical absorption of intense beams. *Phys. Rev. A*, 78, 013823, 2008.

[Correa et al. 2002] D. S. Correa; L. de Boni; D.S. dos Santos jr.; N. M. Barbosa Neto; O. N. Oliveira jr.; L. Misoguti; S. C. Zilio; C. R. Mendonc¸a. Reverse saturable absorption in chlorophyll a solutions. *Appl. Phys. B*, 74:559, 2002.

[Dunning and Hulet 1996] F. B. Dunning; R. G. Hulet. *Atomic, molecular, and optical physics; Atoms and Molecules*. Academic Press Inc., San Diego, California, 1996.

[El Halabieh et al. 2004] H. El Halabieh; O. Mermut; C. J. Barrett. Using light to control physical properties of polymers and surfaces with azobenzene chromophores. *Pure Appl.Chem.*, 76:1445–1465, 2004.

[Finkelmann et al. 2001] H. Finkelmann; E. Nishikawa; G. G. Pereira; M. Warner. A new optomechanical effect in solids. *Phys.Rev.Lett.*, 87, 015501, 2001.

[Heard 2006] D. E. Heard. *Analytical techniques for atmospheric measurement*. Blackwell publishing Ltd., Oxford, 2006.

[Henderson et al. 2007] J. N. Henderson; H. W. Ai; R. E. Campbell; S. J. Remington. Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proc.Natl.Am.Sci.USA*, 104, 6672, 2007.

[Hipkins 1986] M. F. Hipkins; N. R. Baker. *Photosynthesis energy transduction: a practical approach.* IRL Press, Oxford, Oxford, 1986.

[Hogan et al. 2002] P. M. Hogan; A. R. Tajbakhsh; E. M. Terentjev. Uv manipulation of order and macroscopic shape in nematic elastomers. *Phys.Rev.E*, 65, 41720, 2002.

[Jaffe and Orchin 1962] H. H Jaffe; M. Orchin. *Theory and applications of ultraviolet spectroscopy*. Wiley, New York, 1962.

[Lee et al. 2009] Y. J. Lee; S. I. Yanga; D. S. Kangb; S.-W. Joo. Solvent dependent photo-isomerization of 4-dimethylaminoazobenzene carboxylic acid. *Chem. Phys.*, 361, 176–179, 2009.

[McCall and Hahn 1967] S. L. McCall; E. L. Hahn. Self-induced transparency by pulsed coherent light. *Phys.Rev.Lett.*, 18, 908, 1967.

[Mechau et al. 2005] N. Mechau; M. Saphiannikova; D. Neher. Dielectric and mechanical properties of azobenzene polymer layers under visible and ultraviolet irradiation. *Macromolecules*, 38, 3894–3902, 2005.

[Meitzner and Fischer 2002] G. D. Meitzner; D. A. Fischer. Distortions of fluorescence yield x-ray absorption spectra due to sample thickness. *Microchem.J.*, 71, 281, 2002.

[Merbs and Nathans 1992] S. L. Merbs; J. Nathans. Photobleaching difference absorption-spectra of human cone pigments- quantitative analysis and comparison to other methods. *Photochem.Photobiol.*, 56, 869, 1992.

[Mirchin et al. 2003] N. Mirchin; A. Peled; Y. Dror. Modeling and analysis of bleaching processes in photoexcited chlorophyll solutions. *Synthetic Metals*, 138, 323, 2003.

[Mirchin and Peled 2005] N. Mirchin; A. Peled. Photo-bleaching response in chlorophyll solutions. *Appl. Surface Sci.*, 248, 91, 2005.

[Nathan et al. 1985] V. Nathan; A. H. Guenther; S. S. Mitra. Review of multiphoton absorption in crystalline solids. *J.Opt.Soc.Am.B*, 2, 294, 1985.

[Nathanson et al. 1992] A. Natansohn; P. Rochon; J. Gosselin; S. Xie. Azo polymers for reversible optical storage. 1.poly[4'- [[2- (acr yloyloxy)ethyll ethylaminol-4-ni troazo benzene]. *Macromolecules*, 25, 2268– 2273, 1992.

[Nitzan and Ross 1973] A. Nitzan; J. Ross. Oscillations, multiple steady states, and instabilities in illuminated systems. *J.Chem.Phys.*, 59, 241, 1973.

[Rau 1990] H. Rau. Photochemistry of azobenzene. In J. F. Rabek, editor, *Photochemistry and Photophysics*, pages 119–142. CRC press; Boca Raton, 1990.

[Renner and Moroder 2006] C. Renner; L. Moroder. Azobenzene as conformational switch in model peptides. *ChemBioChem*, 7, 868–878, 2006.

[Serdyuk et al. 2007] I. N. Serdyuk; N. R. Zaccai; J. Zaccai. *Methods in molecular biophysics*. Cambridge University Press, Cambridge, 2007.

[Serra and Terentjev 2008a] F. Serra; E. M. Terentjev. Effects of solvent viscosity and polarity on the isomerization of azobenzene. *Macromolecules*, 123, 981–986, 2008.

[Serra and Terentjev 2008b] F. Serra; E. M. Terentjev. Nonlinear dynamics of absorption and photobleaching of dyes. *J.Chem.Phys.*, 123, 224510, 2008.

[Statman and Janossi 2003] D. Statman ; I. Janossi. Study of photoisomerization of azo dyes in liquid crystals. *J.Chem.Phys.*, 118, 3222–3232, 2003.

[Sudesh Kumar and Neckers 1989] G. Sudesh Kumar ; D. C. Neckers. Photochemistry in azobenzenecontaining polymers. *Chem.Rev.*, 89, 1915–1925, 1989.

[Van Oosten et al. 2005] K. D. Harris; R. Cuypers; P. Scheibe; C. L. van Oosten; C.W.M. Bastiaansen; J. Lub; J.D. Broer. Large amplitude light-induced motion in high elastic modulus polymer actuators. *J.Mater.Chem.*, 15, 5043, 2005.

[Van Oosten et al. 2007] C. L. Van Oosten; K. D. Harris; C. W. M. Bastiaansen; J.D. Broer. Glassy photomechanical liquid-crystal network actuators for microscale devices. *Eur.Phys.J.E*, 23:329, 2007.

[Van Oosten et al. 2008] C. L. Van Oosten; D. Corbett; D. Davies; M. Warner; C. W. M. Bastiaansen; D. J. Broer. Bending dynamics and directionality reversal in liquid crystal network photoactuators. *Macromolecules*, 41:8592–8596, 2008.

[Victor and Torkelson 1987] J. G. Victor; J. M. Torkelson. On measuring the distribution of local free volume in glassy polymers by photochromic and fluorescence techniques. *Macromolecules*, 20, 2241–2250, 1987.

[White et al. 2009] T. J. White; S. V. Serak; N. V. Tabiryan; R. A. Vaiaa; T. J. Bunning. Polarization-controlled, photodriven bending in monodomain liquid crystal elastomer cantilevers. *J. Mater. Chem.*, 19, 1080–1085, 2009.

[Wohlgenannt and Vardeny 2003] M. Wohlgenannt; Z. V. Vardeny. Spin-dependent exciton formation rates in π-conjugated materials. *J. Phys. Condens. Matter*, 15, R83–R107, 2003.

[Yu et al. 2004] Y. Yu; M. Nakano; T. Ikeda. Photoinduced bending and unbending behavior of liquidcrystalline gels and elastomers. *Pure Appl.Chem.*, 78, 1467–1477, 2004.

[Zimmerman et al. 1958] G. Zimmerman; L. Y. Chow; U. J. Paik. The photochemical isomerization of azobenzene. *J. Am. Chem. Soc.*, 80, 3528–3531, 1958.

2

Exact Nonlinear Dynamics in Spinor Bose-Einstein Condensates

Jun'ichi Ieda[1] and Miki Wadati[2]

[1]*Institute for Materials Research, Tohoku University,*
[2]*Department of Physics, Tokyo University of Science*
Japan

1. Introduction

Bose–Einstein Condensation (BEC) of atomic gases has attracted a renewed theoretical and experimental interest in quantum many-body systems at extremely low temperatures (Pethick & Smith; 2002). This excitement stems from two favorable features: (1) by applying magnetic fields and lasers, most of the system parameters, such as the shape, dimensionality, internal states of the condensates, and even the strength of the interatomic interactions, are controllable; (2) due to the diluteness, the mean-field theory explains experiments quite well. In particular, the Gross–Pitaevskii (GP) equation demonstrates its validity as a basic equation for the condensate dynamics. The GP equation is the counterpart of the nonlinear Schrödinger (NLS) equation in nonlinear optics. Thus, a study based on nonlinear analysis is possible and important.

In nonlinear physics, a soliton is remarkable object not only for the fact that exact solutions can be obtained but also for its usefulness as a communications tool due to its robustness. In general, solitons are formed under the balance between nonlinearity and dispersion. For atomic condensates, the former is attributed to the interatomic interactions, while the latter comes from the kinetic energy. Either dark or bright solitons are allowable depending on the positive or negative sign of the interatomic coupling constants *g*, respectively, and indeed have been observed in a quasione-dimensional (q1D) optically constructed waveguide (Strecker et al.; 2002) (Khaykovich et al.; 2002). Such matter-wave solitons are expected in atom optics for applications in atom laser, atom interferometry, and coherent atom transport (Meystre; 2002). In this chapter, we extend the analysis of the matter-wave solitons to a multicomponent case by considering the so-called spinor condensate (Stenger et al.; 1998) whose spin degrees of freedom are liberated under optical traps. Based on theoretical and experimental results, we introduce a new integrable model which describes the dynamical properties of the matter-wave soliton of spinor condensates (Ieda et al.; 2004a). We employ the inverse scattering method to solve this model exactly. As a result, we predict the occurrence of undiscovered physical phenomena such as macroscopic spin precession and spin switching.

The chapter is organized as follows. In Sec. 2, the mean field theory of condensate is briefly reviewed. Section 3 introduces an effective interatomic coupling in a q1D condensate. Using these results, we consider a spinor condensate in q1D regime in Sec. 4. Then, in Sec. 5, we show an integrable condition of the coupled nonlinear equations for spinor condensates in which the exact soliton solutions are derived. In Sec. 6 and 7, we analyze the spin properties of one-soliton and two-soliton, respectively. Finally we summarize our findings and remark some current progresses on this topic in Sec. 8.

2. Mean field theory

The dynamics of BEC wave function can be described by an effective mean-field equation known as the Gross-Pitaevskii (GP) equation. This is a classical nonlinear equation that takes into account the effects of interatomic interactions through an effective mean field.

In this section, we derive the GP equation for a single component condensate and discuss the theoretical background of it for later extension to a low dimensional case and a spinor case.

2.1 Hamiltonian

In order to derive the mean-field equation for atomic BECs, we start with the second quantized Hamiltonian. The Hamiltonian for the system of N interacting bosons with the mass m in a trap potential $U_{\text{trap}}(\mathbf{r})$ can be written as

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}}, \tag{1}$$

$$\hat{H}_0 = \int d\mathbf{r} \hat{\Psi}^\dagger(\mathbf{r}) \left[-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{trap}}(\mathbf{r}) \right] \hat{\Psi}(\mathbf{r}), \tag{2}$$

$$\hat{H}_{\text{int}} = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \hat{\Psi}^\dagger(\mathbf{r}) \hat{\Psi}^\dagger(\mathbf{r}') v(\mathbf{r} - \mathbf{r}') \hat{\Psi}(\mathbf{r}) \hat{\Psi}(\mathbf{r}'), \tag{3}$$

where $v(\mathbf{r} - \mathbf{r}')$ expresses the two-body interaction and the bosonic field operators satisfy the equal-time commutation relations:

$$[\hat{\Psi}(\mathbf{r}), \hat{\Psi}^\dagger(\mathbf{r}')] = \delta(\mathbf{r} - \mathbf{r}'), \quad [\hat{\Psi}^\dagger(\mathbf{r}), \hat{\Psi}^\dagger(\mathbf{r}')] = [\hat{\Psi}(\mathbf{r}), \hat{\Psi}(\mathbf{r}')] = 0. \tag{4}$$

In most of the experiments, the trap is well approximated by a harmonic oscillator potential,

$$U_{\text{trap}}(\mathbf{r}) = \frac{1}{2} m(\omega_x^2 x^2 + \omega_y^2 y^2 + \omega_z^2 z^2). \tag{5}$$

Condensates are pancake-shape for $\omega_z \gg \omega_{x,y}$ whereas cigar-shape for $\omega_{x,y} \gg \omega_z$. For other choice of trap potentials, say a linear or a 4-th order potential, the thermodynamic properties can be changed (Ieda et al.; 2001). The discussion about non-harmonic potentials will be given in a later section in connection with an implementation of quasi-one dimensional system.

The atom-atom interaction $v(\mathbf{r} - \mathbf{r}')$ in a dilute and ultracold system can be approximated by

$$v(\mathbf{r} - \mathbf{r}') = g\delta(\mathbf{r} - \mathbf{r}'), \tag{6}$$

$$g = \frac{4\pi\hbar^2 a}{m}, \tag{7}$$

where a is the s-wave scattering length. The scattering length is the controllable parameter which determines the properties of the low energy scattering between cold atoms. The positive (negative) sign of a corresponds to the effectively repulsive (attractive) interaction.

2.2 Bogoliubov theory

The mean-field theory for *weakly interacting dilute Bose gases* (WIDBG) was proposed in Bogoliubov's 1947 work (Pethick & Smith; 2002). The main idea of his approach consists in separating out the condensate contribution from the bosonic field operator:

$$\hat{\Psi}(\mathbf{r},t) \simeq \sqrt{n_0} + \hat{\phi}(\mathbf{r},t), \tag{8}$$

where $n_0 = N_0/\Omega$ is a uniform condensate density (c-number) with N_0 the number of the condensed atoms, Ω the volume of the system, and the quantum part $\hat{\phi}$ is assumed to be a small perturbation. Taking $\hat{\phi}$ and $\hat{\phi}^\dagger$ terms up to quadratic, Bogoliubov built the "first-oder" theory of uniform Bose gas.

This idea can be extended to non-uniform gases in trap potentials. If we introduce the $\mathbf{r}$ dependence of the condensate part, the field operator is expressed as

$$\hat{\Psi}(\mathbf{r},t) \simeq \Phi(\mathbf{r},t) + \hat{\phi}(\mathbf{r},t). \tag{9}$$

The scalar function $\Phi(\mathbf{r}, t)$ is called the *condensate wave function*, which is normalized to be the number of the condensed atoms,

$$\int d\mathbf{r} |\Phi(\mathbf{r})|^2 = N_0. \tag{10}$$

In the case of BEC, the number of the condensed atoms becomes macroscopic, i.e.,

$$N - N_0 \ll N_0 < N. \tag{11}$$

In this sense, the "macroscopic" wave function $\Phi(\mathbf{r}, t)$ is related to the first quantized N-body wave function $\Phi_N(\mathbf{r}_1, \ldots, \mathbf{r}_N; t)$ as

$$\Phi_N(\mathbf{r}_1,\ldots,\mathbf{r}_N;t) \simeq \Pi_i \Phi(\mathbf{r}_i,t), \tag{12}$$

which obviously satisfies the symmetry under exchanges of two bosons.

Following the Bogoliubov prescription, we substitute (9) into (1) and retain $\hat{\phi}$ and $\hat{\phi}^\dagger$ terms up to quadratic;

$$\hat{H} \simeq E_\Phi + \hat{H}_1 + \hat{H}_2, \tag{13}$$

$$E_\Phi = \int d\mathbf{r} \Phi^*(\mathbf{r},t) \left[-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{trap}}(\mathbf{r}) + \frac{g}{2}|\Phi(\mathbf{r},t)|^2 \right] \Phi(\mathbf{r},t), \tag{14}$$

$$\hat{H}_1 = \int d\mathbf{r} \Phi^*(\mathbf{r},t) \left[-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{trap}}(\mathbf{r}) + g|\Phi(\mathbf{r},t)|^2 \right] \hat{\phi}(\mathbf{r},t) + \text{h.c.}, \tag{15}$$

$$\begin{aligned}\hat{H}_2 &= \int d\mathbf{r}\hat{\phi}^\dagger(\mathbf{r},t)\left[-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{trap}}(\mathbf{r}) + 2g|\Phi(\mathbf{r},t)|^2\right]\hat{\phi}(\mathbf{r},t)\\ &+ \frac{g}{2}\int d\mathbf{r}\left[\hat{\phi}^\dagger(\mathbf{r},t)\hat{\phi}^\dagger(\mathbf{r},t)\Phi(\mathbf{r},t)^2 + \Phi^*(\mathbf{r},t)^2\hat{\phi}(\mathbf{r},t)\hat{\phi}(\mathbf{r},t)\right].\end{aligned} \tag{16}$$

Equation (14) is called the Gross-Pitaevskii energy functional. The statistical and dynamical properties of the condensate are determined through a variation of E_Φ while the low-lying excitations from the ground state can be analyzed by diagonalizing $\hat{H}_2$. In the ground state, $\hat{H}_1$ part vanishes identically.

2.3 Gross-Pitaevskii equation

Even at the zero temperature, interactions may cause quantum correlation which gives rise to occupation in the excited states. The assumption that the quantum fluctuation part $\hat{\phi}$ $(\mathbf{r}, t)$ gives a small contribution to the condensate is valid for a dilute system. In particular, if we consider a dilute limit:

$$na^3 \ll 1, \tag{17}$$

where na^3 is the gas parameter with n the particle number density, neglecting $\hat{\phi}$ parts provides an appropriate description of the condensate wave function at zero temperature. By a variational principle,

$$\imath\hbar\frac{\partial}{\partial t}\Phi(\mathbf{r},t) = \frac{\delta E_\Phi}{\delta\Phi^*(\mathbf{r},t)}, \tag{18}$$

we obtain the Gross-Pitaevskii (GP) equation:

$$\imath\hbar\frac{\partial}{\partial t}\Phi(\mathbf{r},t) = \left[-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{trap}}(\mathbf{r}) + g\,|\Phi(\mathbf{r},t)|^2\right]\Phi(\mathbf{r},t). \tag{19}$$

This equation has been derived independently by Gross and Pitaevskii (Pethick & Smith; 2002) to deal with the superfluidity of ^{4}He-II. The GP equation is a classical field equation for a scalar (complex) function Φ but contains $\hbar$ explicitly. In this sense, the description of the condensate in terms of Φ is a manifestation of the macroscopic de Broglie wave, where the corpuscular aspect of matter dose not play a role. Now the modulus and gradient of phase of $\Phi = |\Phi|\exp(i\varphi)$ have a clear physical meaning,

$$n(\mathbf{r},t) = |\Phi(\mathbf{r},t)|^2, \qquad \mathbf{v}(\mathbf{r},t) = \frac{\hbar}{m}\nabla\varphi(\mathbf{r},t), \tag{20}$$

where n and $\mathbf{v}$ denote number density and velocity of the condensate, respectively.

3. Confinement induced resonance

In this section, we derive an effective one-dimensional (1D) Hamiltonian for bosons confined in an elongated trap. The interactions between atoms in the experiments are always three-dimensional (3D) even when the kinetic motion of the atoms in such a tight radial confinement is 1D like. Therefore, the trap-induced corrections to the strength of the atomic interactions should be taken into account properly.

This problem was first solved by Olshanii (Olshanii; 1998) within the pseudopotential approximation, yielding a new type of tuning mechanism for the scattering amplitude, now called *confinement induced resonance* (CIR). In what follows, we show a detailed account of a renormalization of the 3D interaction into an effective 1D interaction, which produces the CIR. This technique plays a crucial role in Sec. 5 in order to realize an integrable condition for spinor GP equations.

3.1 Model Hamiltonian

We start with the following model:

1. The trap potential is composed by an axially symmetric 2D harmonic potential of a frequency $\omega_\perp$ in the x-y plane.
2. Atomic motion for the z direction is free.
3. Interatomic interaction potential is represented by the Fermi-Huang pseudopotential:

$$v(r) = g\delta(\mathbf{r})\frac{\partial}{\partial r}(r\cdot), \tag{21}$$

   where the coupling strength g is expressed by the 3D s-wave scattering length a as eq. (7) (Meystre; 2002).
4. The energy of atoms for both transverse and longitudinal motions is well below the transverse vibrational energy $\hbar\omega_\perp$.

In the harmonic potential we can separate the center of mass and relative motion. Then we consider the Schrödinger equation for the relative motion,

$$\left[-\frac{\hbar^2}{2m_\mathrm{r}}\frac{\partial^2}{\partial z^2} + g\delta(\mathbf{r})\frac{\partial}{\partial r}(r\cdot) + \hat{H}_\perp\right]\Psi(\mathbf{r}) = E\Psi(\mathbf{r}), \tag{22}$$

where the reduced mass $m_\mathrm{r} = m/2$, the relative coordinate $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, and the transverse Hamiltonian:

$$\hat{H}_\perp \equiv -\frac{\hbar^2}{2m_\mathrm{r}}\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right] + \frac{m_\mathrm{r}\omega_\perp^2}{2}(x^2 + y^2). \tag{23}$$

From the above condition 4, we assume that the incident wave is factorized as $\mathrm{e}^{\mathrm{i}k_z z}\phi_{n=0,m_z=0}(r_\perp)$, where $\phi_{n=0,m_z=0}(r_\perp)$ is the transverse ground state ($r_\perp^2 \equiv x^2 + y^2$). The longitudinal kinetic energy is smaller than the energy separation between the ground state and the first axially symmetric excited state:

$$\frac{\hbar^2 k_z^2}{2m_\mathrm{r}} < E_{n=2,m_z=0} - E_{n=0,m_z=0} = 2\hbar\omega_\perp, \tag{24}$$

where $E_{n,m_z} = \hbar\omega_\perp(n+1)$ is the energy spectrum of 2D harmonic oscillator with $n = 0, 1, 2, \ldots$ the principal quantum number, and m_z the angular momentum around the z axis, which takes on values $m_z = 0, 2, 4, \ldots, n$ $(1, 3, 5, \ldots, n)$ for even (odd) n.

3.2 One-dimensional scattering amplitude

The asymptotic form of the scattering wave function is given by

$$\Psi(z,r_\perp) \to \left[\mathrm{e}^{\mathrm{i}k_z z} + f_{\mathrm{even}}\mathrm{e}^{\mathrm{i}k_z|z|} + \mathrm{sgn}(z) f_{\mathrm{odd}}\mathrm{e}^{\mathrm{i}k_z|z|} \right] \phi_{0,0}(r_\perp), \quad \text{as } |z| \to \infty, \tag{25}$$

where f_{even} and f_{odd} denote the one-dimensional scattering amplitudes for the even and odd partial waves, respectively. While the transverse state ($n = m_z = 0$) remains unchanged under the assumption of low energy scattering considered above, the scattering amplitudes $f_{\mathrm{even,odd}}$ are affected by a virtual excited state of the axially symmetric modes ($n > 0, m_z = 0$) during the collision.

To calculate the one-dimensional scattering amplitude we expand the solution,

$$\Psi(z,r_\perp) = \mathrm{e}^{\mathrm{i}k_z z}\phi_{0,0}(r_\perp) + \sum_{n=\mathrm{even}} A_n \mathrm{e}^{\mathrm{i}k_z^{(n)}|z|}\phi_{n,0}(r_\perp), \tag{26}$$

where $\phi_{n=\mathrm{even},0}$ is the (axially symmetric) eigenstate of the transverse Hamiltonian (23), and substitute this expansion into eq. (22) with the eigenvalue $E = \hbar^2 k_z^2/2m_\mathrm{r} + \hbar\omega_\perp$. Operating

$$2\pi \int_0^\infty \mathrm{d}r_\perp r_\perp \phi_{0,0}^*(r_\perp) \int_{-\varepsilon}^{\varepsilon} \mathrm{d}z \tag{27}$$

to both side of the Schrödinger equation and taking the limit, in sequence, $\varepsilon \to 0^+$, $z \to \infty$, along with the asymptotic form (25), we can obtain $k_z^{(0)} = k_z$ and the following expression for the scattering amplitudes:

$$f_{\mathrm{even}} \equiv A_0 = -\frac{\mathrm{i}g m_\mathrm{r}}{\hbar^2 k_z}\phi_{0,0}^*(0)\Psi_{\mathrm{reg}}; \quad f_{\mathrm{odd}} = 0. \tag{28}$$

Here we have used the normalization condition:

$$2\pi \int_0^\infty \mathrm{d}r_\perp r_\perp |\phi_{0,0}(r_\perp)|^2 = 1, \tag{29}$$

and the $r \to 0$ limit of the regular (free of the $1/r$ divergence) part of the solution Ψ,

$$\Psi_{\mathrm{reg}} \equiv \frac{\partial}{\partial r}[r\Psi(\mathbf{r})]|_{r\to 0} = \frac{\partial}{\partial z}[z\Psi(z, r_\perp = 0)]|_{z\to 0^+}. \tag{30}$$

We note that the regularization operator $\frac{\partial}{\partial r}(r\cdot)$ that removes the $1/r$ divergence from the scattered wave plays an important role in this derivation. All the expansion coefficients A_n ($n = 2, 4, \ldots$) in eq. (26) can be obtained in the same procedure for each mode $\phi_{n,0}(r_\perp)$ with the corresponding imaginary wave number:

$$k_z^{(n)} = \frac{2\mathrm{i}}{a_\perp}\left(\frac{n}{2} - \frac{k_z^2 a_\perp^2}{4}\right)^{1/2}, \tag{31}$$

the normalization condition of $\phi_{n,0}(r_\perp)$ and a simple relation: $|\phi_{n,0}(0)|^2 = 1/(\pi a_\perp^2)$. Here $a_\perp$ is the oscillator length of the (relative) transverse motion,

$$a_\perp = \sqrt{\frac{\hbar}{m_\mathrm{r}\omega_\perp}}. \tag{32}$$

Recall that due to the condition (24) the value inside the parentheses in eq. (31) is positive definite. Thus, the expression for the wave function along the z axis reads

$$\Psi(z, r_\perp = 0) = \frac{1}{\sqrt{\pi} a_\perp} \mathrm{e}^{\mathrm{i}k_z z} - \frac{\mathrm{i} g m_\mathrm{r}}{\pi \hbar^2 k_z a_\perp^2} \Psi_\mathrm{reg} \mathrm{e}^{\mathrm{i}k_z |z|} - \frac{g m_\mathrm{r}}{2\pi \hbar^2 a_\perp} \Psi_\mathrm{reg} \Lambda \left[\frac{2|z|}{a_\perp}, -\left(\frac{k_z a_\perp}{2} \right)^2 \right], \quad (33)$$

where the function Λ is defined as

$$\Lambda[\xi, \epsilon] = \sum_{s'=1}^{\infty} \frac{\exp(-\sqrt{s' + \epsilon}\,\xi)}{\sqrt{s' + \epsilon}}; \quad (34)$$

the sum over $s'= n/2$ originates from the sum appearing in eq. (26). We have chosen the value $\phi_{0,0}(0)$ to be real and positive. By subtracting and adding a sum,

$$\sum_{s'=1}^{\infty} \int_{s'-1}^{s'} \mathrm{d}s'' \frac{\exp(-\sqrt{s''}\,\xi)}{\sqrt{s''}} = \frac{2}{\xi}, \quad (35)$$

to the function Λ, and then, collecting ξ^0 term from the Taylor series of $\exp(-\sqrt{s''}\,\xi)$ and $\exp(-\sqrt{s'' + \epsilon}\,\xi)$ with respect to ξ, one can show an expansion,

$$\Lambda[\xi, \epsilon] = \frac{2}{\xi} + \Lambda^{(0)}(\epsilon) + \Lambda^{(1)}(\epsilon)\xi + \ldots. \quad (36)$$

Here the zero-order term of the expansion has a form,

$$\Lambda^{(0)}(\epsilon) = -C + \bar{\Lambda}^{(0)}(\epsilon), \quad (37)$$

with

$$C = \lim_{s \to \infty} \left(\int_0^s \frac{\mathrm{d}s'}{\sqrt{s'}} - \sum_{s'=1}^{s} \frac{1}{\sqrt{s'}} \right) = -\zeta(1/2) = 1.4603\ldots, \quad (38)$$

and

$$\bar{\Lambda}^{(0)}(\epsilon) = \sum_{s'=1}^{\infty} \left(\frac{1}{\sqrt{s' + \epsilon}} - \frac{1}{\sqrt{s'}} \right) = \sum_{j=1}^{\infty} (-1)^j \frac{\zeta[(1+2j)/2](2j-1)!!\epsilon^j}{2^j j!}. \quad (39)$$

Substituting eq. (33) with eq. (36) into eq. (30), we get Ψ_reg in an explicit form.
We then write the final expression of the one-dimensional scattering amplitudes (25) as

$$f_\mathrm{even} = -\frac{1}{1 + \mathrm{i}k_z a_\mathrm{1D} - \underbrace{(\mathrm{i}k_z a_\perp/2)\bar{\Lambda}^{(0)}(-k_z^2 a_\perp^2/4)}_{\mathcal{O}(k_z^3 a_\perp^3)}}, \quad (40)$$

with the 1D scattering length:

$$a_\mathrm{1D} = -\frac{a_\perp^2}{2a} \left(1 - C \frac{a}{a_\perp} \right). \quad (41)$$

3.3 Effective one-dimensional coupling strength

The expression (40) is an exact result for the potential (21) with arbitrary strength of the transverse confinement $a_\perp$. For atoms with the low kinetic energy, we can drop $\mathcal{O}(k_z^3 a_\perp^3)$ term in the denominator of the scattering amplitudes (40), obtaining a one-dimensional contact potential,

$$v_{1\mathrm{D}}(z) = \bar{g}\delta(z), \tag{42}$$

were the coupling strength:

$$\bar{g} = -\frac{\hbar^2}{m_{\mathrm{r}} a_{1\mathrm{D}}} = \frac{4\hbar^2 a}{m a_\perp^2} \frac{1}{1 - C(a/a_\perp)}. \tag{43}$$

Note that a simple average of the three-dimensional coupling $g = 4\pi\hbar^2 a/m$ over the transverse ground state only reproduces the coefficient of (43),

$$2\pi \int_0^\infty \mathrm{d}r_\perp r_\perp |\phi_{0,0}(r_\perp)|^2 \frac{4\pi\hbar^2 a}{m}\delta(\mathbf{r}) = \frac{4\hbar^2 a}{m a_\perp^2}\delta(z). \tag{44}$$

The resonance factor $1/[1 - C(a/a_\perp)]$ implies a possibility to control the strength of atomatom scattering via tuning a confinement potential $a_\perp$. The physical origin of the CIR is attributed to a zero-energy Feshbach resonance in which the transverse modes of the confining potential assume the roles of “open” and “closed” scattering channels.

4. Spinor Bose–Einstein condensate

In this section, we extend the model of a single component condensate discussed in Sec. 2 to that of a multicomponent condensate with the spin degrees of freedom, which we call a *spinor condensate* for short (Pethick & Smith; 2002). In terms of “spin”, we mean the hyperfine spin of atoms in this chapter.

4.1 Hamiltonian

The hyperfine spin f is defined by $f = s + i$, where s and i denote the electronic and nuclear spins of the atoms. For simplicity, we consider bosons with the hyperfine spin $f = 1$. This includes alkalis with nuclear spin $i = 3/2$ such as ^{7}Li, ^{87}Rb, and ^{23}Na. Alkali bosons with $f > 1$ such as ^{85}Rb (with $i = 5/2$), and ^{133}Cs (with $i = 7/2$) may have even richer structures.

Atoms in the $f = 1$ state are characterized by a vectorial field operator with the components subject to the hyperfine spin manifold. The three-component field $\hat{\Psi} = \{\hat{\Psi}_1, \hat{\Psi}_0, \hat{\Psi}_{-1}\}^T$, where the superscript T denotes the transpose, satisfies the bosonic commutation relations:

$$[\hat{\Psi}_\alpha(\mathbf{r},t), \hat{\Psi}_\beta^\dagger(\mathbf{r}',t)] = \delta_{\alpha,\beta}\delta(\mathbf{r} - \mathbf{r}'). \tag{45}$$

In order to discuss the properties of spinor Bose gases, we start with the following second quantized Hamiltonian,

$$\hat{H} = \hat{H}_0 + \hat{H}_{\mathrm{int}} + \hat{H}_{lz} \tag{46}$$

$$\hat{H}_0 = \int \mathrm{d}\mathbf{r} \sum_\alpha \hat{\Psi}_\alpha^\dagger(\mathbf{r},t) \left[-\frac{\hbar^2}{2m}\nabla^2 + U_{\mathrm{trap}}(\mathbf{r}) \right] \hat{\Psi}_\alpha(\mathbf{r},t), \tag{47}$$

$$\hat{H}_{\text{int}} = \frac{1}{2}\int d\mathbf{r}\int d\mathbf{r}' \sum_{\alpha,\alpha',\beta,\beta'} \hat{\Psi}^{\dagger}_{\alpha}(\mathbf{r},t)\hat{\Psi}^{\dagger}_{\beta}(\boldsymbol{r}',t)v(\mathbf{r}-\mathbf{r}')_{\alpha\alpha'\beta\beta'}\hat{\Psi}_{\alpha'}(\mathbf{r},t)\hat{\Psi}_{\beta'}(\mathbf{r}',t), \tag{48}$$

$$\hat{H}_{lz} = -p\int d\mathbf{r}\sum_{\alpha,\beta}\hat{\Psi}^{\dagger}_{\alpha}(\mathbf{r},t)\mathrm{f}^{z}_{\alpha,\beta}\hat{\Psi}_{\beta}(\mathbf{r},t) \tag{49}$$

where $U_{\text{trap}}(\mathbf{r})$ is the external trap potential, $v(\mathbf{r} - \mathbf{r}')$ expresses the two-body interaction and subscripts $\{\alpha, \beta, \alpha', \beta' = 1, 0, -1\}$ denote the components of the spin. The last term in eq. (46), $\hat{H}_{lz}$, is the response to an external magnetic field p (the linear Zeeman effect). This response to the magnetic field necessarily selects one of several possible ground states, or the so-called weak field seeking state, $m_f = -1$ for $f = 1$ case where the spin degrees of freedom are "frozen". We set $p = 0$ throughout this chapter.

Due to the Bose–Einstein statistics, the total spin $F = f_1 + f_2$ of any two bosons whose relative orbital angular momentum is zero should be restricted to even, $F = 2f, 2f-2, \ldots, 0$. Thus, the interatomic interaction $\hat{v}(\mathbf{r} - \mathbf{r}')$ can be divided into several sectors labeled by F as

$$\hat{v}(\mathbf{r}-\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')\sum_{F=\text{even}} g_F \hat{P}_F, \tag{50}$$

where $\hat{P}_F$ is the projection operator and g_F characterizes the strength of the binary interaction between bosonic atoms with the total spin F. This coupling constant g_F is related to the corresponding s-wave scattering length a_F as

$$g_F = \frac{4\pi\hbar^2 a_F}{m}. \tag{51}$$

For $f = 1$ bosons, since F takes only on values 0 and 2, we can rewrite the potential $\hat{v}(\mathbf{r} - \mathbf{r}')$ in a simple form using the following two properties of the projection operators $\hat{P}_0, \hat{P}_2$; the completeness of the operators,

$$\hat{P}_0 + \hat{P}_2 = \hat{I}, \tag{52}$$

where $\hat{I}$ is an identity operator, and the product of the angular momentum operators,

$$\hat{\mathbf{f}}\cdot\hat{\mathbf{f}}' = \frac{1}{2}\left[\hat{\mathbf{F}}^2 - \hat{\mathbf{f}}^2 - \hat{\mathbf{f}}'^2\right] = \sum_{F=0,2}\frac{1}{2}[F(F+1) - 2f(f+1)]\hat{P}_F = \hat{P}_2 - 2\hat{P}_0, \tag{53}$$

where a hat "ˆ" on **f** means an operator as projection. Solving these equations (52), (53) for $\hat{P}_0$ and $\hat{P}_2$, we obtain the form of the interaction in terms of the angular momentum operators,

$$\hat{v}(\mathbf{r}-\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')(c_0\hat{I} + c_1\hat{\mathbf{f}}_1\cdot\hat{\mathbf{f}}_2). \tag{54}$$

In this expression,

$$c_0 = \frac{2g_2 + g_0}{3}, \qquad c_1 = \frac{g_2 - g_0}{3}, \tag{55}$$

which are the magnitude of the density-density interaction and of the spin-spin interaction, respectively. Thus, the interaction Hamiltonian is rewritten as

$$\begin{aligned}\hat{H}_{\text{int}} =&\frac{c_0}{2}\int \mathrm{d}\mathbf{r}\sum_{\alpha,\beta}\hat{\Psi}^{\dagger}_{\alpha}(\mathbf{r},t)\hat{\Psi}^{\dagger}_{\beta}(\mathbf{r},t)\hat{\Psi}_{\alpha}(\mathbf{r},t)\hat{\Psi}_{\beta}(\mathbf{r},t)\\ &+\frac{c_1}{2}\int \mathrm{d}\mathbf{r}\sum_{\alpha,\alpha',\beta,\beta'}\hat{\Psi}^{\dagger}_{\alpha}(\mathbf{r},t)\hat{\Psi}^{\dagger}_{\beta}(\mathbf{r},t)\mathbf{f}_{\alpha\beta}\cdot\mathbf{f}_{\alpha'\beta'}\hat{\Psi}_{\alpha'}(\mathbf{r},t)\hat{\Psi}_{\beta'}(\mathbf{r},t),\end{aligned} \tag{56}$$

where we may use the following expressions of spin-1 matrices **f** = (f^x, f^y, f^z) as

$$\mathrm{f}^x=\frac{1}{\sqrt{2}}\begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix},\quad \mathrm{f}^y=\frac{\mathrm{i}}{\sqrt{2}}\begin{pmatrix}0&-1&0\\1&0&-1\\0&1&0\end{pmatrix},\quad \mathrm{f}^z=\begin{pmatrix}1&0&0\\0&0&0\\0&0&-1\end{pmatrix}. \tag{57}$$

A construction of the interaction Hamiltonian for a general hyperfine spin *f* can be found in (Ueda & Koashi; 2002).

4.2 *f* = 1 spinor condensate in quasi 1D regime

From now on, we assume that the system is quasi-one dimensional: the trapping potential is suitably anisotropic such that the transverse spatial degrees of freedom (*y*-*z* plain) is factorized from the longitudinal (*x* axis) and all the hyperfine states are in transverse ground state.

As derived in Sec. 2, in the mean-field theory of the spinor BEC, the assembly of atoms in the *f* = 1 state is characterized by a vectorial order parameter:

$$\Phi(x,t)\equiv[\Phi_1(x,t),\Phi_0(x,t),\Phi_{-1}(x,t)]^T \tag{58}$$

where the subscripts {1, 0,–1} denote the magnetic quantum numbers with the components subject to the hyperfine spin space. The normalization is imposed as

$$\int \mathrm{d}x\,\Phi(x,t)^{\dagger}\cdot\Phi(x,t)=N_{\mathrm{T}}, \tag{59}$$

where N_T is the total number of atoms.

According to the discussion in Sec. 3, the effective 1D couplings $\bar{g}_0$ and $\bar{g}_2$are represented by

$$\bar{g}_F=\frac{4\hbar^2 a_F}{m a_{\perp}^2}\frac{1}{1-C(a_F/a_{\perp})}, \tag{60}$$

where a_F is the 3D *s*-wave scattering length of the total hyperfine spin *F* = 0, 2 channels, respectively, and $a_{\perp}$ is the size of the ground state in the (relative) transverse motion.

Thus, the Gross-Pitaevskii energy functional of this system is given by

$$E_{\Phi}=\int \mathrm{d}x\left\{\Phi^*_{\alpha}(x,t)\left[-\frac{\hbar^2}{2m}\partial_x^2\right]\Phi_{\alpha}(x,t)+\frac{\bar{c}_0}{2}n^2(x,t)+\frac{\bar{c}_1}{2}\mathbf{f}^2(x,t)\right\}, \tag{61}$$

with the particle number and spin densities, respectively, defined by

$$n(x,t)\equiv\Phi^*_{\alpha}(x,t)\Phi_{\alpha}(x,t),\qquad \mathbf{f}(x,t)\equiv\Phi^*_{\alpha}(x,t)\mathbf{f}_{\alpha\beta}\Phi_{\beta}(x,t). \tag{62}$$

The coupling constants $\bar{c}_0$ and $\bar{c}_1$ are connected to those in eqs. (60) (cf. eq. (43)) as

$$\bar{c}_0 = \frac{2\bar{g}_2 + \bar{g}_0}{3}, \qquad \bar{c}_1 = \frac{\bar{g}_2 - \bar{g}_0}{3}. \tag{63}$$

The time-evolution of spinor condensate wave function $\Phi(x, t)$ can be derived from

$$i\hbar\partial_t \Phi_\alpha(x,t) = \frac{\delta E_\Phi}{\delta \Phi_\alpha^*(x,t)}. \tag{64}$$

Substituting eq. (61) into eq. (64), we get a set of equations for the longitudinal wave functions of the spinor condensate:

$$\begin{aligned} i\hbar\partial_t \Phi_1(x,t) = -\frac{\hbar^2}{2m}\partial_x^2 \Phi_1(x,t) + (\bar{c}_0 + \bar{c}_1)\left[|\Phi_1(x,t)|^2 + |\Phi_0(x,t)|^2\right]\Phi_1(x,t) \\ + (\bar{c}_0 - \bar{c}_1)|\Phi_{-1}(x,t)|^2\Phi_1(x,t) + \bar{c}_1\Phi_{-1}^*(x,t)\Phi_0^2(x,t), \end{aligned} \tag{65a}$$

$$\begin{aligned} i\hbar\partial_t \Phi_0(x,t) = -\frac{\hbar^2}{2m}\partial_x^2 \Phi_0(x,t) + (\bar{c}_0 + \bar{c}_1)\left[|\Phi_1(x,t)|^2 + |\Phi_{-1}(x,t)|^2\right]\Phi_0(x,t) \\ + \bar{c}_0|\Phi_0(x,t)|^2\Phi_0(x,t) + 2\bar{c}_1\Phi_0^*(x,t)\Phi_1(x,t)\Phi_{-1}(x,t), \end{aligned} \tag{65b}$$

$$\begin{aligned} i\hbar\partial_t \Phi_{-1}(x,t) = -\frac{\hbar^2}{2m}\partial_x^2 \Phi_{-1}(x,t) + (\bar{c}_0 + \bar{c}_1)\left[|\Phi_{-1}(x,t)|^2 + |\Phi_0(x,t)|^2\right]\Phi_{-1}(x,t) \\ + (\bar{c}_0 - \bar{c}_1)|\Phi_1(x,t)|^2\Phi_{-1}(x,t) + \bar{c}_1\Phi_1^*(x,t)\Phi_0^2(x,t). \end{aligned} \tag{65c}$$

5. Integrable model

To analyze the dynamical properties of the coupled system (65), we propose an integrable model as follows (Ieda et al.; 2004a,b). We consider the system with the coupling constants,

$$\bar{c}_0 = \bar{c}_1 \equiv -c < 0, \qquad (2\bar{g}_0 = -\bar{g}_2 > 0). \tag{66}$$

This situation corresponds to attractive mean-field interaction $\bar{c}_0 < 0$ and ferromagnetic spin-exchange interaction $\bar{c}_1 < 0$. Note that in preceding investigations of spinor condensates (Pethick & Smith; 2002), mean-field interaction is assumed to be repulsive $c_0 > 0$ and far exceeding spin-exchange interaction in the magnitude $c_0 \gg |c_1|$ in line with experimental data. Thus, the parameter regime (66) was not been explored in detail.

The effective interactions between atoms in a BEC have been tuned with a Feshbach resonance (Pethick & Smith; 2002). In spinor BECs, however, we should extend this to alternative techniques such as an optically induced Feshbach resonance or a confinement induced resonance (Olshanii; 1998), which do not affect the rotational symmetry of the internal spin states. In the latter, the above condition is surely obtained by setting

$$a_\perp = 3C\frac{a_0 a_2}{2a_0 + a_2}, \tag{67}$$

in eq. (60) when

$$a_0 > a_2 > 0 \qquad \text{or} \qquad a_2 > 0 > a_0. \tag{68}$$

It is worth noting that the integrable property itself is independent of the sign of $\bar{c}_0$ ($\bar{c}_1$) as far as their magnitudes are equal to each other. The opposite sign case, i.e., $\bar{c}_0$= $\bar{c}_1 \equiv c$ > 0, can be analyzed in the same manner (Uchiyama et al.; 2006).
In the dimensionless form:

$$\mathbf{\Phi} \to \{\phi_1, \sqrt{2}\phi_0, \phi_{-1}\}^T, \tag{69}$$

where time and length are measured in units of

$$\bar{t} = \frac{\hbar a_\perp}{c}, \qquad \bar{x} = \hbar\sqrt{\frac{a_\perp}{2mc}}, \tag{70}$$

respectively, we rewrite eqs. (65) as follows, (we omit the arguments (*x*, *t*) hereafter.)

$$\mathrm{i}\partial_t\phi_1 = -\partial_x^2\phi_1 - 2\{|\phi_1|^2 + 2|\phi_0|^2\}\phi_1 - 2\phi_{-1}^*\phi_0^2, \tag{71a}$$

$$\mathrm{i}\partial_t\phi_0 = -\partial_x^2\phi_0 - 2\{|\phi_{-1}|^2 + |\phi_0|^2 + |\phi_1|^2\}\phi_0 - 2\phi_0^*\phi_1\phi_{-1}, \tag{71b}$$

$$\mathrm{i}\partial_t\phi_{-1} = -\partial_x^2\phi_{-1} - 2\{|\phi_{-1}|^2 + 2|\phi_0|^2\}\phi_{-1} - 2\phi_1^*\phi_0^2. \tag{71c}$$

Now we find that these coupled equations (71) are equivalent to a 2×2 matrix version of nonlinear Schrödinger (NLS) equation:

$$\mathrm{i}\partial_t Q + \partial_x^2 Q + 2QQ^\dagger Q = O, \tag{72}$$

with an identification,

$$Q = \begin{pmatrix} \phi_1 & \phi_0 \\ \phi_0 & \phi_{-1} \end{pmatrix}. \tag{73}$$

Since the matrix NLS equation (72) is completely integrable (Tsuchida & Wadati; 1998), the integrability of the reduced equations (71) are proved automatically (Ieda et al.; 2004a). Remark that the general *M*× *L* matrix NLS equation is also integrable. It is worthy to search other integrable models for higher spin case (Uchiyama et al.; 2007).

5.1 Soliton solution

We summarize an explicit formula for the soliton solution of the 2 × 2 matrix version of NLS equation (72) with eq. (73) by considering a reduction of a general formula obtained in (Tsuchida & Wadati; 1998).
Under the vanishing boundary condition, one can apply the inverse scattering method (ISM) to the nonlinear time evolution equation (72) associated with the generalized Zakharov-Shabat eigenvalue problem:

$$\partial_x \begin{bmatrix} \Psi_1 \\ \Psi_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} k^* I & 2Q \\ -2Q^\dagger & -k^* I \end{bmatrix}\begin{bmatrix} \Psi_1 \\ \Psi_2 \end{bmatrix}. \tag{74}$$

Here Ψ_1 and Ψ_2 take their values in 2 × 2 matrices. The complex number k is the spectral parameter. I is the 2 × 2 unit matrix. The 2 × 2 matrix Q plays a role as a potential function in this linear system. According to (Tsuchida & Wadati; 1998), N-soliton solution of eq. (72) with eq. (73) is expressed as

$$Q(x,t) = (\underbrace{I\,I \cdots I}_{N})S^{-1}\begin{bmatrix} \Pi_1 \mathrm{e}^{\chi_1} \\ \Pi_2 \mathrm{e}^{\chi_2} \\ \vdots \\ \Pi_N \mathrm{e}^{\chi_N} \end{bmatrix}, \tag{75}$$

where the $2N$ × $2N$ matrix S is given by

$$S_{ij} = \delta_{ij} I + \sum_{l=1}^{N} \frac{\Pi_i \cdot \Pi_l^\dagger}{(k_i + k_l^*)(k_j + k_l^*)} \mathrm{e}^{\chi_i + \chi_l^*}, \qquad 1 \le i,j \le N. \tag{76}$$

Here we have introduced the following parameterizations:

$$\Pi_j = \begin{pmatrix} \beta_j & \alpha_j \\ \alpha_j & \gamma_j \end{pmatrix}, \tag{77}$$

$$\chi_j \equiv \chi_j(x,t) = k_j x + \mathrm{i}k_j^2 t - \epsilon_j. \tag{78}$$

The 2 × 2 matrices Π_j normalized to unity in a sense of the square norm,

$$||\Pi_j||_2 \equiv \sqrt{2|\alpha_j|^2 + |\beta_j|^2 + |\gamma_j|^2} = 1, \tag{79}$$

must take the same form as Q from their definition. We call them "polarization matrices," which determine both the populations of three components {1, 0, –1} within each soliton and the relative phases between them. The complex constants k_j denote discrete eigenvalues, each of which determines a bound state by the potential Q. ϵ_j are real constants which can be used to tune the initial displacements of solitons. It is worth noting that all x and t dependence is only through the variables $\chi_j(x, t)$. As we shall see in Sec. 6, the real part of $\chi_j(x, t)$ represents the coordinate for observing soliton-j's envelope while the imaginary part of it represents the coordinate for observing soliton-j's carrier waves.

The same procedure can be performed for nonvanishing boundary conditions (Ieda et al.; 2007) which is relevant to formation of spinor dark solitons (Uchiyama et al.; 2006).

Equation (72) is a completely integrable system whose initial value problems can be solved via, for example, the ISM (Tsuchida & Wadati; 1998) (Ieda et al.; 2007). The existence of the **r**-matrix for this system guarantees the existence of an infinite number of conservation laws which restrict the dynamics of the system in an essential way. Here we show explicit forms of some conserved quantities, i.e., total number, total spin (magnetization), total momentum and total energy.

$$\textbf{total number:} \quad N_\mathrm{T} = \int \mathrm{d}x\, n(x,t); \tag{80}$$

$$n(x,t) = \Phi^\dagger \cdot \Phi = \mathrm{tr}\{Q^\dagger Q\}. \tag{81}$$

$$\textbf{total spin:}\quad \mathbf{F}_{\mathrm{T}} = \int \mathrm{d}x\, \mathbf{f}(x,t)\,; \tag{82}$$

$$\mathbf{f}(x,t) = \Phi^{\dagger}\cdot\mathbf{f}\cdot\Phi = \mathrm{tr}\{Q^{\dagger}\boldsymbol{\sigma}Q\}. \tag{83}$$

$$\textbf{total momentum:}\quad P_{\mathrm{T}} = \int \mathrm{d}x\, p(x,t)\,; \tag{84}$$

$$p(x,t) = -\mathrm{i}\hbar\Phi^{\dagger}\cdot\partial_x\Phi = -\mathrm{i}\hbar\cdot\mathrm{tr}\{Q^{\dagger}Q_x\}. \tag{85}$$

$$\textbf{total energy:}\quad E_{\mathrm{T}} = \int \mathrm{d}x\, e(x,t)\,; \tag{86}$$

$$e(x,t) = \frac{\hbar^2}{2m}\partial_x\Phi^{\dagger}\cdot\partial_x\Phi - \frac{c}{2}\left[n(x,t)^2 + \mathbf{f}(x,t)^2\right] = c\cdot\mathrm{tr}\{Q_x^{\dagger}Q_x - Q^{\dagger}QQ^{\dagger}Q\}. \tag{87}$$

Here tr{·} denotes the matrix trace and $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)^T$ are the Pauli matrices,

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{88}$$

6. Spin property of one-soliton solution

In this section, we discuss one-soliton solutions and classify them by their spin states. If we set $N = 1$ in the formula (75) we obtain the one-soliton solution:

$$Q = \frac{\mathrm{e}^{\chi}}{\det S}\begin{pmatrix} \beta + \gamma^*\mathrm{e}^{2\chi_{\mathrm{R}}+\rho}\det\Pi & \alpha - \alpha^*\mathrm{e}^{2\chi_{\mathrm{R}}+\rho}\det\Pi \\ \alpha - \alpha^*\mathrm{e}^{2\chi_{\mathrm{R}}+\rho}\det\Pi & \gamma + \beta^*\mathrm{e}^{2\chi_{\mathrm{R}}+\rho}\det\Pi \end{pmatrix}, \tag{89}$$

where

$$\det S = 1 + \mathrm{e}^{2\chi_{\mathrm{R}}+\rho} + \mathrm{e}^{4\chi_{\mathrm{R}}+2\rho}|\det\Pi|^2, \qquad \mathrm{e}^{\rho/2} \equiv \frac{1}{2k_{\mathrm{R}}}, \qquad \Pi \equiv \begin{pmatrix} \beta & \alpha \\ \alpha & \gamma \end{pmatrix}, \tag{90}$$

$$\chi_{\mathrm{R}} \equiv \chi_{\mathrm{R}}(x,t) = k_{\mathrm{R}}(x - 2k_{\mathrm{I}}t) - \epsilon, \qquad \chi_{\mathrm{I}} \equiv \chi_{\mathrm{I}}(x,t) = k_{\mathrm{I}}x + (k_{\mathrm{R}}^2 - k_{\mathrm{I}}^2)t. \tag{91}$$

We have omitted the subscripts of the soliton number. Here and hereafter, the subscripts R and I denote real and imaginary parts, respectively. Throughout this section, we set $k_R > 0$ without loss of generality. We remark the significance of each parameter/coordinate as follows,

Π : polarization matrix of soliton
k_{R} : amplitude of soliton
$2k_{\mathrm{I}}$: velocity of soliton's envelope
χ_{R} : coordinate for observing soliton's envelope
χ_{I} : coordinate for observing soliton's carrier wave.

We use the term "amplitude" to indicate the peak(s) height of soliton's envelope. Actual amplitude should be represented as k_{R} multiplied by a factor from 1 to $\sqrt{2}$ which is determined by the type of polarization matrices. The explicit form will be shown later. As mentioned before, soliton's motion depends on both x and t via variables χ_{R} and χ_{I}, from which we can see the meaning of velocity of soliton.

From a total spin conservation, one-soliton solution can be classified by the spin states. We shall show that the only two spin states are allowable,

$$\mathbf{F}_{\mathrm{T}} = (0,0,0)^T \qquad \text{for} \qquad \det\Pi \neq 0, \tag{92a}$$

$$|\mathbf{F}_{\mathrm{T}}| = N_{\mathrm{T}} \qquad \text{for} \qquad \det\Pi = 0. \tag{92b}$$

Substituting eqs. (89)–(91) into eq. (83), we obtain the local spin density of the one-soliton solution:

$$\mathbf{f}(x,t) = \frac{\mathrm{e}^{2\chi_{\mathrm{R}}}}{(\det S)^2}\left(1 - \mathrm{e}^{4\chi_{\mathrm{R}}+2\rho}|\det\Pi|^2\right)\mathrm{tr}\{\Pi^\dagger\boldsymbol{\sigma}\Pi\}. \tag{93}$$

We also give the explicit form of the number density:

$$n(x,t) = \frac{\mathrm{e}^{2\chi_{\mathrm{R}}}}{(\det S)^2}\left\{1 + \left(4\mathrm{e}^{2\chi_{\mathrm{R}}+\rho} + \mathrm{e}^{4\chi_{\mathrm{R}}+2\rho}\right)|\det\Pi|^2\right\}. \tag{94}$$

To clarify the physical meaning of detΠ, we define here another important local density as

$$\Theta(x,t) \equiv \Phi_0^2 - 2\Phi_1\Phi_{-1} = -2\det Q. \tag{95}$$

This quantity measures the formation of singlet pairs. Note that these "pairs" are distinguished from Cooper pairs of electrons or those of ^{3}He owing to the different statistical properties of ingredient particles. Since $\Theta(x, t)$ does not contribute to the magnetization of the soliton, it is invariant under any spin rotation. As far as ground state properties are concerned, it is not necessary to introduce $\Theta(x, t)$ for a system of spin-1 bosons, while a counterpart to eq. (95) plays a crucial role for spin-2 case (Ueda & Koashi; 2002). As we shall show later, however, it is useful to characterize solitons within energy degenerated states.

In the case of the one-soliton solution (89), the singlet pair density is proportional to the determinant of the polarization matrix Π,

$$\Theta(x,t) = -2\frac{\mathrm{e}^{2\chi}}{\det S}\det\Pi. \tag{96}$$

This suggests that detΠ represents the magnitude of the singlet pairs. For the general N-soliton case, this singlet pair density can vary after each collision of solitons and is not the conserved density. The detail will be discussed at the end of this section.

In what follows, we classify spin states of the one-soliton solution based on the values of detΠ.

6.1 Ferromagnetic state

Let detΠ = 0, then eq. (89) becomes a simple form:

$$Q = k_{\mathrm{R}}\mathrm{sech}(\chi_{\mathrm{R}} + \rho/2)\Pi\mathrm{e}^{\mathrm{i}\chi_{\mathrm{I}}}. \tag{97}$$

Now all of $m_F = 0, \pm1$ components share the same wave function. Their distribution in the internal state reflects directly the elements of the polarization matrix Π. One can see the meaning of each parameter listed above. By definition, the singlet pair density (96) vanishes everywhere. Thus, this type of soliton belongs to the *ferromagnetic* state and will be referred to as a ferromagnetic soliton. The total number of atoms is given by integrating eq. (94) as

$$N_{\mathrm{T}} = 2k_{\mathrm{R}}. \tag{98}$$

The total magnetization (82) becomes

$$\mathbf{F}_{\mathrm{T}} = 2k_{\mathrm{R}}\left(2\,\mathrm{Re}\{\alpha^*(\beta+\gamma)\}, -2\,\mathrm{Im}\{\alpha^*(\beta-\gamma)\}, |\beta|^2 - |\gamma|^2\right)^T, \tag{99}$$

with the modulus, $|\mathbf{F}_{\mathrm{T}}| = N_{\mathrm{T}}$. Equation (99) is connected to $\mathbf{F}'_{\mathrm{T}} = 2k_{\mathrm{R}}(0,0,1)^T$ through a gauge transformation and a spin rotation.

Next, we calculate the total momentum and the total energy of the ferromagnetic soliton. Substituting eq. (97) into eqs. (84), (86) and using $\det\Pi = 0$, we obtain

$$P_{\mathrm{T}}^{\mathrm{f}} = N_{\mathrm{T}}\hbar k_{\mathrm{I}}, \qquad E_{\mathrm{T}}^{\mathrm{f}} = N_{\mathrm{T}} c\left(k_{\mathrm{I}}^2 - \frac{k_{\mathrm{R}}^2}{3}\right), \tag{100}$$

respectively. In infinite homogeneous 1D space as considered here, it can be shown that a single component GP equation for BEC with attractive interactions, i.e., the self-focusing NLS equation possesses the one-soliton solution that minimizes the total energy for fixed number of particles and total momentum. This remains true for the spinor GP equations (71). As we will see later, for given number of N_{T}, the stationary ($k_{\mathrm{I}} = 0$) one-soliton solution in the ferromagnetic state is the ground state of this system. On the other hand, in finite 1D space case, the ground state is subject to a quantum phase transition between uniform and soliton states (Kanamoto et al.; 2002).

6.2 Polar state

If $\det\Pi \neq 0$, the local spin density has one node, i.e., $\mathbf{f}(x_0, t) = 0$ at a point:

$$x_0 = 2k_{\mathrm{I}}t + \frac{1}{2k_{\mathrm{R}}}\left(\ln\frac{4k_{\mathrm{R}}^2}{|\det\Pi|} + 2\epsilon\right), \tag{101}$$

for each moment t. Setting $x' = x - x_0$ and $A^{-1} \equiv 2\,|\det\Pi\,|$, we get

$$\mathbf{f}(x') = -\frac{4k_{\mathrm{R}}^2 A\sinh(2k_{\mathrm{R}}x')}{\left[\cosh(2k_{\mathrm{R}}x') + A\right]^2}\begin{pmatrix} 2\,\mathrm{Re}\{\alpha^*(\beta+\gamma)\} \\ -2\,\mathrm{Im}\{\alpha^*(\beta-\gamma)\} \\ |\beta|^2 - |\gamma|^2 \end{pmatrix}. \tag{102}$$

Since each component of the local spin density is an odd function of x', its average value is zero,

$$\mathbf{F}_{\mathrm{T}} = \int \mathrm{d}x'\,\mathbf{f}(x') = (0,0,0)^T. \tag{103}$$

This implies that this type of soliton, on the average, belongs to the *polar* state (Pethick & Smith; 2002). Let us also rewrite the number density (94) as

$$n(x') = \frac{4k_{\mathrm{R}}^2 \left[A \cosh 2k_{\mathrm{R}} x' + 1\right]}{\left[\cosh 2k_{\mathrm{R}} x' + A\right]^2}. \tag{104}$$

To elaborate on this type of soliton, we further divide into two cases.

(i) $A^{-1} = 2|\det\Pi| = 1$ ($\alpha\beta^* + \alpha^*\gamma$=0).

Under this constraint, we find the local spin (102) vanishes everywhere. Solitons in this state possess the symmetry of polar state locally. We, therefore, refer to only those solitons as polar solitons. Considering eq. (89) with the above condition, we recover a normal sech-type soliton solution:

$$Q = \sqrt{2} k_{\mathrm{R}} \operatorname{sech}(k_{\mathrm{R}} x') \Pi \mathrm{e}^{\mathrm{i}\chi_{\mathrm{I}}}. \tag{105}$$

Note that the amplitude of soliton is different from that of the ferromagnetic soliton, which leads to a relation between the total number and the spectral parameter as

$$N_{\mathrm{T}} = 4k_{\mathrm{R}}. \tag{106}$$

The total momentum and the total energy are given by

$$P_{\mathrm{T}}^{\mathrm{p}} = N_{\mathrm{T}} \hbar k_{\mathrm{I}}, \qquad E_{\mathrm{T}}^{\mathrm{p}} = N_{\mathrm{T}} c \left(k_{\mathrm{I}}^2 - \frac{k_{\mathrm{R}}^2}{3} \right), \tag{107}$$

respectively. The difference between ferromagnetic soliton energy and polar soliton energy with the same number of atoms N_T is

$$E_{\mathrm{T}}^{\mathrm{f}} - E_{\mathrm{T}}^{\mathrm{p}} = -\frac{N_{\mathrm{T}}^3 c}{16} < 0, \tag{108}$$

which is a natural consequence of the ferromagnetic interaction, i.e., $\bar{c}_1 = -c < 0$.

(ii) $A^{-1} = 2|\det\Pi| < 1$.

In this case, the local spin retains nonzero value, although the average spin amounts to be zero. The density profile (104) has the following structure. When $A > 2$, a peak of the density splits into two (Fig. 1) due to different density profiles of $m_F = 0, \pm 1$ components.

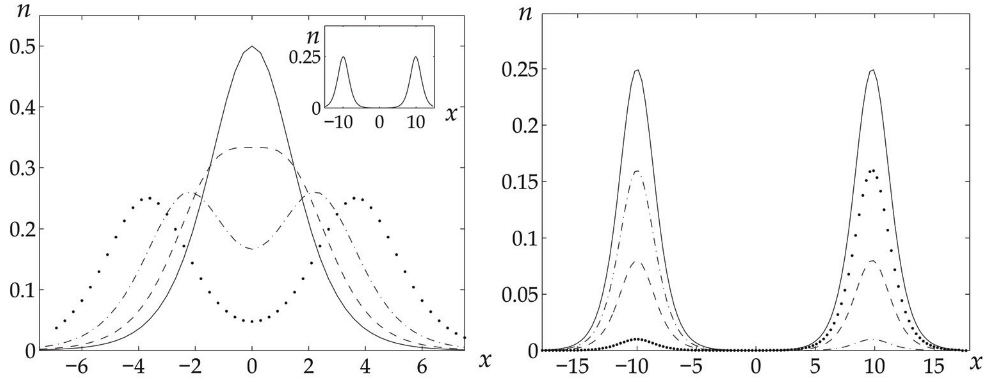
For a large value of A, namely, when $\det\Pi$ gets close to zero, such twin peaks separate away. In consequence, they behave as if a pair of two distinct ferromagnetic solitons with antiparallel spins, traveling in parallel with the same velocity and the amplitudes half as much as that of the polar soliton ($A = 1$) in the density profile [see the inset of Fig. 1(a) and Fig. 1 (b)].

Hence, solitons of this type will be referred to as split solitons. The total number is the same as the case (i),

$$N_{\mathrm{T}} = 4k_{\mathrm{R}}. \tag{109}$$

The total momentum and the total energy are the same values as those in the case (i):

$$P_{\mathrm{T}}^{\mathrm{s}} = N_{\mathrm{T}} \hbar k_{\mathrm{I}}, E_{\mathrm{T}}^{\mathrm{s}} = N_{\mathrm{T}} c \left(k_{\mathrm{I}}^2 - \frac{k_{\mathrm{R}}^2}{3} \right). \tag{110}$$



(a) (b)

Fig. 1. The density profiles of eq. (104). (a)We set $k_{\mathrm{R}} = 0.5$, and $A = 1$ (solid line), 2 (dashed line), 5 (dash-dot line), 20 (dotted line). The inset shows a split soliton for $A = 10^4$, consisting of two ferromagnetic like solitons with the same velocity. (b) The density profiles of eq. (104) (solid line) for $k_{\mathrm{R}} = 0.5$ and $A = 10^4$, and the three components, $m_F = 0$ (dashed line), $m_F = 1$ (dotted line) and $m_F = -1$ (dash-dot line) are shown simultaneously.

This degeneracy is ascribed to the integrable condition for the coupling constants, i.e., $\bar{c}_0 = \bar{c}_1$. Comparing case (i) with case (ii), we find that a variety of dissimilar shaped solitons are degenerated in the polar state. To characterize them, we can use, instead of A, a physical quantity defined as

$$\mathcal{S} \equiv \int \mathrm{d}x |\Theta(x,t)| = N_{\mathrm{T}} \frac{2\tan^{-1}\left(\sqrt{\frac{A-1}{A+1}}\right)}{\sqrt{A^2-1}}, \tag{111}$$

which is a monotone decreasing function of $A \in [1, \infty)$; the maximum value, N_{T}, at $A = 1$ (polar soliton) and limiting to 0 at $A \to \infty$ (ferromagnetic soliton). In this sense, $\mathcal{S}$ has the meaning as the "total singlet pairs" of the whole system. As noted above, $\mathcal{S}$ is not the conserved quantity in general ($N \geq 2$); all the conserved densities should be expressed by the matrix trace of products of $Q^\dagger$, Q and their derivatives (Tsuchida & Wadati; 1998) as eqs. (81), (83), (85), and (87) while $|\Theta(x, t)|$ is not. Nevertheless, $\mathcal{S}$ can be used to label solitons in the polar state because it dose not change in the meanwhile prior to the subsequent collision.

7. Two-soliton collision and spin dynamics

In this section, we analyze two-soliton collisions in the spinor model. The two-soliton solutions can be obtained by setting $N = 2$ in eq. (75). The derivation is straightforward but rather lengthy. An explicit expression of the two-soliton solution is given in Appendix of (Ieda et al.; 2004b) and, here, compute asymptotic forms of specific two-soliton solutions as $t \to \mp \infty$, which define the collision properties of two-soliton in the spinor model.

For simplicity, we restrict the spectral parameters to regions:

$$k_{1\mathrm{R}} > 0, \qquad k_{2\mathrm{R}} < 0, \qquad k_{1\mathrm{I}} < 0, \qquad k_{2\mathrm{I}} > 0. \tag{112a}$$

Under the conditions, we calculate the asymptotic forms in the final state ($t\to\infty$) from those in the initial state ($t\to-\infty$). Since each soliton's envelope is located around $x \simeq 2k_{j\mathrm{I}}t$, soliton-1 and soliton-2 are initially isolated at $x \to \pm\infty$, and then, travel to the opposite directions at a velocity of $2k_{1\mathrm{I}}$ and $2k_{2\mathrm{I}}$, respectively. After a head-on collision, they pass through without changing their velocities and arrive at $x\to\mp\infty$ in the final state. Collisional effects appear not only as usual phase shifts of solitons but also as a rotation of their polarization.

According to the classification of one-soliton solutions in the previous section, we choose the following three cases: i) Polar-polar solitons collision, ii) Polar-ferromagnetic solitons collision, iii) Ferromagnetic-ferromagnetic solitons collision. As we shall see later, the polar soliton dose not affect the polarization of the other solitons apart from the total phase factor. On the other hand, ferromagnetic solitons can 'rotate' their partners' polarization, which allows for switching among the internal states.

7.1 Polar-polar solitons collision

We first deal with a collision between two polar solitons defined by k_j and Π_j (j = 1, 2) with the conditions (112) and $\alpha_j\beta_j^* + \alpha_j^*\gamma_j = 0$, equivalently,

$$|\det\Pi_1| = |\det\Pi_2| \equiv \frac{1}{2}. \tag{113}$$

In the asymptotic regions, we can consider each soliton separately. Thus, the initial state is given by the sum of two polar solitons as

$$Q \simeq Q_1^{\mathrm{in}} + Q_2^{\mathrm{in}}, \tag{114}$$

where the asymptotic form of soliton-j (j = 1, 2) is

$$Q_j^{\mathrm{in}} = \sqrt{2}k_{j\mathrm{R}}\mathrm{sech}(\chi_{j\mathrm{R}} + \rho_j/2)\Pi_j\mathrm{e}^{\mathrm{i}\chi_{j\mathrm{I}}}. \tag{115a}$$

These can be proved by taking the limit $\chi_{2\mathrm{R}} \to -\infty$ with keeping $\chi_{1\mathrm{R}}$ finite and, vice versa, $\chi_{1\mathrm{R}} \to -\infty$ with $\chi_{2\mathrm{R}}$ fixed. Phase factors which come from the values of $|\det\Pi_j|$ are absorbed by the arbitrary constants ϵ_j inside $\chi_{j\mathrm{R}}$. In the final state, the opposite limit $\chi_{2\mathrm{R}} \to \infty$ with keeping $\chi_{1\mathrm{R}}$ finite and $\chi_{1\mathrm{R}} \to \infty$ with $|\chi_{2\mathrm{R}}| < \infty$ yields

$$Q \simeq Q_1^{\mathrm{fin}} + Q_2^{\mathrm{fin}}, \tag{116}$$

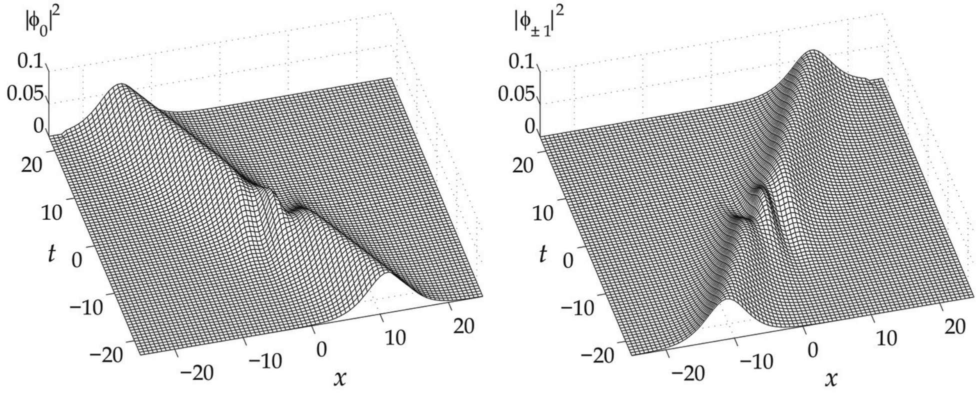
where

$$Q_j^{\mathrm{fin}} = \sqrt{2}k_{j\mathrm{R}}\mathrm{sech}\left(\chi_{j\mathrm{R}} + \rho_j/2 + r\right)\Pi_j\mathrm{e}^{\mathrm{i}(\chi_{j\mathrm{I}}+\sigma_j)}, \tag{117}$$

with

$$r = 2\ln\left|\frac{k_1 - k_2}{k_1 + k_2^*}\right|, \tag{118}$$

$$\sigma_1 = 2\arg\left(\frac{k_1 - k_2}{k_1 + k_2^*}\right), \sigma_2 = 2\arg\left(\frac{k_2 - k_1}{k_2 + k_1^*}\right). \tag{119}$$



(a) (b)

Fig. 2. Density plots of $|\phi_0|^2$ (a) and $|\phi_{\pm1}|^2$ (b) for a polar-polar collision. Soliton 1 (left mover) carries only 0 component and soliton 2 (right mover) consists of ±1 components. The parameters used here are $k_1 = 0.25 - 0.25\mathrm{i}$, $k_2 = -0.5 + 0.25\mathrm{i}$, $\alpha_1 = 1/\sqrt{2}$, $\beta_1 = \gamma_1 = 0$, $\alpha_2 = 0$, $\beta_2 = \gamma_2 = 1/\sqrt{2}$.

Equations (115) and (117) are the same form as polar one-soliton solution (105). Collisional effects appear only in the position shift (118) and the phase shifts (119). In Figs. 2, we show the polar-polar collision with $\alpha_1 = 1/\sqrt{2}$, $\beta_1 = \gamma_1 = 0$ and $\alpha_2 = 0$, $\beta_2 = \gamma_2 = 1/\sqrt{2}$. Thus, the partial number N_j, magnetization $\mathbf{F}_j$, momentum P_j, and energy E_j are defined for the asymptotic form of soliton-j and calculated in the same manner as the previous section. The integrals of motion are represented by the sum of those quantities for each soliton. Moreover, we can prove that

$$N_j = 4|k_{j\mathrm{R}}|, \qquad |\mathbf{F}_j| = 0, \qquad P_j = N_j \hbar k_{j\mathrm{I}}, \qquad E_j = N_j c(k_{j\mathrm{I}}^2 - k_{j\mathrm{R}}^2/3), \tag{120}$$

which are by themselves conserved through the collision. In this sense, the polar-polar collision is basically the same as that of the single-component NLS equation.

7.2 Polar-ferromagnetic solitons collision

Under the condition (112), we set soliton 1 to be polar soliton and soliton 2 to be ferromagnetic soliton:

$$|\det\Pi_1| = 1/2, \qquad |\det\Pi_2| = 0. \tag{121}$$

Then, the initial state is represented by eq. (114) with

$$Q_1^{\mathrm{in}} = \sqrt{2}k_{1\mathrm{R}}\mathrm{sech}(\chi_{1\mathrm{R}} + \rho_1/2)\Pi_1 \mathrm{e}^{\mathrm{i}\chi_{1\mathrm{I}}}, \tag{122a}$$

$$Q_2^{\mathrm{in}} = k_{2\mathrm{R}}\mathrm{sech}(\chi_{2\mathrm{R}} + \rho_2/2)\Pi_2 \mathrm{e}^{\mathrm{i}\chi_{2\mathrm{I}}}. \tag{122b}$$

The final state is given by eq. (116) with

$$Q_1^{\mathrm{fin}} = 2k_{1\mathrm{R}} \frac{\tilde{\Pi}_1 \mathrm{e}^{-(\chi_{1\mathrm{R}}+\rho_1/2+\delta)} + \left(\sigma_y \tilde{\Pi}_1^{\dagger} \sigma_y\right) \det\tilde{\Pi}_1 \mathrm{e}^{\chi_{1\mathrm{R}}+\rho_1/2+\delta}}{\mathrm{e}^{-(2\chi_{1\mathrm{R}}+\rho_1+2\delta)} + 1 + \mathrm{e}^{2\chi_{1\mathrm{R}}+\rho_1+2\delta} |\det\tilde{\Pi}_1|^2} \mathrm{e}^{\mathrm{i}\chi_{1\mathrm{I}}}, \tag{123a}$$

$$Q_2^{\mathrm{fin}} = k_{2\mathrm{R}} \operatorname{sech}\left(\chi_{2\mathrm{R}} + \rho_2/2 + r\right) \Pi_2 \mathrm{e}^{\mathrm{i}(\chi_{2\mathrm{I}}+\sigma_2)}. \tag{123b}$$

Here we have defined

$$\mathrm{e}^{2\delta} = \left| \frac{k_1 - k_2}{k_1 + k_2^*} \right|^2 \left\{ 1 + \frac{(k_1 + k_1^*)^2 (k_2 + k_2^*)^2}{|k_1 + k_2^*|^2 |k_1 - k_2|^2} \left| \mathrm{tr}\left(\Pi_1 \Pi_2^{\dagger}\right) \right|^2 \right\}, \tag{124}$$

$$\tilde{\Pi}_1 = \mathrm{e}^{-\delta} \left\{ \Pi_1 - \frac{k_2 + k_2^*}{k_1 + k_2^*} \left(\Pi_1 \Pi_2^{\dagger} \Pi_2 + \Pi_2 \Pi_2^{\dagger} \Pi_1 \right) + \left(\frac{k_2 + k_2^*}{k_1 + k_2^*} \right)^2 \mathrm{tr}\left(\Pi_1 \Pi_2^{\dagger}\right) \Pi_2 \right\}, \tag{125}$$

and also used eqs. (118), (119). Normalization of the new polarization matrix (125) turns out to be unity,

$$||\tilde{\Pi}_1||_2 = \sqrt{2|\tilde{\alpha}_1|^2 + |\tilde{\beta}_1|^2 + |\tilde{\gamma}_1|^2} = 1. \tag{126}$$

The determinant of it becomes

$$\det\tilde{\Pi}_1 = \mathrm{e}^{-2\delta} \left(\frac{k_1 - k_2}{k_1 + k_2^*} \right)^2 \det\Pi_1. \tag{127}$$

We can see clearly that the initial polar soliton breaks into a split type, $\tilde{A}_1 \equiv (2|\det\tilde{\Pi}_1|)^{-1} > 1$, after the collision with a ferromagnetic one. Only when $\left|\mathrm{tr}\left(\Pi_1 \Pi_2^{\dagger}\right)\right| = 0$, where the spinor part of wave function of two initial solitons is orthogonal to each other, we have $\tilde{A}_1 = 1$. Then, eqs. (123) are reduced to

$$Q_1^{\mathrm{fin}} = \sqrt{2} k_{1\mathrm{R}} \operatorname{sech}\left(\chi_{1\mathrm{R}} + \rho_1/2 + r\right) \Pi_1 \mathrm{e}^{\mathrm{i}(\chi_{1\mathrm{I}}+\sigma_1)}, \tag{128a}$$

$$Q_2^{\mathrm{fin}} = k_{2\mathrm{R}} \operatorname{sech}\left(\chi_{2\mathrm{R}} + \rho_2/2 + r\right) \Pi_2 \mathrm{e}^{\mathrm{i}(\chi_{2\mathrm{I}}+\sigma_2)}. \tag{128b}$$

which means that the polar soliton keeps its shape against the collision and shows no mixing among the internal states except for the total phase shift. On the other hand, because of the total spin conservation, the ferromagnetic soliton always retains its polarization matrix and shows only the position and phase shifts similar to those of the polar-polar case.
In Fig. 3, we have density plots of a polar-ferromagnetic collision with the parameters shown in the caption. These pictures correspond to each component of the exact two-soliton solution for one collisional run. For simplicity, we choose the parameters to have $|\phi_1| = |\phi_{-1}|$. The polar soliton (soliton 1) initially prepared in $m_F = \pm 1$ are switched into a soliton with a large population in $m_F = 0$ and the remnant of $m_F = \pm 1$ after the collision. Through the collision, the ferromagnetic soliton (soliton 2) plays only a switcher, showing no mixing in the internal state of itself outside the collisional region, as clearly seen in eq. (123b). In general, this kind of a drastic internal shift of polar soliton is likely observed for large values of $\left|\mathrm{tr}\left(\Pi_1 \Pi_2^{\dagger}\right)\right|$ which appears in eqs. (124), (125). Although all the conserved quantities such as the number of particles and the averaged spin of individual solitons are

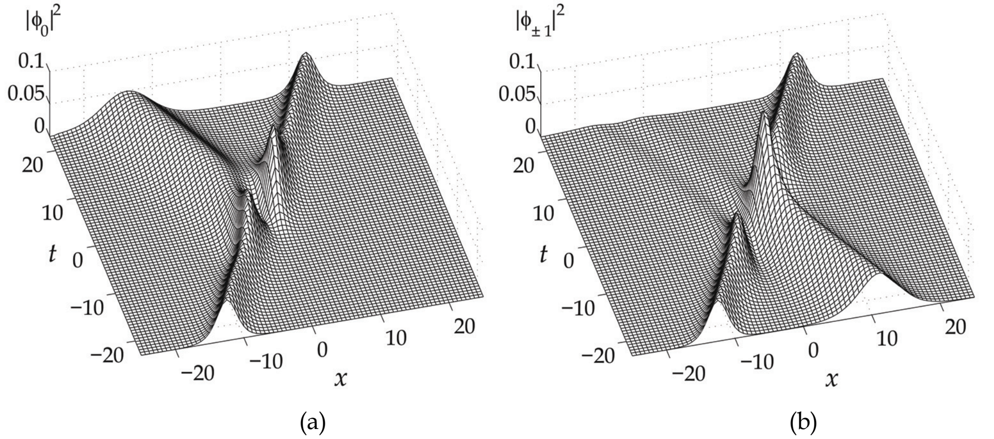


Fig. 3. Density plots of $|\phi_0|^2$ (a) and $|\phi_{\pm1}|^2$ (b) for a polar-ferromagnetic collision. Soliton 1 (left mover) is a polar soliton and soliton 2 (right mover) is a ferromagnetic soliton. The parameters used here are $k_1 = 0.25 - 0.25\mathrm{i}$, $k_2 = -0.5 + 0.25\mathrm{i}$, $\alpha_1 = 0$, $\beta_1 = \gamma_1 = 1/\sqrt{2}$, $\alpha_2 = \beta_2 = \gamma_2 = 1/2$.

invariant during this type of collision, the fraction of each component can vary not only in each soliton level but also in the total after the collision. This contrasts to an intensity coupled multicomponent NLS equation in which the total distribution among all components is invariant throughout soliton collisions while a switching phenomenon similar to Fig. 3 can be observed (Radhakrishnan et al.; 1997).

7.3 Ferromagnetic-ferromagnetic solitons collision

Finally, we discuss the collision between two ferromagnetic solitons,

$$\det\Pi_1 = \det\Pi_2 = 0. \tag{129}$$

The asymptotic forms are obtained for the initial state, $Q \simeq Q_1^{\mathrm{in}} + Q_2^{\mathrm{in}}$ where

$$Q_j^{\mathrm{in}} = k_{j\mathrm{R}}\mathrm{sech}(\chi_{j\mathrm{R}} + \rho_j/2)\Pi_j \mathrm{e}^{\mathrm{i}\chi_{j\mathrm{I}}}. \tag{130}$$

and for the final state, $Q \simeq Q_1^{\mathrm{fin}} + Q_2^{\mathrm{fin}}$ where

$$Q_j^{\mathrm{fin}} = k_{j\mathrm{R}}\mathrm{sech}\left(\chi_{j\mathrm{R}} + \rho_j/2 + s\right)\tilde{\Pi}_j \mathrm{e}^{\mathrm{i}\chi_{j\mathrm{I}}}. \tag{131a}$$

Here we have defined

$$s = \ln\left\{1 - \frac{(k_1 + k_1^*)(k_2 + k_2^*)}{|k_1 + k_2^*|^2}\left|\mathrm{tr}(\Pi_1\Pi_2^\dagger)\right|\right\}, \tag{132}$$

and, for (j, l) = (1,2) or (2,1),

$$\tilde{\Pi}_j = \mathrm{e}^{-s}\left\{\Pi_j - \frac{k_l + k_l^*}{k_j + k_l^*}\left(\Pi_j\Pi_l^\dagger\Pi_l + \Pi_l\Pi_l^\dagger\Pi_j\right) + \left(\frac{k_l + k_l^*}{k_j + k_l^*}\right)^2 \mathrm{tr}\left(\Pi_j\Pi_l^\dagger\right)\Pi_l\right\}, \tag{133}$$

which are shown to be normalized in unity,

$$||\tilde{\Pi}_j||_2 = \sqrt{2|\tilde{\alpha}_j|^2 + |\tilde{\beta}_j|^2 + |\tilde{\gamma}_j|^2} = 1. \tag{134}$$

Each polarization matrix Π_j of a ferromagnetic soliton can be expressed by three real variables τ_j, θ_j, φ_j as

$$\Pi_j = \mathrm{e}^{\mathrm{i}\tau_j} \begin{pmatrix} \cos^2\frac{\theta_j}{2}\mathrm{e}^{-\mathrm{i}\varphi_j} & \cos\frac{\theta_j}{2}\sin\frac{\theta_j}{2} \\ \cos\frac{\theta_j}{2}\sin\frac{\theta_j}{2} & \sin^2\frac{\theta_j}{2}\mathrm{e}^{\mathrm{i}\varphi_j} \end{pmatrix}. \tag{135}$$

In this expression, the polarization matrices in the initial state Π_j and in the final state $\tilde{\Pi}_j$ are given by

$$\Pi_j = \mathrm{e}^{\mathrm{i}\tau_j}\mathbf{u}_j \cdot \mathbf{u}_j^T, \qquad \tilde{\Pi}_j = \mathrm{e}^{-s+\mathrm{i}\tau_j}\tilde{\mathbf{u}}_j \cdot \tilde{\mathbf{u}}_j^T, \tag{136}$$

where, with (j, l) = (1,2), (2,1),

$$\mathbf{u}_j = \left(\cos\frac{\theta_j}{2}\mathrm{e}^{-\mathrm{i}\frac{\varphi_j}{2}}, \sin\frac{\theta_j}{2}\mathrm{e}^{\mathrm{i}\frac{\varphi_j}{2}}\right)^T, \qquad \tilde{\mathbf{u}}_j = \mathbf{u}_j - \frac{k_l + k_l^*}{k_j + k_l^*}\left(\mathbf{u}_l^\dagger \cdot \mathbf{u}_j\right)\mathbf{u}_l. \tag{137}$$

This defines the collision property for the ferromagnetic-ferromagnetic soliton collision.

We can gain a better understanding of the collision between two ferromagnetic solitons by recasting it in terms of the spin dynamics. The total spin conservation restricts the motion of the spin of each soliton on a circumference around the total spin axis [Fig. 4(a)]. It will be interpreted as a spin precession around the total magnetization.

We calculate the magnetization for each soliton to investigate their collision. In the initial state, following eq. (99), we have the spin of soliton-j as

$$\mathbf{F}_j = 2|k_{j\mathrm{R}}|\left(\sin\theta_j\cos\varphi_j, \sin\theta_j\sin\varphi_j, \cos\theta_j\right)^T. \tag{138}$$

Thanks to the scattering property (137), the final state spins can be obtained through $\mathbf{F}_{1,2}$ by

$$\tilde{\mathbf{F}}_j = \mathrm{e}^{-s}\left[\left\{1 - \frac{2k_{l\mathrm{R}}(k_{1\mathrm{R}} + k_{2\mathrm{R}})}{|k_1 + k_2^*|^2}\right\}\mathbf{F}_j + \frac{k_{2\mathrm{I}} - k_{1\mathrm{I}}}{|k_1 + k_2^*|^2}(\mathbf{F}_j \times \mathbf{F}_l) + \left\{\mathrm{e}^s - 1 + \frac{2k_{j\mathrm{R}}(k_{1\mathrm{R}} + k_{2\mathrm{R}})}{|k_1 + k_2^*|^2}\right\}\mathbf{F}_l\right], \tag{139}$$

where

$$\mathrm{e}^s = 1 + \frac{|\mathbf{F}_\mathrm{T}|^2 - (|\mathbf{F}_1| - |\mathbf{F}_2|)^2}{4|k_1 + k_2^*|^2}. \tag{140}$$

The conserved total spin, $\mathbf{F}_\mathrm{T} \equiv \mathbf{F}_1 + \mathbf{F}_2 = \tilde{\mathbf{F}}_1 + \tilde{\mathbf{F}}_2$, is given by

$$\mathbf{F}_\mathrm{T} = \begin{pmatrix} 2|k_{1\mathrm{R}}|\sin\theta_1\cos\varphi_1 + 2|k_{2\mathrm{R}}|\sin\theta_2\cos\varphi_2 \\ 2|k_{1\mathrm{R}}|\sin\theta_1\sin\varphi_1 + 2|k_{2\mathrm{R}}|\sin\theta_2\sin\varphi_2 \\ 2|k_{1\mathrm{R}}|\cos\theta_1 + 2|k_{2\mathrm{R}}|\cos\theta_2 \end{pmatrix}. \tag{141}$$

Considering spin rotation around the total spin $\mathbf{F}_\mathrm{T}$, we can find 'rotated spin' as

$$\mathbf{F}_j^{\mathrm{rot}} = f^{-1}(\varphi)h^{-1}(\theta)f(\omega)h(\theta)f(\varphi)\mathbf{F}_j, \tag{142}$$

where

$$f(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad h(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}, \tag{143}$$

with

$$\varphi = \arctan\frac{F_\mathrm{T}^y}{F_\mathrm{T}^x}, \qquad \theta = \arccos\frac{F_\mathrm{T}^z}{|\mathbf{F}_\mathrm{T}|}. \tag{144}$$

The rotation angle ω is determined by setting $\mathbf{F}_1^{\mathrm{rot}} = \tilde{\mathbf{F}}_1$ through eqs. (139) and (142),

$$\cos\omega = \frac{4(k_{1\mathrm{I}} - k_{2\mathrm{I}})^2 - |\mathbf{F}_\mathrm{T}|^2}{4(k_{1\mathrm{I}} - k_{2\mathrm{I}})^2 + |\mathbf{F}_\mathrm{T}|^2}, \qquad \sin\omega = \frac{4(k_{2\mathrm{I}} - k_{1\mathrm{I}})|\mathbf{F}_\mathrm{T}|}{4(k_{1\mathrm{I}} - k_{2\mathrm{I}})^2 + |\mathbf{F}_\mathrm{T}|^2}. \tag{145}$$

For the case that the magnitudes of the amplitude and velocity for each ferromagnetic soliton are, respectively, identical with each other, $|k_{1\mathrm{R}}| = |k_{2\mathrm{R}}| \equiv N_\mathrm{T}/4$, $|k_{1\mathrm{I}}| = |k_{2\mathrm{I}}| \equiv k_\mathrm{I}$, the final state magnetizations (139) are given by

$$\tilde{\mathbf{F}}_j = \cos^2\frac{\omega}{2}\mathbf{F}_j + \sin\omega\frac{(\mathbf{F}_j \times \mathbf{F}_l)}{|\mathbf{F}_\mathrm{T}|} + \sin^2\frac{\omega}{2}\mathbf{F}_l, \tag{146}$$

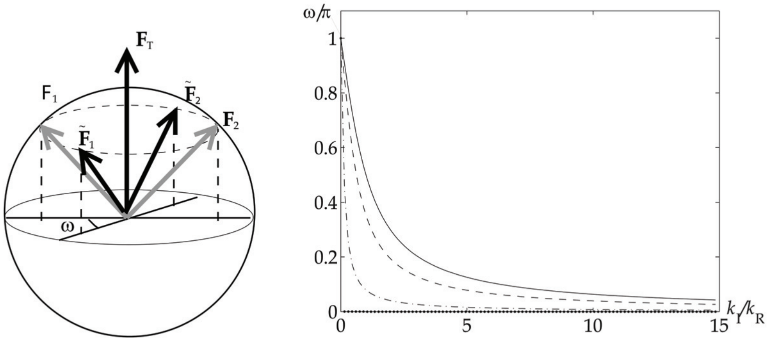
where $(j, l) = (1,2), (2,1)$. The rotation angle ω depends only on the ratio $k_\mathrm{I}/k_\mathrm{R}$ and the magnitude of the normalized total magnetization, $\mathcal{F} \equiv |\mathbf{F}_\mathrm{T}|/N_\mathrm{T}$, as

$$\omega = 2\arccos\left(\left[1 + \left(\frac{k_\mathrm{R}}{k_\mathrm{I}}\right)^2 \mathcal{F}^2\right]^{-1/2}\right). \tag{147}$$

The principal value should be taken for the arccosine function: $0 \le \arccos x \le \pi$.

Setting $k_\mathrm{I} \gg k_\mathrm{R}$ in eq. (147), one gets the small rotation angle, $\omega \simeq 0$. In the opposite case, $k_\mathrm{I} \ll k_\mathrm{R}$, each spin of two colliding solitons almost reverses its orientation, $\omega \simeq \pi$. Recall that k_I is the speed of soliton. We can understand these phenomena since a slower soliton spends the longer time inside the collisional region. Figure 4 shows the velocity dependence of the rotation angle for various initial normalized spins. When $\mathcal{F} = 1$, which corresponds to the case of antiparallel spin collision, the spin precession can not occur as shown by the dotted line in Fig. 4(b).

In Fig. 5–Fig. 7, we give examples of this type of collisions for different k_I, with the other conditions fixed, to illustrate the velocity dependence. The initial normalized spin for the parameter set given in the captions is $\mathcal{F} = 0.5$. The rotation angles are $\omega \simeq 0.2\pi$, 0.5π and 0.9π for Fig. 5, Fig. 6 and Fig. 7, respectively. The internal shift $\phi_1 \rightarrow \phi_{-1}$, and vice versa, gradually increase by slowing down the velocity of the solitons.



(a) (b)

Fig. 4. (a) Schematic of spin precession of two colliding ferromagnetic solitons. (b) Velocity dependence of the rotational angle in spin precession for the different initial relative angles, $\mathcal{F}$ = 1 (solid line), 0.5 (dashed line), 0.0157π (dash-dot line) and 0 (dotted line).

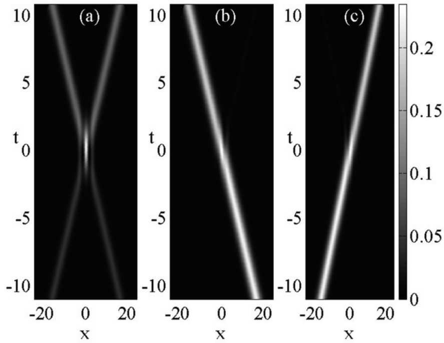


Fig. 5. Density plots of (a) $|\phi_0|^2$, (b) $|\phi_1|^2$ and (c) $|\phi_{-1}|^2$ for a fast ferromagnetic-ferromagnetic collision. The parameters used here are k_1 = 0.5 – 0.75i, k_2 = –0.5 + 0.75i, α_1 = 4/17, β_1 = 16/17, γ_1 = 1/17, α_2 = 4/17, β_2 = 1/17, γ_2 = 16/17.

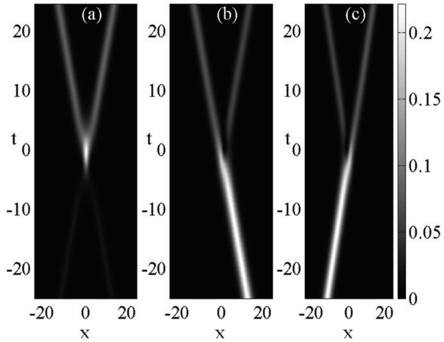


Fig. 6. Density plots of (a) $|\phi_0|^2$, (b) $|\phi_1|^2$ and (c) $|\phi_{-1}|^2$ for a medium speed ferromagnetic-ferromagnetic collision. The parameters are the same as those of Fig. 5 except for k_{1I} = –0.25, k_{2I} = 0.25.

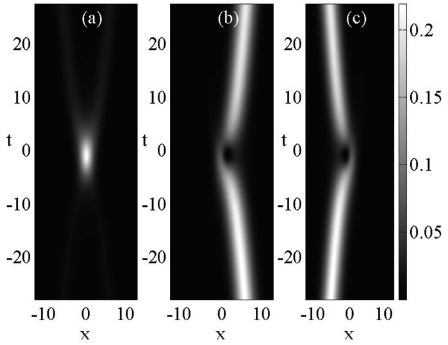


Fig. 7. Density plots of (a) $|\phi_0|^2$, (b) $|\phi_1|^2$ and (c) $|\phi_{-1}|^2$ for a slow ferromagnetic-ferromagnetic collision. The parameters are the same as those of Fig. 5 except for $k_{1I} = -0.05$, $k_{2I} = 0.05$.

8. Concluding remarks

The soliton properties in spinor Bose–Einstein condensates have been investigated. Considering two experimental achievements in atomic condensates, the matter-wave soliton and the spinor condensate, at the same time, we have predicted some new phenomena.

Based on the results provided in Sec. 2–4, in Sec. 5 we have introduced the new integrable model which describes the dynamics of the multicomponent matter-wave soliton. The key idea is finding the integrable condition of the original coupled nonlinear equations, i.e., the spinor GP equations derived in Sec. 4. The integrable condition expressed by the coupling constants, which is accessible via the confinement induced resonance explained in Sec. 3.

In Sec. 6, we classify the one-soliton solution. There exist two distinct spin states: ferromagnetic, $|\mathbf{F}_T| = N_T$ and polar, $|\mathbf{F}_T| = 0$. In the ferromagnetic state, the spatial part and the spinor part of the soliton are factorized (ferromagnetic soliton). In the polar state, dissimilar shaped solitons which we call polar soliton for $\mathbf{f}(x) = 0$ and split soliton otherwise are energetically degenerate. The polar soliton has one peak and the space-spinor factorization holds. On the other hand, a split soliton consists of twin peaks and the three components show different profiles. Changing the polarization parameters one may control the peak distance continuously.

In Sec. 7, we have analyzed two-soliton solutions which rule collisional phenomena of the multiple solitons. Specifying the initial conditions, we have demonstrated two-soliton collisions in three characteristic cases: polar-polar, polar-ferromagnetic, ferromagnetic-ferromagnetic. In their collisions, the polar soliton is always "passive" which means that it does not rotate its partner's polarization while the ferromagnetic soliton does. Thus, in the polar-ferromagnetic collision, one can use the polar soliton as a signal and ferromagnetic soliton as a switch to realize a coherent matter-wave switching device. Collision of two ferromagnetic solitons can be interpreted as the spin precession around the total spin. The rotation angle depends on the total spin, amplitude and velocity of the solitons. Only varying the velocity induces drastic change of the population shifts among the components.

Stability of spinor solitons has been investigated numerically and perturbatively (Li et al.; 2005) (Dabrowska-Wüster et al.; 2007) (Doktorov et al.; 2008). It is also interesting to pursue the soliton dynamics of spinor condensates under longitudinal harmonic trap (Zhang et al.; 2007). Recently, the integrability of the spinor GP equation has been studied in detail (Gerdjikov et al.; 2009). The behavior of spinor solitons shows a variety of nonlinear dynamics and it is worth exploring them experimentally.

9. Acknowledgment

This work was supported by Grant-in-Aid for Scientific Research No. 20740182 from MEXT.

10. References

Dabrowska-Wüster, B. J.; Ostrovskaya, E. A.; Alexander, T. J.; Kivshar, Y. S. (2007). Multicomponent gap solitons in spinor Bose–Einstein condensates. *Physical Review A*, 75, 2, (April 2008) 023617-1–023617-11, ISSN 1094-1622.

Doktorov, E. V.; Wang, J. D.; Yang, J. K. (2008). Perturbation theory for bright spinor Bose-Einstein condensate solitons. *Physical Review A*, 77, 4, (April 2008) 043617-1–043617-11, ISSN 1094-1622.

Gerdjikov, V.S.; Kostov, N.A.; Valchev, T. I. (2009). Solutions of multi-component NLS models and Spinor Bose–Einstein condensates. *Physica D: Nonlinear Phenomena*, 238, 15, (July 2009) 1306–1310, ISSN 0167-2789.

Ieda, J.; Tsurumi, T.; Wadati, M. (2001). Bose–Einstein Condensation of Ideal Bose Gases. *Journal of the Physical Society of Japan*, 70, 5, (May 2001) 1256–1259, ISSN 1347-4073.

Ieda, J.; Miyakawa, T.; Wadati, M. (2004). Exact Analysis of Soliton Dynamics in Spinor Bose–Einstein Condensates. *Physical Review Letters*, 93, 19, (November 2004) 194102-1–194102-4, ISSN 1079-7114.

Ieda, J.; Miyakawa, T.; Wadati, M. (2004). Matter-Wave Solitons in an F=1 Spinor Bose–Einstein Condensate. *Journal of the Physical Society of Japan*, 73, 11, (November 2004) 2996–3007, ISSN 1347-4073.

Ieda, J.; Uchiyama, M.; Wadati, M. (2007). Inverse scattering method for square matrix nonlinear Schrodinger equation under nonvanishing boundary conditions. *Journal of Mathematical Physics*, 48, 1, (January 2007) 013507-1–013507-19, ISSN 1089-7658.

Khaykovich, L.; Schreck, F.; Ferrari, G.; Bourdel, T.; Cubizolles, J.; Carr, L. D.; Castin, Y.; Salomon, C. (2002). Formation of a Matter-Wave Bright Soliton. *Science*, 96, 5571, (May 2002) 1290–1293, ISSN 1095-9203.

Kanamoto, R; Saito, H; Ueda, M. (2003). Quantum phase transition in one-dimensional Bose– Einstein condensates with attractive interactions. *Physical Review A*, 67, 1, (January 2003) 013608-1–013608-7, ISSN 1094-1622.

Li, L.; Li, Z.; Malomed, B. A.; Mihalache, D.; Liu, W. M. (2005). Exact soliton solutions and nonlinear modulation instability in spinor Bose–Einstein condensates. *Physical Review A*, 72, 3, (September 2005) 033611-1–033611-11, ISSN 1094-1622.

Meystre, P. (2001). *Atom Optics*, Springer-Verlag New York, Inc., New York.

Olshanii, M. (1998). Atomic Scattering in the Presence of an External Confinement and a Gas of Impenetrable Bosons. *Physical Review Letters*, 81, 5, (August 1998) 938–941, ISSN 1079-7114.

Pethick, C. J. & Smith, H. (2002). *Bose–Einstein condensation in dilute Gases*, Cambridge University Press, Cambridge. Also 2nd ed. (2008).

Stenger, J.; Inoue, S.; Stamper-Kurn, D. M.; Miesner, M. R.; Chikkatur, A. P.; Ketterle, W. (1998). Spin domains in ground-state Bose–Einstein condenstate. *Nature*, 396, 6709, (November 1998) 345–348, ISSN 0028-0836.

Radhakrishnan, R.; Lakshmanan, M.; Hietarinta, J. (1997). Inelastic collision and switching of coupled bright solitons in optical fibers. *Physical Review E*, 56, 2, (August 1997) 2213–2216, ISSN 1550-2376.

Strecker, K. E.; Partridge, G. B.; Truscott, A. G.; Hulet, R. G. (2002). Formation and propagation of matter-wave soliton trains. *Nature*, 417, (May 2002) 150–153, ISSN 0028-0836.

Tsuchida, T. & Wadati, M. (1998). The Coupled Modified Korteweg–de Vries Equations. *Journal of the Physical Society of Japan*, 67, 4, (April 1998) 1175–1187, ISSN 1347-4073.

Uchiyama, M.; Ieda, J.; Wadati, M. (2006). Dark Solitons in F=1 Spinor Bose–Einstein Condensate. *Journal of the Physical Society of Japan*, 75, 6, (June 2006), 064002-1–064002-9, ISSN 1347-4073.

Uchiyama, M.; Ieda, J.; Wadati, M. (2007). Multicomponent Bright Solitons in F=2 Spinor Bose– Einstein Condensates. *Journal of the Physical Society of Japan*, 76, 7, (July 2007), 074005- 1–074005-6, ISSN 1347-4073.

Ueda, M. & Koashi, M. (2002). Theory of spin-2 Bose–Einstein condensates: Spin correlations, magnetic response, and excitation spectra. *Physical Review A*, 65, 6, (May 2002), 063602-1–063602-22, ISSN 1094-1622.

Zhang, W.; Müstecaplıoğlu, Ö. E.; You, L. (2007). Solitons in a trapped spin-1 atomic condensate. *Physical Review A*, 75, 4, (April 2007), 043601-1–043601-8, ISSN 1094-1622.

A Conceptual Model for the Nonlinear Dynamics of Edge-localized Modes in Tokamak Plasmas

Todd E. Evans[1], Andreas Wingen[2],
Jon G. Watkins[3] and Karl Heinz Spatschek[2]
[1]*General Atomics,*
[2]*Institute for Theoretical Physics, Heinrich-Heine-University,*
[3]*Sandia National Laboratory,*
[1,3]*United States*
[2]*Germany*

1. Introduction

High performance magnetically confined toroidal plasmas, such as those required for the operation of a tokamak based fusion power plant, suffer from a troubling type of repetitive edge instability known as edge-localized modes (ELMs). Magnetohydrodynamic (MHD), peeling-ballooning, theory predicts that these instabilities are driven by a large current density and pressure gradient that forms at the plasma edge as a consequence of the enhanced confinement levels achieved in high performance H-mode plasmas. Although ELMs are a common feature of high confinement tokamak plasmas, there are significant gaps in our understanding of how these instabilities scale with the geometry of the plasma and operating conditions expected in large tokamaks that are required for the generation of fusion power. Thus, there is an urgent need for a model that can be validated with experimental data from existing smaller tokamaks.

Here, we present a conceptual model describing the topological evolution of the magnetic separatrix, in a tokamak plasma with a dominate lower hyperbolic point. Subsequently, the nonlinear dynamics of the ELM instability, prescribed by the evolving separatrix topology, is discussed. The model invokes a feedback amplification mechanism that causes the stable and unstable invariant manifolds of the separatrix, comprising a "homoclinic tangle" (Guckenheimer & Holmes, 1983), to grow explosively as the topology of the separatrix manifolds unfold. The amplification process is driven by the rapid growth of helical, field-aligned, thermoelectric currents that flow through relatively short edge plasma flux tubes connecting high heat flux wall structures, known as divertor target plates, on both sides of the plasma. These thermoelectric currents produce magnetic fields that couple to the separatrix and modify its 3D (topological) structure. As the lobes of the separatrix tangle grow, their area of intersection with the divertor target plates increases along with the size of the flux tubes connecting target plates on both sides of the plasma. This increases the thermoelectric current flow and completes the feedback loop. Numerical simulations have shown that our model is consistent with measurements of the currents flowing between the target plates and with camera images of the heat flux patterns on the divertor target plates

(Wingen, et al. 2009a). In addition, this model suggests that nonaxisymmetric external magnetic coils can be used to force higher order separatrix bifurcations to prevent ELMs.

In the following sections, we discuss why ELMs are an important problem in tokamaks, review what is known experimentally and theoretically about the characteristics of ELMs, present a conceptual framework for the nonlinear evolution of an ELM and discuss results from numerical simulations of the proposed model. We show a sequence of topological bifurcations that are observed in the numerical simulations and discuss how these result in a separatrix topology that produces heat flux patterns which are consistent with those measured by infrared cameras in the DIII-D (Luxon, 2002) tokamak. A general description of tokamaks and tokamak physics is given by Callen et al., (1992), Evans (2008) and Wesson (2004).

2. Properties of ELMs in high performance tokamak plasmas

2.1 ELM dynamics

Type-I ELMs are naturally occurring MHD instabilities that release large bursts of particles and energy from the boundary of the plasma (Suttrop 2000). These very fast growing instabilities share properties that are somewhat similar to the eruption of prominences and flares from the solar photosphere (Evans, et al., 1996). More specifically, expanding hot plasma filaments carrying energy, particles and momentum away from the confined plasma volume into the surrounding space are associated with these complex dynamical plasma events that form on the surface of the sun and at the edge of a tokamak discharge. Tokamaks operating in high confinement H-modes, with strong edge transport barriers, rely exclusively on the formation of a large pressure gradient near the surface of the plasma to obtain sufficiently high central temperatures and densities to carry out fusion energy experiments in these devices. These large pressure gradients are believed to drive edge MHD instabilities, referred to as peeling-ballooning modes, that are responsible for the onset of ELMs (Snyder, et al., 2005). Since ELMs periodically release particles and energy from the edge of the plasma, they limit the size of the pressure gradient that can be obtained in tokamaks (Fenstermacher, et al. 2003). Scaling studies suggest that this limits the maximum temperature of the core plasma and thus the ultimate performance of the tokamak (Loarte, et al. 2003). In addition, the impulsive energy and particle flux released by ELMs can cause a significant enhancement in the erosion of solid surfaces that make up the internal walls and divertor components of the tokamak. The impulsive loading of these structures due to ELMs releases non-hydrogenic impurities as a result of enhanced solid surface erosion. These impurities change the properties of the divertor plasma and can be transported out of the divertor chamber into the region of the scrape-off layer (SOL) plasma located between the separatrix in the main chamber walls of the tokamak. While ELMs tend to prevent the eroded divertor impurities from penetrating deeply into the high temperature region of the core plasma, located inside the steep pressure gradient region referred to as pedestal plasma, these impurities can accumulate in steady-state discharges and affect the plasma performance unless the tokamak pumping system is capable of removing this additional particle flux.

ELMs are typically classified by the amount of energy they eject from the pedestal plasma and their dynamical properties. The largest of these instabilities, referred to as type-I ELMs, are capable of reducing the energy stored in the pedestal plasma by as much as 20-25% in tokamaks operating at the highest performance levels (Loarte, et al. 2003). In the largest tokamaks operating at the present time, this amounts to the ejection of up to 1 MJ of energy within a period of about 200-300 μs. In the next generation of tokamaks that are under construction or being planned, the energy ejected by a single ELM is expected to increase by about a factor of 20. These type-I ELMs are also characterized by an increase in frequency f_{ELM} as the injected power level of the neutral beam heating system P_{NBI} increases and they do not tend to have any clearly identifiable coherent magnetic fluctuations prior to their onset (*i.e.*, magnetic precursors) although an increase in the level of broadband plasma turbulence is sometimes observed prior to their onset. Profiles of the edge electron density (n_e) and temperature (T_e) just before an ELM are shown for a typical DIII-D type-I ELMing discharge in Fig. 1(a) while Fig. 1(b) shows how the n_e profile changes immediately following a type-I ELM.

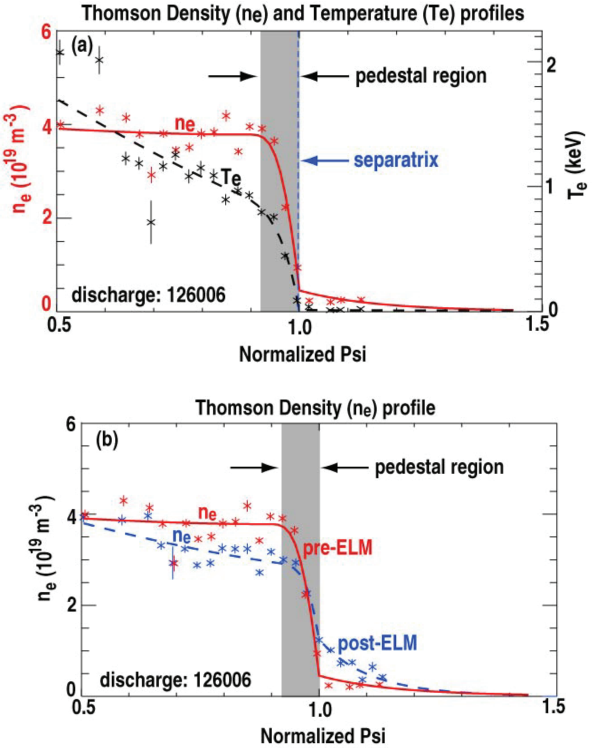


Fig. 1. (a) An example of the steep n_e and T_e profiles, as a function of the normalized poloidal magnetic flux Psi (ψ_N), across the outer region of the plasma inside the separatrix and outside the separatrix in a region referred to as the SOL and (b) the n_e profile before and after an ELM in DIII-D discharge 126006.

As seen in Fig. 1(b), plasma density from the top of the pedestal region inward to approximately 1/2 the radius of the core plasma, at a normalized poloidal magnetic flux Psi (ψ_N) equal 0.5, is ejected into the region outside the separatrix, referred to as the SOL, during the explosive growth period of the ELM instability in a typical high performance, low collisionality, DIII-D type-I ELMing discharge. Type-I ELMs typically have frequencies f_{ELM}

= 10-200 Hz and are triggered when the pedestal pressure gradient ($\alpha_{\rm ped} \propto \nabla p_{\rm ped}$) approaches the critical ideal MHD ballooning limit $\alpha_{\rm ped} \sim 2\text{–}3\ \alpha_{\rm crit\_ball}$ (Osborne et al., 2000). They are often characterized by an isolated, very rapid, increase in the deuterium recycling emissions when the particles ejected from inside the separatrix arrive at the divertor target plates or the walls of the tokamak as shown in Fig. 2.

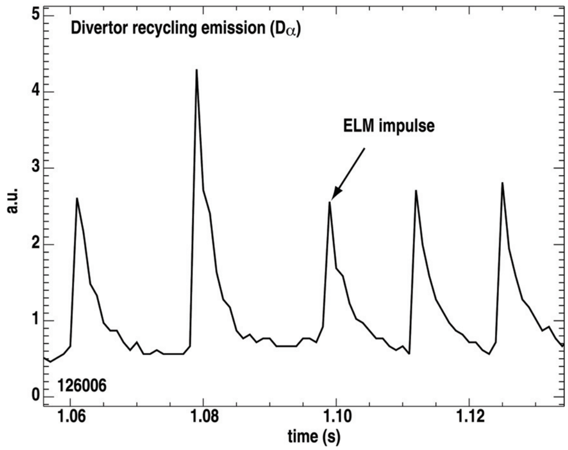


Fig. 2. A series of type-I ELM impulses seen in the lower (primary) divertor deuterium (D_α) recycling emissions during DIII-D discharge 126006 where $f_{\rm ELM}$ = 50→75 Hz is correlated to an increase in the stored energy of the plasma.

A significant fraction of the energy released from the pedestal during type-I ELMs strikes the divertor target plates [see Fig. 5(b) for a view of the DIII-D lower divertor and divertor target plates] along with the particle flux responsible for the spikes in the recycling emissions (Fig. 2). It is this combined, highly impulsive, heat and particle flux that can cause enhanced erosion of the divertor targets and walls in large, high performance, tokamaks leading to a substantial increase of non-hydrogenic impurities released into the divertor and SOL plasmas.

Also associated with these impulsive heat and particle fluxes are large, rapidly growing, electric currents that are an intrinsic part of type-I ELM dynamics. These currents are, in fact, a basic element of the nonlinear model described below. In DIII-D Langmuir probes are used to measure the parallel ion saturation current flowing through the divertor target plates at several radial positions. These measurements show that these currents grow explosively to a saturated amplitude exceeding 1 A/mm^2 in 50 μs or less during the nonlinear growth of a type-I ELM. Measurements of the toroidal distribution and dynamics of these currents with a toroidal tile current array in DIII-D has shown that they are strongly non-axisymmetric with dominate toroidal mode numbers consisting primarily of n=1 and 2 components (Evans et al., 1995) while the presence of higher n modes has been observed during ELM precursors in some DIII-D discharges (Osborne et al., 2000). As an example of the dynamics involved in the evolution of this current, the data shown in Fig. 3 demonstrates the explosive growth of the instability followed by a slow decay during a single ELM. This data is obtained with two lower divertor Langmuir probes located just outside the 15 mm SOL flux surface with major radii R = 1.500 m and 1.528 m in a double null (upper and lower hyperbolic point) DIII-D discharge biased upward by 11 mm. The radial structure of this current indicates that the ELM produces a strong interaction with the top of a pump duct more than 120 mm away from the point of intersection of the separatrix with the lower divertor target (Fig. 5 shows a layout of the lower divertor geometry). This data also suggests that current flowing in a type-I ELM may be associated with a relatively ridged structure that forms during its initial explosive growth phase. Finally, it suggests that as the current in this structure decays it appears to rotate past the two Langmuir probes with a rather regular period of $\Delta t \sim 480\ \mu s$ (*i.e.*, a toroidal rotation velocity $v_{ELM} = 2\pi R/\Delta t \sim$ 19.6 km/s where R is the radial position of one of the Langmuir probes) as indicated by a series of fairly regularly spaced peaks shown in Fig. 3. Signatures such as these have also been seen in the DIII-D midplane reciprocating Langmuir probe data where v_{ELM} = 13.5 km/s was observed in plasmas with edge toroidal carbon rotation velocities v_{carbon} = 22 km/s (Boedo et al., 2005; Yu et al., 2008).

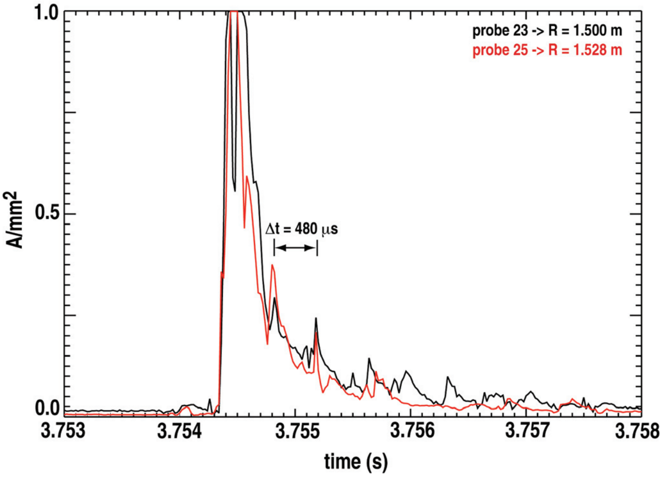


Fig. 3. Time evolution of the plasma current measured by a pair of lower divertor Langmuir probes located 28 mm apart in major radius (R) during an ELM DIII-D discharge 138229.

Type-II ELMs are sometimes observed as the axisymmetric plasma shape becomes more triangular and elongated. They often appear as small, irregular, fluctuations in the D_α data interspersed between the large type-I ELM D_α spikes. Type-II ELMs do not appear to have a distinct P_{NBI} dependence or any signatures associated with coherent MHD precursors and do not seem to be associated with a specific α_{crit} limit (Zohm, 1996). Type-III ELMs are small, relatively high f_{ELM}, instabilities that tend to have lower frequencies as P_{NBI} increases (Osborne et al., 2000). They typically have coherent magnetic precursor modes with frequencies in the 50 kHz range and have low to intermediate toroidal mode numbers (n=5-10). They are often found in relatively high-density, lower P_{NBI}, plasmas with α_{ped} ranging from about 30% to 50% of α_{crit} (Suttrop, 2000). Other types of small ELMs (e.g., type-V) have been identified in spherical tokamaks where they appear only in lower single null plasmas and are often interspersed between large type-I ELMs (Maingi et al., 2005).

The conceptual model proposed here deals exclusively with the nonlinear dynamics and associated topological evolution of the explosive growth seen during the initial growth of type-I ELMs. The dynamics of the different types of ELMs outlined above, as well as those observed during intermittent transport bursts in low confinement modes (L-modes) are, at the most fundamental level, required to conform to the general framework of this model *i.e.*, the fact that a divergence free vector field, such as the equilibrium magnetic field in a tokamak or stellarator, must ultimately be consistent with a (conservative) Hamiltonian representation such as that prescribed by dynamical systems theory (Guckenheimer & Holmes, 1983; Lichtenberg & Lieberman, 1992).

2.2 ELM topology

Before elaborating the details of the nonlinear type-I ELM model below, it is instructive to briefly describe the global topology of these instabilities. Fortunately, spherical tokamaks such as MAST (Kirk et al., 2004; Kirk et al., 2007) and NSTX (Maingi et al., 2005) are equipped with visible light fast framing cameras that can capture images of type-I ELMs. Figure 4 provides a full view of the plasmas captured during a type-I ELM in MAST.

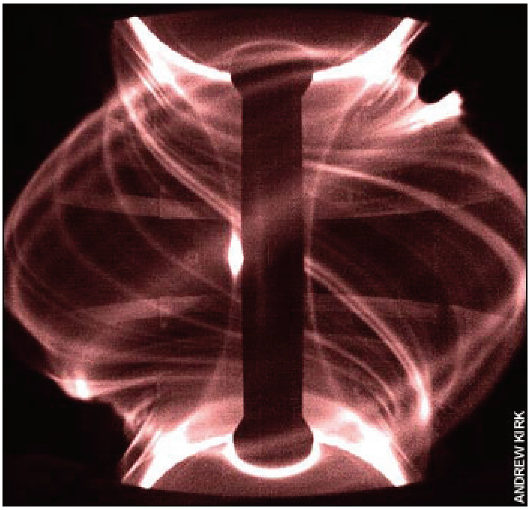


Fig. 4. A wide-angle view of the MAST plasma at one instant in time during the evolution of a type-I ELM (Courtesy of A. Kirk, Culham Laboratory, UK).

Here, the bright emission bands, referred to as ELM filaments, wrap around the outer surface of the plasma in helical patterns that connect the upper and lower divertors. The pitch of these filaments is aligned with the local magnetic field, which typically has a rather steep angle with respect to the equatorial plane of the plasma due to the relative strength of the poloidal field compared to that of the toroidal field in spherical tokamaks such as MAST. Note that the intensity of the emission in these filaments is not uniform along their helical axis and that these structures are seen to protrude from the surface as they approach the upper hyperbolic point where they become much more toroidally aligned. These protrusions are consistent with the type of structure predicted by the topology of homoclinic and heteroclinic tangles invoked in the ELM model presented below. Here, the protrusions correspond to the lobes of the tangle, which become narrower in the poloidal direction and more extended in the radial direction as they approach a hyperbolic point (Fig. 5 shows the lobes calculated when an n=1 homoclinic tangle is found in the DIII-D tokamak). This poloidal compression, accompanied by a radial expansion, is a consequence of the preservation of a constant value of the magnetic flux contained inside each lobe of the structure as prescribed by the Hamiltonian nature of the tangle in the model as it approaches a region of weak poloidal magnetic field near the hyperbolic points. These protruding lobes form a spiraling magnetic footprint that converges to the unperturbed intersection of the separatrix with the divertor target plate similar to the one shown in Fig. 4 of Roeder et al., (2003) for an n=1 homoclinic tangle in DIII-D. These magnetic footprints are essential elements of the nonlinear ELM model presented below.

2.3 ELM theories

Linear ELM theories tend to fall into three general categories. The first and most well developed of these includes ideal and resistive ballooning MHD models. These involve pressure driven modes that couple to external kink modes (sometimes referred to as peeling modes). The second involves dynamics described by a bifurcation of the confinement from an H-mode to an L-mode forming a dynamical state described by a restricted type of limit cycle. The third combines elements taken from the MHD and limit cycle models to construct an appropriate set of dynamics. Each of these models is reviewed in a paper by Conner (1998). Nonlinear ELM models are relatively sparse due to the complex nature of the dynamics and topology involved in this phase of the instability. One example invokes the explosive growth of a narrow finger of hot plasma that pushes its way through other field lines (nonlinear ballooning) from a small region in the plasma interior and spreads across a large section of the surface of the plasma (Cowley et al., 2003). These models are difficult to validate in any practical way with tokamak data due to a lack of specific predictions on how they relate to the various types of ELMs and operating regimes found in high performance tokamak discharges. Clearly, a more quantitative model is needed. Thus, there is strong motivation to develop a model that can be more easily tested with experimental data. The model presented below provides a step in this direction since it can be used to numerically calculate the global topology of ELMs including the distribution and size of magnetic footprints that can be directly compared to divertor diagnostic data (Wingen et al., 2009a).

3. Description of the proposed nonlinear ELM model

3.1 Hamiltonian description of the separatrix topology in poloidally diverted tokamaks

Poloidally diverted tokamaks are formed by a set of external axisymmetric coils that result in poloidal magnetic field nulls when superimposed on the magnetic field due to a toroidal plasma current flow in the discharge. These poloidal field nulls, in combination with magnetic fields from other shaping coils in the tokamak, form hyperbolic points (Zaslavsky, 2005) of the system along with their associated separatrices that divide field line trajectories into trapped (inside the separatrix) versus passing (outside the separatrix) regions of space (Evans, 2008). In an ideal axisymmetric poloidally diverted tokamak the trapped and passing field line regions are referred to as “closed” and “open“ field line regions respectively. This is because field lines outside the separatrix intersect the walls of the tokamak and thus are “open“ with respect to the loss of heat and particles that flow parallel to the field lines. Alternatively, field lines inside the ideal axisymmetric separatrix do not intersect the walls of the tokamak and thus are considered “closed“ in terms of heat and particle transport parallel to these field lines. As discussed below this terminology is no longer applicable when small non-axsymmetric magnetic pertrubations are present in the system.

A fundamental element of the ELM model discussed here is the nonlinear evolution of the separatrix topology in a poloidally diverted tokamak. Here, the growth of a small topological defect known as a homoclinic tangle (Guckenheimer & Holmes, 1983), formed by the separatrix due to intersections of stable and unstable invariant manifolds associated with a hyperbolic point of the system, is the basic dynamical process invoked by the model. In a poloidally diverted tokamak, following along the spatial trajectory of a stable invariant manifold in the forward direction results in a series of converging steps that approaches the hyperbolic point associated with the manifold. Similarly, following the unstable invariant manifold in the opposite (backward) direction produces a series of converging steps toward the hyperbolic point from the opposite side. Thus, the splitting of trajectories into stable and unstable manifolds due to non-axisymmetric perturbations introduces a directional dependence into the spatial trajectories of the field lines implying that following field lines in opposite directions leads to very different spatial locations in the plasma (with the exception of homoclinic points where the stable and unstable invariant manifolds intersect). In general, a homoclinic (self-intersecting) tangle results in a 3D separatrix topology that is a generic property of perturbed hyperbolic conservative systems which are composed of divergence-free vector fields. The dynamics of such a system is described by integrating Hamilton's equations of motion. Theoretically, it is well known that when sufficiently small perturbations are introduced into such a system it remains Hamiltonian in nature and preserves its well-behaved (deterministic) dynamics (Dankowicz, 1997; Lichtenberg & Lieberman, 1992). Such systems are commonly referred to as "near integrable" and generically have non-degenerate, transversely self-intersecting, separatrix manifold topologies that form the lobes of the homoclinic tangle. Separatrix structures such as these have been studied extensively in physics, mathematics, astrophysics, engineering and neuroscience (Dankowicz, 1997; Guckenheimer & Holmes, 1983; Simiu, 2002). Additionally, it is well known from conservative dynamical systems theory that the topology of a homoclinic tangle is the fundamental element that dictates the behavior of the trajectories which form the solutions to the differential equations describing the dynamics of the system (Guckenheimer & Holmes, 1983). In toroidal plasma confinement devices such as stellarators and tokamaks, the 3D topology of the field lines at any instant in time is found by integrating a set of magnetic differential equations that are formulated in terms of the toroidal (χ) and poloidal (ψ) magnetic flux coordinates (D'haeseler et al., 1991). Here, ψ is associated with the Hamiltonian H while χ serves as the canonical momentum of the system and the equations describing the 3D spatial trajectories of the field lines are given in Hamilton–Jacobi form as:

$$\frac{d\theta}{d\phi}=\frac{dH}{d\chi} \tag{1}$$

$$\frac{d\chi}{d\phi}=-\frac{dH}{d\theta} \tag{2}$$

where θ and ϕ are the poloidal and toroidal angles respectively (Evans, 2008). The usual Hamiltonian is recognized in terms of the familiar canonical coordinates p,q by substituting

$\chi \rightarrow p$ and $\theta \rightarrow q$ and associating ϕ with time (t). In a tokamak $2\pi\chi$ is the toroidal magnetic flux enclosed by a surface of constant χ and $2\pi\psi = 2\pi H$ is the poloidal magnetic flux inside a surface of constant H.

Equations (1) and (2) are generally integrable given an axisymmetric plasma equilibrium but the addition of small non-axisymmetric magnetic fields transforms the Hamiltonian into an arbitrary function of the toroidal and poloidal angles. In this system a symmetry breaking, non-axisymmetric, magnetic perturbation can be expressed in terms of a perturbed Hamiltonian $\varepsilon H_1(\chi,\phi,\theta)$ where ε is a small dimensionless perturbation parameter. Then, the total Hamiltonian H is given as:

$$H = H_0(\chi) + \varepsilon H_1(\chi,\phi,\theta) \tag{3}$$

or the sum of the axisymmetric part H_0 and the non-axisymmetric perturbed part H_1. The perturbed part of the Hamiltonian can be expressed in terms of a Fourier series as:

$$H_1(\chi,\phi,\theta) = \sum_{m,n} H_{m,n}(\chi)\cos(m\theta - n\phi + \chi_{m,n}) \tag{4}$$

where m and n are the poloidal and toroidal mode numbers respectively (Abdullaev, 2006).

In realistic tokamaks, the nominally degenerate invariant manifolds that form an ideally axisymmetric separatrix are transformed into an infinite set of homoclinic intersections by small field-errors associated with non-axisymmetric toroidal and poloidal magnetic field coil positions and other random magnetic pertrubations that are an intrinsic part of the tokamak environment (Evans et al., 2005). In addition, externally applied low toroidal mode number (n=1) non-axisymmetric magnetic fields are commonly used to "correct" ambient field-errors that amplify MHD modes in the core plasma.

Figure 5 shows a poloidal projection of the 3D separatrix structure at one toroidal angle in the DIII-D tokamak during the application of an n=1 field-error correction perturbation. As seen in the lower part of Fig. 5(a) just above the divertor region the lobes of the homoclinic tangle intersect the high-field side (HFS) wall (R = 1.02 m, Z = -1.17 m) while on the low-field side (LFS) the lobes intersect the horizontal divertor target plate (R = 1.35 m, Z = -1.36 m). Here, R,Z are cylindrical coordinates representing the distance from the toroidal axis of the tokamak and the displacement from equatorial plane respectively. A magnified view of the lower divertor region is shown in Fig. 5(b) with a 45° divertor tile (dashed line) connecting the HFS wall (R = 1.02 m) to the horizontal target plate tile (Z = -1.37 m). The entrance to the pump duct is shown on the right-hand side of Fig. 5(b) (R ≥ 1.36 m) with the top of the pump duct located at Z = -1.25 m. The connection length L_c of magnetic flux tubes between the LFS divertor target plate and the HFS wall is shown by the color bars in each part of the figure.

This discharge is an example of a double null plasma equilibrium with the balance between the upper and lower hyperbolic points displaced slightly downward. Here, the upper hyperbolic point is located at R = 1.27 m, Z = 1.11 m while the lower hyperbolic point is located at R = 1.28 m, Z = -1.13 m. The topology of the hetroclinic tangles formed in double null equilibria has been shown to be a sensitive function of the relative positions of the upper (secondary) and lower (primary) hyperbolic points (Evans et al., 2004). For a downward biased equilibrium such as that shown in Fig. 5, the homoclinic tangle associated with the lower hyperbolic point dominates the 3D topology of the separatrix and creates a dramatic change in the magnetic topology inside the separatrix. Here, one of the lobes of the tangle intersects the horizontal divertor target plate at R = 1.36 m, Z = -1.36 m while another lobe intersects the vertical wall located at R = 1.02 m. The intersection of these homoclinic lobes with the divertor target plate and wall ″opens″ some of the field lines that were previously in the ″closed″ region inside the separatrix prior to the application of the n=1 magnetic perturbation field from the field-error correction coil (although some field lines are always open due to intrinsic non-axisymmetric field errors without the correction coils). This topological change creates a set of highly complex field line trajectories that traverse the plasma volume inside the separatrix and connect the vertical high-field side (HFS) wall to the low-field side (LFS) horizontal divertor target plate. This topology is similar to that of "line-tying" found in the solar photosphere (Gibons & Spicer, 1981). Additionally, the field line topology formed by this "line-tying" type of bifurcation is composed of a mixture of stochastic fields, with a wide range of connection lengths (L_c) that form fractal distributions (Abdullaev, 2006), embedded inside a set of coherent flux tubes with short connection lengths (Wingen et al., 2009b; Wingen et al., 2009c). It is the short L_c flux tubes that play a fundamental role in the nonlinear ELM model discussed below.

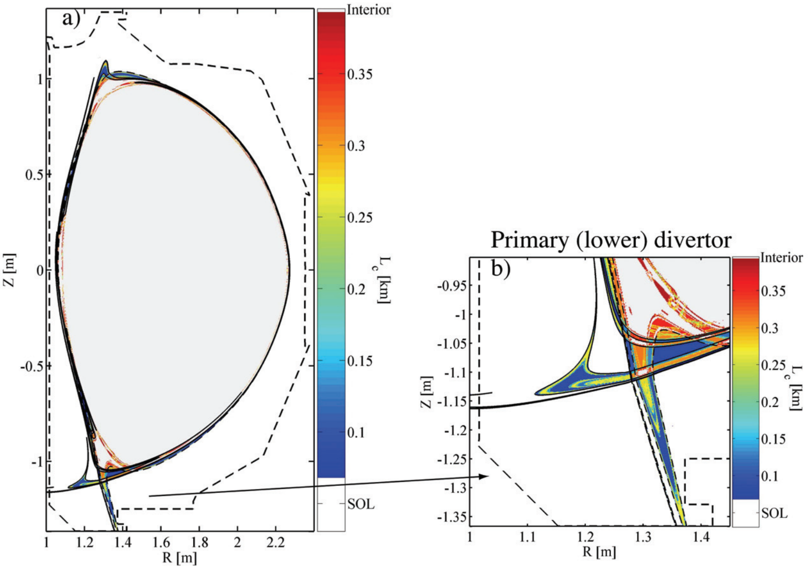


Fig. 5. (a) Full poloidal cross sectional view of a separatrix homoclinic tangle formed by an applied external n=1 magnetic perturbation due to the DIII-D field-error correction coil with a current of 8 kAt in discharge 133908 at t = 2000 ms. (b) An expanded cross sectional view of the primary divertor. L_c is the field line connection length between the horizontal target plate (Z = -1.36 m) and the vertical wall (R = 1.02 m).

The intersection of the homoclinic and heteroclinic lobes with divertor targets and walls forms objects referred to as magnetic footprints on the R,ϕ and Z,ϕ planes of the divertor targets and walls respectively as shown in Fig. 6(a) for the HFS wall and Fig. 6(b) for the LFS divertor target plate for the same conditions as in Fig. 5. Measurements of the heat (Evans et al., 2005, Evans et al., 2007) and particle (Schmitz et al., 2008) flux distributions on the

DIII-D divertors have been shown to be qualitatively consistent with numerical calculations of magnetic footprints produced by homoclinic lobes during experiments with applied non-axisymmetric magnetic fields from field-error correction and edge MHD (ELM) suppression coils (Evans et al., 2006). Similar heat flux patterns have been observed in the ASDEX-U tokamak (Eich et al., 2005) during ELMs. Quantitative comparisons of heat flux measurements inside these footprints with numerical simulations indicate that the separation between adjacent lobe intersections with the divertor targets can be a factor of 2-3 times larger than that predicted suggesting that there is a significant amplification of the homoclinic tangle structure from the applied n=1 field due to the response of the plasma (Evans et al., 2007). There are also indications that the topology of the lobes is affected by magnetic perturbations from MHD modes deep in the core plasma (Evans et al., 2005).

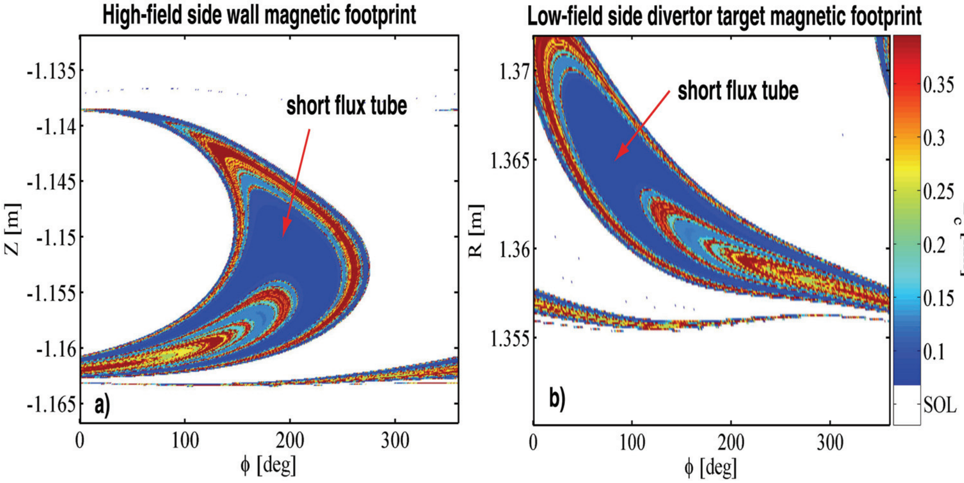


Fig. 6. Lower divertor (a) magnetic footprint formed on HFS vertical wall by an externally applied n=1 perturbation (no plasma response) from the DIII-D field-error correction coil with a current of 8 kAt in discharge 133908 at t = 2000 ms and (b) the LFS magnetic footprint formed on the horizontal divertor target plate. These footprints define the open field line hit points due to the intersection of the lobes of the homoclinic tangle shown in Fig. 5 with the target plate and wall. As in Fig. 5, L_c is the field line connection length between the horizontal target plate (Z = -1.36 m) and the vertical wall (R = 1.02 m).

3.2 Description of the temporal evolution prescribed by the model

A conceptual model describing the dynamics of the edge plasma and the evolution of the pedestal magnetic topology following the linear growth phase of a type-I ELM is presented. Understanding the physics, topology and dynamics of ELMs during their post-linear growth phase is essential for predicting the characteristics of these instabilities as a funtion of the pedestal plasma conditions. In particular, a model is needed that can be used to predict the temporal evolution of the plasma heat and particle distributions on the vessel wall and divertor components.

As discussed above, small quasi-static homoclinic and heteroclinic tangles result naturally from a variety of non-axisymmetric magnetic field perturbations commonly found in high performance poloidally diverted tokamaks. Examples of these non-axisymmetric field perturbations include toroidal field ripple, field-errors, core and edge MHD modes, 3D electromagnetic field control (trim) coils and small, spatially random, 3D field components due to magnetic materials and tolerence build-ups in the electromagnetic coils used to confine and shape the plasma (Evans et al., 2005; Evans et al., 2007). Thus, it is not unreasonable to expect the formation of separatrix homoclinic and heteroclinic tangles to be the norm rather than the exception, whether in a low confinement L-mode or in high confinement H-mode plasma as well as between and during ELMs. It is the existence of the separatrix topology associated with these tangles between ELMs that forms the basis of the model (Evans et al., 2009) described here.

Given the basic topology shown in Fig. 5, the model assumes that small fluctuations in the pedestal plasma pressure initiate a linearly growing MHD instability as the equilibrium conditions in a narrow region just inside the separatrix approach a marginal stability point. An example of this process is described by ideal MHD peeling-ballooning theory (Snyder et al., 2005; Wilson et al., 2006) which presumes that linearly growing intermediate n modes lead to the onset of the nonlinear growth phase. Peeling-ballooning theory predicts that the onset of this edge MHD mode significantly increases the radial heat and particle transport. Our model assumes that the energy associated with the linearly growing MHD mode flows into the coherent, short connection length, homoclinic flux tubes connecting the HFS wall and the LFS divertor target. At this point, fast parallel transport along these homoclinic flux tubes causes a rapid increase in the electron temperature (T_e) inside the magnetic footprints near the wall and divertor surfaces. Experimental measurements taken during the early growth of an ELM demonstrate that there is a rapid release of thermal energy from the area located near the steep gradient region leading up to the top of the pedestal just inside the separatrix (Kirk et al., 2007, Neuhauser et al., 2008). These observations are consistent with our requirement of a rapid increase in the radial energy transport during this time. These rapid bursts of energy flowing from the pedestal into the divertor appear to be correlated with an increase in broadband magnetic fluctuations in the pedestal starting about 10 μs before the onset of the nonlinear growth phase (Neuhauser et al., 2008) suggesting that currents in this region may play a key role in the onset of the nonlinear growth phase.

In our model, it is these inital heat pulses associated with the linearly growing MHD instability that provide the mechanism needed to form a feedback amplification loop. It is this feedback loop that causes the stable and unstable invariant manifolds of the initial homoclinic tangle to grow explosively. Here, it is presumed that the amplification process is triggered by the formation of field-aligned thermoelectric currents that flow through the short, pedestal plasma, homoclinic flux tubes connecting the inner wall and outer divertor target plate. These thermoelectric currents form when T_e at one end of a flux tube increases relative to T_e at the other end (Staebler & Hinton, 1989). Since part of the heat pulse enters the short flux tube near the equatorial plane on the LFS of the discharge it is expected to arrive at the LFS target well before arriving at the HFS wall. Numerical simulations of these two-poloidal-turn helical flux tubes (Wingen, et al. 2009a) show that the distance from the LFS equatorial plane to the LFS target plate is ~25 m while the distance to the HFS wall is ~75 m. In DIII-D H-mode plasmas, T_e just inside the separatrix, where these flux tubes reside, is ~400-500 eV. Thus, given an electron thermal velocity $v_{\mathrm{Te}} = (kT_e/m_e)^{1/2} = 8.4\times10^6$ m/s where k is Boltzmann's constant and m_e is the mass of an electron, these heat pulses arrive at the LFS target plate approximately 6 μs before reaching the HFS wall. This causes T_e near the LFS target plate to increase relative to that near the HFS wall and initiates the flow of a thermoelectric current from the LFS target to the HFS wall with a return current connecting through the lower divertor vessel structure.

Thus, the model assumes that following the release of the initial heat pulse from the linearly growing MHD mode a small field-aligned thermoelectric current begins to flow in a helical flux tube formed by a small preexisting homoclinic tangle. Although the time evolution of the current growth is not specifically predicted by the model, it is assumed that as the current grows its magnetic field perturbs the upper and lower hyperbolic points causing the lobes of the homoclinic tangle and their associated magnetic footprints to increase in size. Simulations have been carried out assuming the current density in the flux tube is limited to approximately 1/2 the initial ion saturation current density (~70-80 mA/mm^2) during the nonlinear phase of the instability (Wingen, et al., 2009a). These simulations have shown that the magnetic footprint, associated with a single n=1 flux tube connecting the primary (lower) LFS divertor target with the HFS wall, grows from an area of 1760 mm^2, with a current of 135.5 A, to an area of 3564 mm^2 with a current of 274.4 A. During this process, a topological bifurcation takes place that creates a new set of n=2 flux tubes connecting the primary divertor LFS target to the HFS wall (Wingen, et al., 2009a). It is then assumed, that as the thermoelectric current grows with increasing footprint area there is a commensurately increasing flow of energy from the pedestal into the flux tube that maintains the constant current density. Here, the working hypothesis is that the growing helical thermoelectric current filaments associated with the short connection length flux tubes also produce resonant magnetic field components that open magnetic islands (*i.e.*, Poincaré islands) on rational surfaces across the pedestal region in addtion to perturbing the nominally axisymmetric hyperbolic points. As these islands grow and overlap they produce an increase in the local magnetic field line stochasticity which enhances the effective radial heat tranport into the homoclinic flux tube containing the thermoelectric current. This completes the feedback amplification loop and results in the initial explosive growth phase of the topological instability.

During the next step in the process, the initial helical current filament grows explosively and acts to amplify the lobes of the homoclinic tangle while inducing a growing level of pedestal stochasticity that penetrates deeper into the core plasma as it grows. This process results in a self-amplification of the lobes due to a positive feedback loop between the size of the tangle lobes, an increasing stochastic layer width and an increase in the heat flux to the target plates that drives an increasing flow of current. This process takes on the appearance of growing helical filaments that protrude beyond the edge of the plasma and seem to propagate radially outward as they grow. A key feature of the processes involved up to this point is that there is no need to invoke field line tearing and reconnection during the evolution of an ELM. The entire process can be described using ideal MHD theory without requiring resistive or dissipative effects that would cause the filaments to tear and separate from the edge of the plasma. Such a process would rapidly shut down the thermal transport responsible for the growth of the instability. This is seen by comparing the 1-2 ms decay time following the current peak in Fig. 3 to a tearing mode growth time $\gamma^{-1} \sim \tau_r^{3/5}\tau_A^{2/5}$ s where γ is the growth rate of the tearing mode, τ_r is the resistive time and τ_A is the Alfvén time in the pedestal. We find that $\gamma^{-1} \leq 0.1$ ms where $\tau_r = 1.2\text{x}10^{-4}$ s and $\tau_A = 5.8\text{x}10^{-5}$ s or approximately an order of magnitude shorter than the current decay time. Therefore, the growth of a tearing mode following the peak in the current would cause a separation of the filament from the pedestal and a rapid, sub-millisecond, termination of the of the current. Instead, we see that the ELM is a radially extended structure, as indicated by the relatively constant ratio of the signals from two adjacent Langmuir probes in Fig. 3, which is reminiscent of the lobes of a homoclinic tangle and that this radial structure persists as the current slowly decays. This relatively slow decay of the current is more consistent with a slow shutdown of the heat flux from the pedestal as the energy reservoir in this region is slowly depleted and T_e in the short flux tubes drops. This reduction in T_e causes an increase in resistivity in the short flux tubes which, when coupled with a cooling of the plasma in front of the divertor target plate due to an increase in particle recycling, as shown in Fig. 2, slowly reduces the thermoelectric current flowing between the target plate and the wall.

Numerical simulations of the growth experienced by a pre-existing, field-error related, homoclinic tangle have been carried out using current filaments that are proportional to the area of the magnetic footprint on the divertor target plate. Results from these simulations demonstrate that the calculated nonlinear dynamics of the tangle's topology are consistent with the heat flux patterns measured in the DIII-D divertor during a type-I ELM (Wingen et al., 2009a). A key question studied during these simulations addresses how the topological evolution prescribed by the model conforms to experimental measurements of type-I ELM dynamics. In particular, data such as that shown in Fig. 4 suggest that the peak in the toroidal mode spectrum of an ELM increases in mode number during the nonlinear growth phase. As discussed below, a bifurcation in the separatrix topology has been identified during the early growth phase of the instability. This bifurcation involves the appearance of heteroclinic invariant manifolds associated with the upper (secondary) hyberbolic point.

3.3 Dynamics of an ELM-induced homoclinic-to-heteroclinic separatrix bifurcation

Here, we describe the appearance of a homoclinic-to-heteroclinic bifurcation as the total current flowing in a short flux tube, connecting the LFS divertor target plate to the HFS wall, increases from 100 to 300 A. The simulation starts with an axisymmetric plasma equilibrium. We then superimpose a spectrum of nonaxisymmetric magnetic perturbations due to field-errors that have been systematically measured in the DIII-D tokamak (Luxon et al., 2003) along with a 3D magnetic perturbation field produced by a field-error correction coil (referred to as the I-coil) in DIII-D discharge 133908 at t = 2000 ms (Wingen et al., 2009a). Note that this is the same plasma equilibrium shown in Fig. 5 but there an artificial n=1 nonaxisymmetric magnetic field is applied by a coil referred to as the C-coil with a relatively large current in order to highlight the properties of the homoclinic tangle. In the simulation discussed here, we use the actual coil currents that were employed during the experiment in discharge 133908.

As a starting point for this simulation, the shortest flux tube produced by the field-errors and an n=1 correction coil is selected. Initially, there is only one relatively small flux tube connecting the LFS side lower (primary) divertor target plate with the HFS wall. We refer to this as flux tube number 1. This flux tube makes two poloidal revolutions along its path through the pedestal plasma just inside the separatrix and has a total length from the target plate to the wall of ~100 m. The current flowing in a large divertor tile sensor is used to establish a current density calibration for the simulation. This is done by calculating the area of overlap between the tile sensor and the magnetic footprint at the toroidal and radial position of the tile when the maximum current during an ELM is reached. Using the calculated area of intersection with the tile sensor and an assumed current density of 77 mA/mm^2 we get 200 A which agrees with the measured current in this tile sensor at the peak amplitude of the ELM. The assumed current density (about 1/2 the pre-ELM ion saturation current) is held fixed throughout the remainder of simulation while the topology of the separatrix unfolds. We start with a relatively small current in flux tube number 1 and increase the current in steps. With each iteration of the code, the area of the magnetic footprints increases as the size of the lobes produced by the homoclinic tangle associated with the primary divertor hyperbolic point increases. The current is increased until the total area of all the magnetic footprints overlapping the tile sensor equals ~3000 mm^2. At this point, the area of the three footprints associated with flux tubes 1, 2 and 3 is calculated and using the assumed current density of 77 mA/mm^2 a total current of ~4.9 kA is obtained (Wingen et al., 2009a).

During the sequence of iterations in the current flowing in flux tube number 1, a new pair of flux tubes is formed followed by the formation of a fourth flux tube at a higher current. The first pair of flux tubes, referred to as flux tube number 2 and 3, connect the primary LFS divertor target plate to the HFS wall after one poloidal turn and have a length of ~50 m. Flux tubes 2 and 3 are formed during a bifurcation of the separatrix topology that involves a splitting of the invariant manifolds, caused by the presence of the secondary (upper) hyperbolic point, into a higher order set of stable and unstable branches of the original manifold topology. We refer to this as a homoclinic-to-heteroclinic bifurcation although here we focus only on the increased complexity of the homoclinic tangle associated with the primary (lower) hyperbolic point.

Figure 7(b) shows the structure of the manifolds produced by the primary (lower divertor) hyperbolic point in the secondary divertor region near the upper hyperbolic point with a current of 100 A flowing in flux tube number 1. Flux tube number 1 is not large enough to be clearly identified at this level of current. With this current, the initial formation of flux tubes 2 and 3 has begun. Here, flux tubes are formed in the area between intersecting stable and unstable manifolds. As seen in Fig. 8(a) flux tube number 3 is completed at 130 A when the stable and unstable manifolds intersect while flux tube number 2 is not yet fully formed at 150 A in Fig. 8(b).

As the current in flux tube number 1 is increased from 150 A to 200 A, flux tube number 2 is completed and a new partially formed flux tube appears, flux tube number 4 as shown in Fig. 9(a), on each side of flux tube number 2. Between 200 A and 300 A flux tube number 4 is completed and manifold connections are made between the secondary (upper) divertor LFS target plate and the primary (lower) HFS wall as well as between the primary LFS target plate and the secondary HFS wall as shown in Fig. 9(b).

From this point on in the simulation a current proportional to the area of intersection of flux tubes 2 and 3 with the primary LFS divertor target, having a current density of 77 mA/mm^2, is included at each subsequent step until the current limit discussed above is reached. As the simulation proceeds flux tubes 2 and 3, which form a pair of single poloidal turn helical structures that are displaced from each other toroidally by 180°, produce an n=2 perturbation that dominates the growth of the lobes and the primary divertor LFS target plate magnetic footprints.

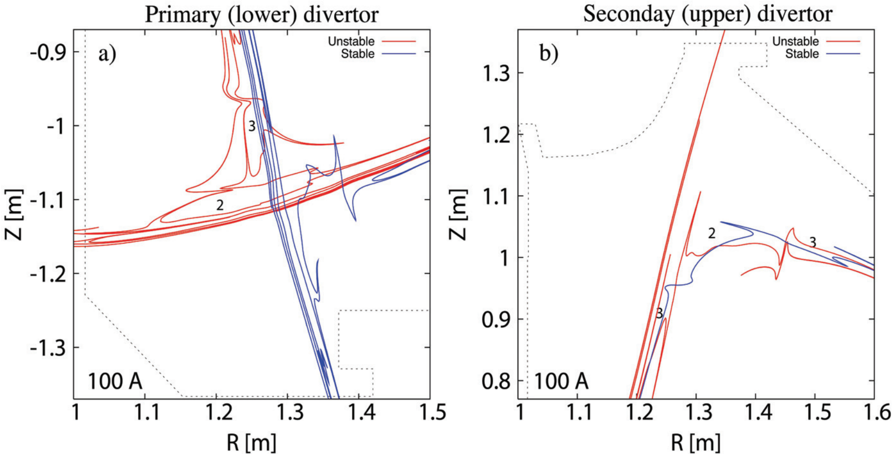


Fig. 7. Poincaré plots of (a) the calculated structure of the stable and unstable invariant manifolds in the primary divertor with a current of 100 A in flux tube number 1 (not clearly visible) and (b) the corresponding structure of the manifolds in the secondary divertor. The numbers 2 and 3 indicate regions where flux tube number 2 and 3 will form as the current in flux tube number 1 is increased in the simulation once the stable and unstable manifolds intersect.

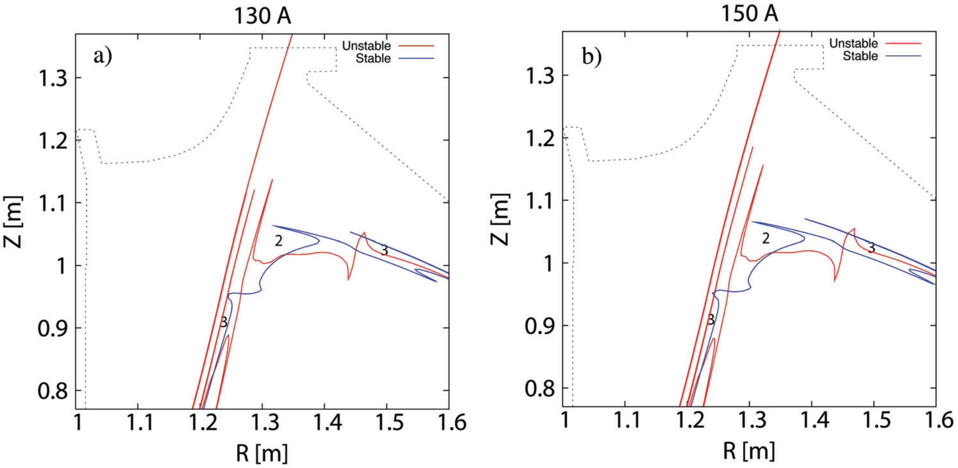


Fig. 8. Poincaré plots of (a) the formation of flux tube number 3 in the secondary divertor as the current in flux tube 1 is increased to 130 A, (b) flux tube number 2 is not completely formed at 150 A.

4. Discussion and conclusion

A conceptual model describing the nonlinear gowth of type-I ELMs in high performance tokamak plasmas has been presented along with a numerical simulation of the separatrix evolution, described by the model, during an ELM in a typical DIII-D H-mode plasma. The

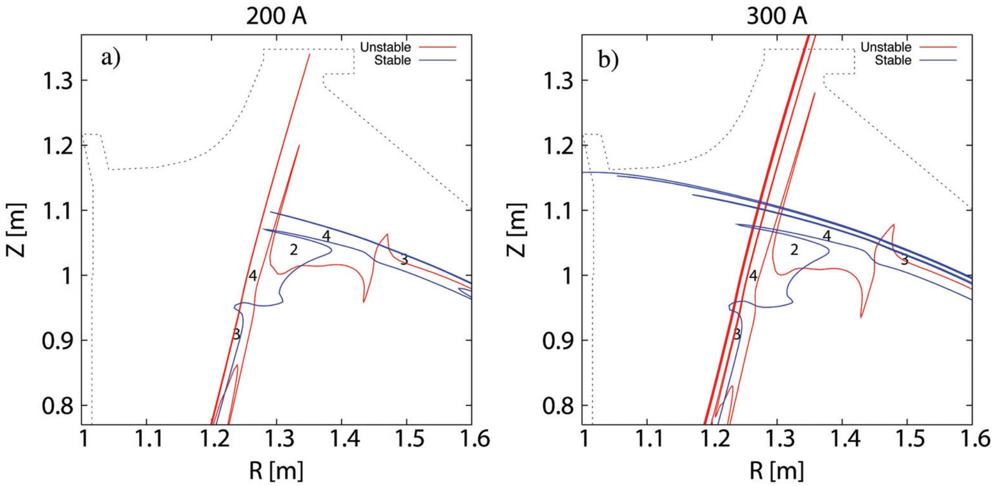


Fig. 9. Poincaré plots of (a) the formation of flux tube number 2 in the secondary divertor at 200 A in flux tube number 1 and the appearance of a new partially formed flux tube (number 4) while (b) at 300 A in flux tube number 1 all of the new flux tubes (numbers 2, 3 and 4) are fully formed.

temporal evolution of the separatrix is driven by a plasma instability resulting from a rapidly growing current that flows through the pedestal region of the plasma and changes the global topology of the manifolds that make up the separatrix. This topological change involves a homoclinic-to-heteroclinic bifurcation of the secondary (upper) hyperbolic point in the equilibrium magnetic field. The bifurcation creates an n=2 helical structure, consisting of two independent flux tubes separated by 180^{o} toroidally, early in nonlinear growth phase when a small, 150-200 A, field-aligned current flows in the original n=1 flux tube created by field-errors and a field-error correction coil. Although reversing the direction of the current in the n=1 flux tube does not have a significant effect on the structure of the invariant manifolds associated with the tangle structures, distributing the current into multiple filaments rather than allowing it in the single filament, as in the simulation shown in Sec. 3.3, results in a much more complex topology that has significantly more lobes intersecting the primary LFS divertor target plate. Thus, the model predicts the formation of a new set of invariant manifolds associated with the secondary hyperbolic point. These new invariant manifolds intersect the upper (secondary) divertor target plate and the HFS wall in DIII-D during the nonlinear growth phase of an ELM. This has significant implications for fusion reactor designs since it implies that complex heat and particle flux striations, associated with the magnetic footprints and flux tubes due to these new separatrix manifolds, should cause large impulsive energy bursts on secondary plasma facing surfaces that are not typically designed to handle such interactions with high energy density plasmas. Therfore, it is important that these predictions be tested using high time resolution measurements of the transient heat and particle flux interactions with plasma facing components near the secondary hyperbolic point during ELMs.

Another important question to ask of the model is whether it can be used to shed light on the physics of ELM suppression when small (~50 G), stationary, n=3 magnetic pertrubations are applied to ELMing H-modes in the DIII-D tokamak (Evans et al, 2006). Here, an interesting hypothessis that can be tested is that the n=3 field interacts with the lobes of the n=1 field-error tangle causing them to break up into relatively small scale structures that are more effective for dissapating the steady-state heat and particle flux over a much larger area of the divertor. This is expected to provide additional control over the pedestal transport that could be used to keep the pressure gradient below the threshold required for the onset of the linearly growing MHD installability. Constructive and destructive interference between nonaxisymmetric magnetic pertrubations from various coils in DIII-D has been studied previously. This work demonstrated that such interactions lead to much more complex lobe structures (Wingen et al., 2009b) that tends to spread the footprints and open more flux tubes (Evans et al, 2007) which can be used to fine tune the pedestal transport. Extending this hypothesis to higher n homoclinic structures, such as n=4 up to n=6 or 7 or combinations of structures with toroidal mode numbers ranging from 1 thorugh 7, suggests that the effect may provide much better control over the height and width of the pedestal region thus allowing the possibility of fine tuning of the pressure gradient profile. With advanced realtime pedestal profile diagnostics, it should be possible to combine this multimode perturbation field approach with an edge pressure and current gradient tracking algorithm to obtain a desired set of pedestal properties, particularly if an edge-localized heating and current drive system such as electron cyclotron system, were to be included as part of the feedback loop.

In general, the model presented here qualitatively fits some of the observed experimental attributes of large type-I ELMs such as a slow decay rate of the current in the flux tube as seen in Fig. 3. The model also has elements that may explain the variability seen in ELM signatures such as the divertor recycling emissions when effects such as type-II ELMs between the type-I ELMs are included. Other effects, such as an increase in the frequency of the ELMs with increasing heating power and the apprent rotation of the ELM structure during the nonlinear growth phase, have not yet been addressed by the model. These will be the focus of future work along with more detailed experimental comparisons.

5. Acknowledgments

This work was supported by the US Department of Energy under DE-FC02-04ER54698, DE-AC04-94AL85000, and DE-AC52-07NA27344.

6. References

Abdullaev, S.S. (2006). *Construction of Mappings for Hamiltonian Systems and Their Applications,* Lecture Notes in Physics, Vol. 691, Springer, ISBN-10 3-540-30915-2, Berlin.

Boedo, J.A.; Rudakov, D.L.; Hollmann, E.M.; Gray, D.S.; Burrell, K.H.; *et al.* (2005). Edge-localized mode dynamics and transport in the scrape-off layer of the DIII-D tokamak. *Physics of Plasmas* 12, 072516:1-11.

Callen, J.D.; Carreras, B.A.; & Stambaugh, R.D. (1992) Stability and transport processes in tokamak plasmas. *Physics Today* 45, 34-42.

Connor, J.W. (1998). A review of models for ELMs. *Plasma Physics and Controlled Fusion* 40, 191-213.

Cowley, S.C.; Wilson, H.; Hurricane, O & Fong, B. (2003). Explosive instabilities: from solar flares to edge localized modes in tokamaks. *Plasma Physics and Controlled Fusion* 45, A31-A38.

Dankowicz, H. (1997). *Chaotic Dynamics in Hamiltonian Systems with applications to celestial mechanics*, World Scientific Series on Nonlinear Science, Series A, Vol. 25, World Scientific, ISBN 9810232217, Singapore.

D'haeseleer, W.D.; Hitchon, W.N.G.; Callen, J.D. & Shohet, J.L. (1991). *Flux coordinates and magnetic field structure, a guide to a fundamental tool for plasma theroy*, Springer Series in Computational Physics, Springer-Verlag ISBN 3-540-52419-3, Berlin.

Eich, T.; Herrmann, A.; Neuhauser, J.; Dux, R.; Fuchs, J.C.; *et al.* (2005). Type-I ELM structure on divertor target plates in ASDEX Upgrade. *Plasma Physics and Controlled Fusion* 47, 815-842.

Evans, T.E.; Lasnier, C.J.; Hill, D.N.; Leonard, A.W.; Fenstermacher, M.E.; et al. (1995). Measurements of non-axisymmetric effects in the DIII-D divertor. *Journal of Nuclear Materials* 220-222, 235-239.

Evans, T.E.; Moyer, R.A.; Stephan, E.A.; Snider, R.T. & Coles, W.A. (1996). Causal spacio-temporal correlations of short scale length solar wind accelaration and heating mechanisma with a solar event correlation analyzer (SECA) instrument package. *Robotic exploration close to the sun: scientific basis*, AIP Conference Proceedings 385, pp. 145-152, Editor: S. R. Habbal, American Institute of Physics ISBN 1-56396-618-2, New York.

Evans, T.E.; Roeder, R.K.W.; Carter, J.A. & Rapoport, B.I. (2004). Homoclinic tangles, bifurcations and stochasticity in poloidally diverted tokamaks. *Contributions to Plasma Physics* 44, 235-240.

Evans, T.E.; Roeder, R.K.W.; Carter, J.A.; Rapoport, B.I.; Fenstermacher, M.E.; & Lasnier, C.J. (2005). Experimental signatures of homoclinic tangles in poloidally diverted tokamaks. *Journal of Physics: Confonference Proceedings Series* 7, 174-190.

Evans, T.E.; Moyer, R.A.; Burrell, K.H.; Fenstermacher, M.E.; Joseph, I.; et al. (2006). Edge stability and transport control with resonant magnetic pertrubations in collisionless tokamak plasmas. *Nature Physics* 2, 419-423.

Evans, T.E.; Joseph, I.; Moyer, R.A.; Fenstermacher, M.E.; Lasnier, C.J.; Yan, L.W. (2007). Experimental and numerical studies of seperatrix splitting and magnetic footfrints in DIII-D. *Journal of Nuclear Materials* 363-365, 570-574.

Evans, T.E. (2008). Implications of topological complexity and Hamiltonian chaos in the edge magnetic field of toroidal fusion plasmas. *Chaos, Complexity and Transport: Theory and Applications*, pp. 147-176 Edited by: Chandre C.; Leoncini, X. & Zaslavsky, G. World Scientific Press, ISBN-13 978-981-281-897-9, Singapore.

Evans, T.E.; Yu, J.H.; Jakubowski, M.W.; Schmitz, O.; Watkins, J.G. & Moyer, R.A. (2009). A coneptual model of the magnetic topology and nonlinear dynamics of ELMs. *Journal of Nuclear Materials* 390-391, 789-792.

Fenstermacher, M.E.; Leonard, A.W.; Snyder, P.B.; Boedo, J.A.; Brooks, N.H.; et al., (2003). ELM particle and energy transport in the SOL and divertor of DIII-D. *Plasma Physics and Controlled Fusion* 45, 1597-1626.

Gibons, M. & Spicer, D.S. (1981). On line tying. *Solar Physics* 69, 57-61.

Guckenheimer, J. & Holmes, P. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields,* Applied Mathematical Science, Vol. 42 Springer-Verlag, ISBN 0-387-90819-6, New York.

Kirk, A.; Wilson, H.R.; Counsell, G.F.; Akers, R.; Arends, E.; et al. (2004). Spatial and temporal structure of edge-localized modes. *Physical Review Letters* 92, 245002:1-4.

Kirk, A.; Counsell, G.F.; Cunningham, G.; Dowling, J.; Dunstan, M.; et al. (2007). Evolution of the pedestal on MAST and the implications fo ELM power loadings. *Plasma Physics and Controlled Fusion* 49, 1259-1275.

Lichtenberg, A.J. & Lieberman, M.A. (1992). *Regular and Chaotic Dynamics*, Applied Mathematical Sciences, Vol. 38 second edition Springer-Verlag, ISBN 0-387-97745-7, New York.

Loarte, A.; Saibene, G.; Sartori, R.; Becoulét, M.; Horton, L.; et al. (2003). ELM energy and particle losses and their extrapolation to burning plasma experiments. *Journal of Nuclear Materials* 313-316, 962-966.

Luxon, J.L. (2002). A design retrospective of the DIII-D tokamak. *Nuclear Fusion* 42, 614-633.

Luxon, J.L.; Schaffer, M.J.; Jackson, G.L.; Leuer, J.A.; Nagy, A.; et al. (2003). Anomalies in the applied magnetic fields in DIII-D and their implications for the understanding of stability experiments. *Nuclear Fusion* 43, 1813-1828.

Maingi, R.; Bush, C.E.; Fredrickson, E.D.; Gates, D.A.; Kaye, S.M.; (2005). H-mode pedestal, ELM and power threshold studies in NSTX. *Nuclear Fusion* 45, 1066-1077.

Neuhauser, J.; Bobkov, V.; Conway, G.D.; et al. (2008). Structure and dynamics of spontaneous and induced ELMs on ASDEX Upgrade. *Nuclear Fusion* 48, 045005:1-15.

Osborne, T. H.; Ferron, J.R.; Groebner, R.J.; Lao, L.L.; Leonard, A.W.; et al. (2000). The effect of plasma shape on H-mode pedestal characteristics on DIII-D. *Plasma Physics and Controlled Fusion* 42, A175-A184.

Roeder, R.K.W.; Rapoport, B.I. & Evans, T.E.; (2003). Elplicit calcualtions of homoclinic tangles in tokamaks. *Physics of Plasmas* 10, 3796-3799.

Simiu, E. (2002). *Chaotic Transitions in Deterministic and Stochastic Dynamical Systems*, Princeton Series in Applied Mathematics, Princeton University Press, ISBN 0-691-05094-5, Princeton, New Jersey.

Schmitz, O.; Evans, T.E.; Fenstermacher, M.E.; Frerichs, H.; Jakubowski, M.W.; et al. (2008). Aspects of three dimensional transport for ELM control experiments in ITER-similar shape plasmas at low collisionality in DIII-D. *Plasma Physics and Controlled Fusion* 50, 124029:1-19.

Snyder, P.B.; Wilson, H.R. & Xu, X.Q. (2005). Progress in the peeling-ballooning model of edge localized modes: Numerical studies of the nonlinear dynamics. *Physics of Plasmas* 12, 056115:1-7.

Staebler , G.M. & Hinton, F.L. (1989). Currents in the scrape-off layer of diverted tokamaks. *Nuclear Fusion* 29, 1820-1824.

Suttrop, W. (2000). The physics of large and small edge localized modes. *Plasma Physics and Controlled Fusion* 42 pp. A1-A14.

Wesson, J. (2004). *Tokamaks*, 3rd Ed. Oxford University Press, ISBN 0-19-8509227, New York.

Wilson, H.R.; Cowley, S.C.; Kirk, A.; & Snyder, P.B. (2006). Magnetohydrodynamic stability of the H-mode transport barrier as a model for edge localized modes: an overview. *Plasma Physics and Controlled Fusion* 48, A71-A84.

Wingen, A.; Evans, T.E.; Lasnier, C.J. & Spatschek, K.H. (2009a). Numerical modelling of the nonlinear ELM cycle in tokamaks. *Physical Review Letters* Vol. -- pp. – (submitted).

Wingen, A.; Evans, T.E. & Spatschek, K.H. (2009b). High resolution numerical studies of the separatrix splitting due to non-axisymmetric perturbation in DIII-D. *Nuclear Fusion* 49, 055027:1-8.

Wingen, A.; Evans, T.E. & Spatschek, K.H. (2009c). Footprint structures due to resonant magnetic perturbations in DIII-D. *Physics of Plasmas* 16, 042504:1-5.

Yu, J.H.; Boedo, J.A.; Hollmann, E.M.; Moyer, R.A. & Rudakov, D.L. (2008). Fast imaging of edge localized mode structure and dynamics in DIII-D. *Physics of Plasmas* 15, 032504:1-7.

Zaslavsky, G.M. (2005). *Hamiltonian chaos & fractional dynamics*, Oxford University Press ISBN 0-19-852604-0, Oxford.

Zohm, H. (1996). Edge localized modes (ELMs). *Plasma Physics and Controlled Fusion* 38, 105-128.

4

Nonlinear Dynamics of Cantilever Tip-Sample Surface Interactions in Atomic Force Microscopy

John H. Cantrell[1] and Sean A. Cantrell[2]
[1]*NASA Langley Research Center*
[2]*Johns Hopkins University*
USA

1. Introduction

The atomic force microscope (AFM) (Bennig et al., 1986) has become an important nanoscale characterization tool for the development of novel materials and devices. The rapid development of new materials produced by the embedding of nanostructural constituents into matrix materials has placed increasing demands on the development of new nanoscale measurement methods and techniques to assess the microstructure-physical property relationships of such materials. Dynamic implementations of the AFM (known variously as acoustic-atomic force microscopies or A-AFM and scanning probe acoustic microscopies or SPAM) utilize the interaction force between the cantilever tip and the sample surface to extract information about sample material properties. Such properties include sample elastic moduli, adhesion, surface viscoelasticity, embedded particle distributions, and topography. The most commonly used A-AFM modalities include various implementations of amplitude modulation-atomic force microscopy (AM-AFM) (including intermittent contact mode or tapping mode) (Zhong et al., 1993), force modulation microscopy (FMM) (Maivald et al., 1991), atomic force acoustic microscopy (AFAM) (Rabe & Arnold, 1994; Rabe et al., 2002), ultrasonic force microscopy (UFM) (Kolosov & Yamanaka, 1993; Yamanaka et al., 1994), heterodyne force microscopy (HFM) (Cuberes et al., 2000; Shekhawat & Dravid, 2005), resonant difference-frequency atomic force ultrasonic microscopy (RDF-AFUM) (Cantrell et al., 2007) and variations of these techniques (Muthuswami & Geer, 2004; Hurley et al., 2003; Geer et al., 2002; Kolosov et al., 1998; Yaralioglu et al., 2000; Zheng et al., 1006; Kopycinska-Müller et al., 2006; Cuberes, 2009).

Central to all A-AFM modalities is the AFM. As illustrated in Fig. 1, the basic AFM consists of a scan head, an AFM controller, and an image processor. The scan head consists of a cantilever with a sharp tip, a piezoelement stack attached to the cantilever to control the distance between the cantilever tip and sample surface (separation distance), and a light beam from a laser source that reflects off the cantilever surface to a photo-diode detector used to monitor the motion of the cantilever as the scan head moves over the sample surface. The output from the photo-diode is used in the image processor to generate the micrograph.

The AFM output signal is derived from the interaction between the cantilever tip and the sample surface. The interaction produces an interaction force that is highly dependent on the tip-sample separation distance. A typical force-separation curve is shown in Fig. 2. Above the separation distance z_A the interaction force is negative, hence attractive, and below z_A the interaction force is positive, hence repulsive. The separation distance z_B is the point on the curve at which the maximum rate of change of the slope of the curve occurs and is thus the point of maximum nonlinearity on the curve (the maximum nonlinearity regime).

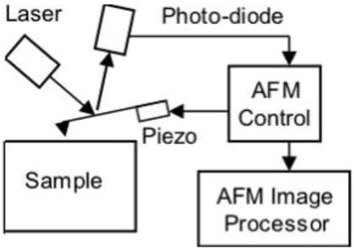


Fig. 1. Schematic of the basic atomic force microscope.

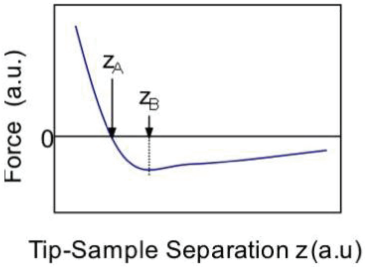


Fig. 2. Interaction force plotted as a function of the separation distance z between cantilever tip and sample surface.

Modalities, such as AFM and AM-AFM, are available for near-surface characterization, while UFM, AFAM, FMM, HFM, and RDF-AFUM are generally used to assess deeper (subsurface) features at the nanoscale. The nanoscale subsurface imaging modalities combine the lateral resolution of the atomic force microscope with the nondestructive capability of acoustic methodologies. The utilization of the AFM in principle provides the necessary lateral resolution for obtaining subsurface images at the nanoscale, but the AFM alone does not enable subsurface imaging. The propagation of acoustic waves through the bulk of the specimen and the impinging of those waves on the specimen surface in contact with the AFM cantilever enable such imaging. The use of acoustic waves in the ultrasonic range of frequencies leads to a more optimal resolution, since both the intensity and the phase variation of waves scattered from nanoscale, subsurface structures increase with increasing frequency (Überall, 1997).

A schematic of the equipment arrangement for the various A-AFM modalities is shown in Fig. 3. The arrangement used for AFAM and FMM is shown in Fig. 3 where the indicated switches are in the open positions. AFAM and FMM utilize ultrasonic waves transmitted into the material by a transducer attached to the bottom of the sample. After propagating through the bulk of the sample, the wave impinges on the sample top surface where it excites the engaged cantilever. For AFAM and FMM the cantilever tip is set to engage the sample surface in hard contact corresponding to the roughly linear interaction region below

z_A of the force-separation curve. The basic equipment arrangement used for UFM is the same as that for AFAM and FMM, except that the cantilever tip for UFM is set to engage the sample in the maximum nonlinearity regime of the force-separation curve. The UFM output signal is a static or "dc" signal resulting from the interaction nonlinearity.

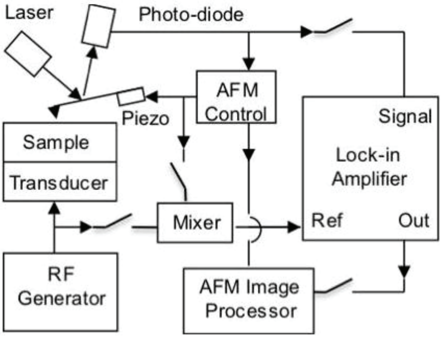


Fig. 3. Acoustic-atomic force microscope equipment configuration. Switches are open for AFAM, FMM, and UFM. Switches are closed for HFM and RDF-AFUM.

The equipment arrangement for RDF-AFUM and HFM is shown in Fig. 3 where the indicated switches are in the closed positions. Similar to the AFAM, FMM and UFM modalities, RDF-AFUM and HFM employ ultrasonic waves launched from the bottom of the sample. However, in contrast to the AFAM, FMM and UFM modalities, the cantilever in RDF-AFUM and HFM is also driven into oscillation. RDF-AFUM and HFM operate in the maximum nonlinearity regime of the force-separation curve, so the nonlinear interaction of the surface and cantilever oscillations produces a strong difference-frequency output signal. For the AM-AFM modality only the cantilever is driven into oscillation and the tip-sample separation distance may be set to any position on the force-separation curve. In one mode of AM-AFM operation the rest or quiescent separation distance z_0 lies well beyond the region of strong tip-sample interaction, i.e. the quiescent separation $z_0 >> z_B$.

Various approaches to assessing the nonlinear behavior of the cantilever probe dynamics have been published (Kolosov & Yamanaka, 1993; Yamanaka et al., 1994; Nony et al., 1999; Yagasaki, 2004; Lee et al., 2006; Kokavecz et al., 2006; Wolf & Gottlieb, 2002; Turner, 2004; Stark & Heckl, 2003; Stark et al., 2004; Hölscher et al., 1999; Garcia & Perez, 2002). We present here a general, yet detailed, analytical treatment of the cantilever and the sample as independent systems in which the nonlinear interaction force provides a coupling between the cantilever tip and the small volume element of sample surface involved in the coupling. The sample volume element is itself subject to a restoring force from the remainder of the sample. We consider only the lowest-order terms in the cantilever tip-sample surface, interaction force nonlinearity. Such terms are sufficient to account for the most important operational characteristics and material properties obtained from each of the various acoustic-atomic force microscopies cited above. A particular advantage of the coupled independent systems model is that the equations are valid for all regions of the force-separation curve and emphasize the local curvature properties (functional form) of the curve. Another advantage is that the dynamics of the sample, hence energy transfer characteristics, can be extracted straightforwardly from the solution set using the same mathematical procedure as that for the cantilever.

We begin by developing a mathematical model of the interaction between the cantilever tip and the sample surface that involves a coupling, via the nonlinear interaction force, of separate dynamical equations for the cantilever and the sample surface. A general solution is found that accounts for the positions of the excitation force (e.g., a piezo-transducer) and the cantilever tip along the length of the cantilever as well as for the position of the laser probe on the cantilever surface. The solution contains static terms (including static terms generated by the nonlinearity), linear oscillatory terms, and nonlinear oscillatory terms. Individual or various combinations of these terms are shown to apply as appropriate to a quantitative description of signal generation for AM-AFM and RDF-AFUM as representatives of the various A-AFM modalities. The two modalities represent opposite extremes in measurement complexity, both in instrumentation and in the analytical expressions used to calculate the output signal. This is followed by a quantitative analysis of image contrast for the A-AFM techniques. As a test of the validity of the present model, comparative measurements of the maximum fractional variation of the Young modulus in a film of LaRC^{TM}-CP2 polyimide polymer are presented using the RDF-AFUM and AM-AFM modalities.

2. Analytical model of nonlinear cantilever dynamics

2.1 General dynamical equations

The cantilever of the AFM is able to vibrate in a number of different modes in free space corresponding to various displacement types (flexural, longitudinal, shear, etc.), resonant frequencies, and effective stiffness constants. Although any shape or oscillation mode of the cantilever can in principle be used in the analysis to follow, for definiteness and expediency we consider only the flexural modes of a cantilever modeled as a rectangular, elastic beam of length L, width a, and height b. We assume the beam to be clamped at the position x = 0 and unclamped at the position x = L, as indicated in Fig. 4. We consider the flexural displacement y(x,t) of the beam to be subjected to some general force per unit length H(x,t), where x is the position along the beam and t is time. The dynamical equation for such a beam is

$$E_B I \frac{\partial^4 y(x,t)}{\partial x^4} + \rho_B A_B \frac{\partial^2 y(x,t)}{\partial t^2} = H(x,t) \tag{1}$$

where E_B is the elastic modulus of the beam, $I = ab^3/12$ is the bending moment of inertia, ρ_B is the beam mass density, and $A_B = ab$ is the cross-sectional area of the beam.

The solution to Eq. (1) may be obtained as a superposition of the natural vibrational modes of the unforced cantilever as

$$y(x,t) = \sum_{n=1}^{\infty} Y_n(x)\eta_{cn}(t) \tag{2}$$

where η_{cn} is the nth mode cantilever displacement (n = 1, 2, 3, ⋯) and the spatial eigenfunctions $Y_n(x)$ form an orthogonal basis set given by (Meirovitch, 1967)

$$Y_n(x) = \left(\frac{\sin q_n x - \sinh q_n x}{\cos q_n x + \cosh q_n x}\right)(\sin q_n x - \sinh q_n x) + (\cos q_n x - \cosh q_n x)\,. \tag{3}$$

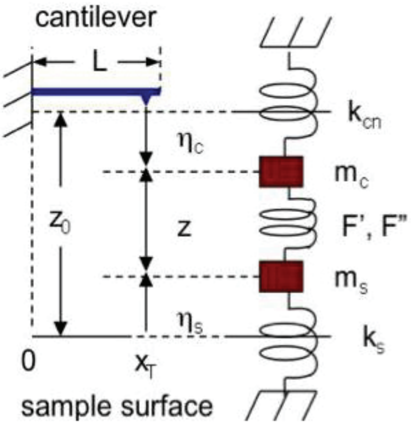


Fig. 4. Schematic of cantilever tip-sample surface interaction: z_0 is the quiescent (rest) tip-surface separation distance (setpoint), z the oscillating tip-surface separation distance, η_c the displacement (positive down) of the cantilever tip, η_s the displacement of the sample surface (positive up), k_{cn} is the nth mode cantilever stiffness constant (represented as an nth mode spring), m_c the cantilever mass, k_s the sample stiffness constant (represented as a single spring), m_s the active sample mass, and $F'(z_0)$ and $F''(z_0)$ are the linear and first-order nonlinear interaction force stiffness constants, respectively, at z_0.

The flexural wave numbers q_n in Eq.(3) are determined from the boundary conditions as $\cos(q_nL)\cosh(q_nL) = -1$ and are related to the corresponding modal angular frequencies ω_n via the dispersion relation $q_n^4 = \omega_n^2 \rho_B A_B / E_B I$. The general force per unit length H(x,t) can also be expanded in terms of the spatial eigenfunctions as (Sokolnikoff & Redheffer, 1958)

$$H(x,t) = \sum_{n=1}^{\infty} B_n(t) Y_n(x) \, . \tag{4}$$

Applying the orthogonality condition

$$\int_0^L Y_m(x) Y_n dx = L\delta_{mn} \tag{5}$$

(δ_{mn} are the Kronecker deltas) to Eq. (4), we obtain

$$B_n(t) = \int_0^L H(\xi,t) Y_n(\xi) d\xi \, . \tag{6}$$

We now assume that the general force per unit length acting on the cantilever is composed of (1) a cantilever driving force per unit length $H_c(x,t)$, (2) an interaction force per unit length $H_T(x,t)$ between the cantilever tip and the sample surface, and (3) a dissipative force per unit length $H_d(x,t)$. Thus, the general force per unit length $H(x,t) = H_c(x,t) + H_T(x,t) + H_d(x,t)$. We now assume that the driving force per unit length is a purely sinusoidal oscillation of angular frequency ω_c and magnitude P_c. We also assume the driving force to result from a drive element (e.g., a piezo-transducer) applied at the point x_c along the cantilever length. We thus write $H_c(x,t) = P_c e^{i\omega_c t}\delta(x - x_c)$ where $\delta(x - x_c)$ is the Dirac delta function. The interaction force per unit length $H_T(x,t)$ of magnitude P_T is applied at the cantilever tip at $x = x_T$ and is not a direct function of time, since it serves as a passive coupling between the independent cantilever and sample systems. We thus write the interaction force per unit length as $H_T(x,t) = P_T\delta(x - x_T)$. We assume the modal dissipation force per unit length $H_d(x,t)$ to be a product of the spatial eigenfunction and the cantilever displacement velocity given as $H_d(x,t) = -P_d Y_n(x)(d\eta_{cn}/dt)$. The coefficient $B_n(t)$ is then obtained from Eq. (6) as

$$B_n(t) = P_c e^{i\omega_c t} Y_n(x_c) + P_T Y_n(x_T) - [P_d \int Y_n(x)dx](d\eta_{cn}/dt) \tag{7}$$

where the integration in the last term is taken over the range x = 0 to x = L. Substituting Eqs. (2) and (4) into Eq.(1) and collecting terms, we find that the dynamics for each mode n must independently satisfy the relation

$$\rho_B A_B Y_n(x)\frac{d^2\eta_{cn}(t)}{dt^2} + E_B I\frac{d^4 Y_n(x)}{dx^4}\eta_{cn} = P_c e^{i\omega_c t} Y_n(x_c) Y_n(x) \tag{8}$$

$$+ P_T Y_n(x_T) Y_n(x) + [P_d \int_0^L Y_n(x)dx]\frac{d\eta_{cn}}{dt}.$$

From Eq. (3) we write $d^4 Y_n / dx^4 = q_n^4 Y_n$. Using this relation and the dispersion relation between q_n and ω_n, we obtain that the coefficient of η_{cn} in Eq.(8) is given by $E_B I(d^4 Y_n / dx^4) = \omega_n^2 \rho_B A_B$. Multiplying Eq. (8) by $Y_m(x)$ and integrating from x = 0 to x = L, we obtain

$$m_c\ddot{\eta}_{cn} + \gamma_c\dot{\eta}_{cn} + k_{cn}\eta_{cn} = F_c e^{i\omega_c t} + F \tag{9}$$

where the overdot denotes derivative with respect to time, $m_c = \rho_B A_B L$ is the total mass of the cantilever and $F_c = P_B L Y_n(x_c)$. The tip-sample interaction force F is defined by $F = P_T L Y_n(x_T)$ and the cantilever stiffness constant k_{cn} is defined by $k_{cn} = m_c\omega_n^2$. The damping coefficient γ_c of the cantilever is defined as $\gamma_c = P_d L \int Y_n(x)dx$. Note that, with regard to the coupled system response, for a given mode n the effective magnitudes of the driving term F_c and the interaction force F are dependent via $Y_n(x_c)$ and $Y_n(x_T)$, respectively, on the positions x_c and x_T at which the forces are applied. The damping factor, in contrast, results from a more general dependence on x via the integral of $Y_n(x)$ over the range zero to L. If the excitation force per unit length is a distributed force over the cantilever surface rather than at a point, then the resulting calculation for F_c would involve an integral over $Y_n(x)$ as obtained for the damping coefficient.

The interaction force F in Eq. (9) is derived without regard to the cantilever tip-sample surface separation distance z. Realistically, the magnitude of F is quite dependent on the separation distance. In particular, various parameters derived from the force-separation curve play an essential role in the response of the cantilever to all driving forces. We further consider that the interaction force not only involves the cantilever at the tip position x_T but also some elemental volume of material at the sample surface. To maintain equilibrium it is appropriate to view the elemental volume of sample surface as a mass element m_s (active mass) that, in addition to the interaction force, is subjected to a linear restoring force from material in the remainder of the sample. We assume that the restoring force per unit displacement of m_s in the direction z toward the cantilever tip is described by the sample stiffness constant k_s.

The interaction force F between the cantilever tip and the mass element m_s is in general a nonlinear function of the cantilever tip-sample surface separation distance z. A typical nonlinear interaction force F(z) is shown schematically in Fig. 2 plotted as a function of the cantilever tip-sample surface separation distance z. The interaction force results from a number of possible fundamental mechanisms including electrostatic forces, van der Waals forces, interatomic repulsive (e.g., Born-Mayer) potentials, and Casimir forces (Law & Rieutord, 2002; Lantz et al., 2001; Polesel-Maris et al., 2003; Eguchi & Hasegawa, 2002; Saint Jean et al.,; Chan et al., 2001). It is also influenced by chemical potentials as well as hydroxyl groups formed from atmospheric moisture accumulation on the cantilever tip and sample surface (Cantrell, 2004).

Since the force F(z) is common to the cantilever tip and the sample surface element, the cantilever and the sample form a coupled dynamical system. We thus consider the cantilever and the sample as independent dynamical systems coupled by their common interaction force F(z). Fig. 4 shows a schematic representation of the various elements of the coupled system. The dynamical equations expressing the responses of the cantilever and the sample surface to all driving and damping forces may be written for each mode n of the coupled system as

$$m_c\ddot{\eta}_{cn} + \gamma_c\dot{\eta}_{cn} + k_{cn}\eta_{cn} = F(z) + F_c\cos\omega_c t \tag{10}$$

$$m_s\ddot{\eta}_{sn} + \gamma_s\dot{\eta}_{sn} + k_s\eta_{sn} = F(z) + F_s\cos(\omega_s t + \theta) \tag{11}$$

where η_{cn} (positive down) is the cantilever tip displacement for mode n, η_{sn} (positive up) is the sample surface displacement for mode n, ω_c is the angular frequency of the cantilever oscillations, ω_s is the angular frequency of the sample surface vibrations, γ_c is the damping coefficient for the cantilever, γ_s is the damping coefficient for the sample surface, F_c is the magnitude of the cantilever driving force, F_s is the magnitude of the sample driving force that we assume here to result from an incident ultrasonic wave generated at the opposite surface of the sample. The factor θ is a phase contribution resulting from the propagation of the ultrasonic wave through the sample material and is considered in more detail in Section 2.2.

Eqs. (10) and (11) are coupled equations representing the cantilever tip-sample surface dynamics resulting from the nonlinear interaction forces. The equations govern the cantilever and surface displacements η_{cn} and η_{sn}, respectively at $x = x_T$. In a realistic AFM measurement of the cantilever response to the driving forces, the measurement point is not generally at $x = x_T$, but at the point $x = x_L$ at which the laser beam of the AFM optical detector system strikes the cantilever surface. The cantilever response at $x = x_L$ is found from Eq. (2) to be

$$y(x_L, t) = \eta_c(t) = \sum_{n=1}^{\infty} Y_n(x_L)\eta_{cn}(t) \tag{12}$$

We note from Fig. 4 that for a given mode n, $z = z_o - (\eta_{cn} + \eta_{sn})$, where z_0 is the quiescent separation distance between the cantilever tip and the sample surface (setpoint distance). We use this relationship in a power series expansion of $F(z)$ about z_o to obtain

$$F(z) = F(z_0) + F'(z_0)(z - z_0) + \frac{1}{2}F''(z_0)(z - z_0)^2 + \cdots \tag{13}$$

$$= F(z_0) - F'(z_0)(\eta_{cn} + \eta_{sn}) + \frac{1}{2}F''(z_0)(\eta_{cn} + \eta_{sn})^2 + \cdots$$

where the prime denotes derivative with respect to z. Substitution of Eq. (13) into Eqs. (10) and (11) gives

$$m_c\ddot{\eta}_{cn} + \gamma_c\dot{\eta}_{cn} + [k_{cn} + F'(z_0)]\eta_{cn} + F'(z_0)\eta_{sn} = F(z_0) + F_c\cos\omega_c t \tag{14}$$

$$+\frac{1}{2}F''(z_0)(\eta_{cn} + \eta_{sn})^2 + \cdots$$

$$m_s\ddot{\eta}_{sn} + \gamma_s\dot{\eta}_{sn} + [k_s + F'(z_0)]\eta_{sn} + F'(z_0)\eta_{cn} = F(z_0) + F_s\cos(\omega_s t + \theta) \tag{15}$$

$$+\frac{1}{2}F''(z_0)(\eta_{cn} + \eta_{sn})^2 + \cdots.$$

It is of interest to note that Eqs. (14) and (15) were obtained assuming that the cantilever is a rectangular beam of constant cross-section. Such a restriction is not necessary, since the mathematical procedure leading to Eqs. (14) and (15) is based on the assumption that the general displacement of the cantilever can be expanded in terms of a set of eigenfunctions that form an orthogonal basis set for the problem. For the beam cantilever the eigenfunctions are $Y_n(x)$. For some other cantilever shape a different orthogonal basis set of eigenfunctions would be appropriate. However, the mathematical procedure used here would lead again to Eqs. (14) and (15) with values of the coefficients appropriate to the different cantilever geometry.

2.2 Variations in signal amplitude and phase from subsurface features

We consider a traveling stress wave of unit amplitude of the form $e^{-\alpha x}\cos(\omega_s t - kx) = \mathrm{Re}[e^{-\alpha x}e^{i(\omega_s t - kx)}]$, where α is the attenuation coefficient, x is the propagation distance, ω_s is the angular frequency, t is time, $k = \omega_s / c$, and c is the phase velocity, propagating through a sample of thickness a/2. We assume that the wave is generated at the bottom surface of the sample at the position x = 0 and that the wave is reflected between the top and bottom surfaces of the sample. We assume that the effect of the reflections is simply to change the direction of wave propagation.

For continuous waves the complex waveform at a point x in the material consists of the sum of all contributions resulting from waves which had been generated at the point x = 0 and have propagated to the point x after multiple reflections from the sample boundaries. We thus write the complex wave $\overline{A}(t)$ as

$$\overline{A}(t) = e^{-\alpha x} e^{i(\omega_s t - kx)}[1 + e^{-(\alpha a + ika)} + \cdots + e^{-n(\alpha a + ika)} + \cdots]$$

$$= e^{-\alpha x} e^{i(\omega_s t - kx)} \sum_{n=0}^{\infty} \left[e^{-(\alpha a + ika)} \right]^n = e^{-\alpha x} e^{i(\omega_s t - kx)} \frac{1}{1 - e^{-(\alpha a + ika)}} \tag{16}$$

where the last equality follows from the geometric series generated by the infinite sum. The real waveform A(t) is obtained from Eq. (16) as

$$A(t) = \mathrm{Re}[\overline{A}(t)] = e^{-\alpha x} (A_1^2 + A_2^2)^{1/2} \cos(\omega_s t - kx - \phi) = e^{-\alpha x} B \cos(\omega_s t - kx - \phi) \tag{17}$$

where

$$A_1 = \frac{e^{\alpha a} - \cos ka}{2(\cosh \alpha a - \cos ka)}, \tag{18}$$

$$A_2 = -\frac{\sin ka}{2(\cosh \alpha a - \cos ka)}, \tag{19}$$

$$\phi = \tan^{-1} \frac{\sin ka}{e^{\alpha a} - \cos ka}, \tag{20}$$

and

$$B = (A_1^2 + A_2^2)^{1/2} = (1 + e^{-2\alpha a} - 2e^{-\alpha a} \cos ka)^{-1/2} . \tag{21}$$

The evaluation (detection) of a continuous wave at the end of the sample opposite that of the source is obtained by setting x = a/2 in the above equations. It is at x = a/2 that the AFM cantilever engages the sample surface. In the following equations we set x = a/2.

The above results are derived for a homogeneous specimen. Consider now that the specimen of thickness a/2 having phase velocity c contains embedded material of thickness d/2 having phase velocity c_d. The phase factor $ka = \omega_s a / c$ in Eqs.(17)-(21) must then be replaced by ka - ψ where

$$\psi = \omega_s d \left(\frac{1}{c} - \frac{1}{c_d} \right) = \omega_s d \frac{\Delta c}{c_d c} = kd \frac{\Delta c}{c_d} \tag{22}$$

and Δc = c_d – c. We thus set x = a/2 and re-write Eqs.(17), (20), and (21) as

$$A'(t) = e^{-\alpha a/2} B' \cos[\omega_s t - \frac{(ka - \psi)}{2} - \phi'] \tag{23}$$

where

$$\phi' = \tan^{-1} \frac{\sin(ka - \psi)}{e^{\alpha a} - \cos(ka - \psi)}, \tag{24}$$

and

$$B' = [1 + e^{-2\alpha a} - 2e^{-\alpha a}\cos(ka - \psi)]^{-1/2}\,. \tag{25}$$

We have assumed in obtaining the above equations that the change in the attenuation coefficient resulting from the embedded material is negligible.

For small ψ we may expand Eq. (23) in a power series about ψ = 0. Keeping only terms to first order, we obtain

$$\phi' = \phi + \Delta\phi \tag{26}$$

where

$$\Delta\phi = -\psi\left[\frac{e^{\alpha a}\cos ka - 1}{(e^{\alpha a} - \cos ka)^2 + \sin^2 ka}\right]. \tag{27}$$

Eq.(22) is thus approximated as

$$A'(t) = e^{-\alpha a/2}B'\cos(\omega_s t - \frac{ka}{2} - \phi + \frac{\psi}{2} - \Delta\phi) = e^{-\alpha a/2}B'\cos(\omega_s t + \theta) \tag{28}$$

where

$$\theta = -(\chi + \Delta\chi) = -\left(\frac{ka}{2} + \phi - \frac{\psi}{2} + \Delta\phi\right), \tag{29}$$

$$\chi = \frac{ka}{2} + \phi \tag{30}$$

and

$$\Delta\chi = -\frac{\psi}{2} + \Delta\phi = -\psi\left[\frac{1}{2} + \frac{e^{\alpha a}\cos ka - 1}{(e^{\alpha a} - \cos ka)^2 + \sin^2 ka}\right]. \tag{31}$$

Equation (28) reveals that the total phase contribution at x = a/2 is θ and from Eqs.(29) and (31) that the phase variation resulting from embedded material is $-\Delta\chi$.

The fractional change in the Young modulus $\Delta E/E$ is related to the fractional change in the ultrasonic longitudinal velocity $\Delta c/c$ as $\Delta E/E \approx \Delta C_{11}/C_{11} = (2\Delta c/c) + (\Delta\rho/\rho)$ where ρ is the mass density of the sample and C_{11} is the Brugger longitudinal elastic constant. Assuming that the fractional change in the mass density is small compared to the fractional change in the wave velocity, we may estimate the relationship between $\Delta E/E$ and $\Delta c/c$ as $\Delta E/E \approx 2\Delta c/c$. This relationship may be used to express ψ, given in Eq.(22) in terms of $\Delta c/c_d = (c/c_d)(\Delta c/c)$, in terms of $\Delta E/E$.

2.3 Solution to the general dynamical equations

We solve the coupled nonlinear Eqs. (14) and (15) for the steady-state solution by writing the coupled equations in matrix form and using an iteration procedure commonly employed in the physics literature (Schiff, 1968) to solve the matrix expression. The first iteration involves solving the equations for which the nonlinear terms are neglected. The second iteration is obtained by substituting the first iterative solution into the nonlinear terms of Eqs.(14) and

(15) and solving the resulting equations. The procedure provides solutions both for the cantilever tip and the sample surface displacements. Since the procedure is much too lengthy to reproduce here in full detail, only the salient features of the procedure leading to the steady state solution for the cantilever displacement $\eta_c = \sum Y_n \eta_{cn}$ are given. We begin by writing

$$\eta_{cn} = \varepsilon_{cn} + \xi_{cn} + \zeta_{cn} \tag{32}$$

and

$$\eta_{sn} = \varepsilon_{sn} + \xi_{sn} + \zeta_{sn} \tag{33}$$

where ε_{cn} and ξ_{cn} represent the first iteration (i.e. linear) static and oscillatory solutions, respectively, for the nth mode cantilever displacement, ζ_{cn} represents the second iteration (i.e., nonlinear) solution for the nth mode cantilever displacement, and ε_{sn} , ξ_{sn} , and ζ_{sn} are the corresponding first and second iteration nth mode displacements for the sample surface.

We note that for the range of frequencies generally employed in A-AFM the contribution from terms in the solution set involving the mass of the sample element m_s is small compared to the remaining terms and may to an excellent approximation be neglected. We thus neglect the terms involving m_s in the following equations.

2.3.1 First iterative solution

The first iterative solution is obtained by linearizing Eqs.(14) and (15), writing the resulting expression in matrix form, and solving the matrix expression assuming sinusoidal driving terms $F_c e^{i\omega_c t}$ and $F_s e^{i\omega_s t}$ for the cantilever and sample surface, respectively. The first iteration yields a static solution ε_{cn} and an oscillatory solution ξ_{cn} for the cantilever. The static solution is given by

$$\varepsilon_{cn} = \frac{k_s F(z_o)}{k_{cn}k_s + F'(z_o)(k_{cn} + k_s)} \, . \tag{34}$$

The first iterative oscillatory solution is given by

$$\xi_{cn} = Q_{cc}\cos(\omega_c t + \alpha_{cc} - \phi_{cc}) + Q_{cs}\cos(\omega_s t - \phi_{ss} + \theta) \tag{35}$$

where

$$\phi_{cc} \approx \tan^{-1}\frac{(\gamma_c k_s + \gamma_s k_{cn})\omega_c - \gamma_s m_c \omega_c^3 + F'(z_0)(\gamma_c + \gamma_s)\omega_c}{k_{cn}k_s - (m_c k_s + \gamma_c\gamma_s)\omega_c^2 + F'(z_0)(k_{cn} + k_s - m_c\omega_c^2)}, \tag{36}$$

$$\phi_{ss} \approx \tan^{-1}\frac{(\gamma_c k_s + \gamma_s k_{cn})\omega_s - \gamma_s m_c \omega_s^3 + F'(z_0)(\gamma_c + \gamma_s)\omega_s}{k_{cn}k_s - (m_c k_s + \gamma_c\gamma_s)\omega_s^2 + F'(z_0)(k_{cn} + k_s - m_c\omega_s^2)}, \tag{37}$$

$$Q_{cc} \approx F_c\{[k_s + F'(z_o)]^2 + \gamma_s^2\omega_c^2\}^{1/2}\{[k_{cn}k_s - \omega_c^2(m_c k_s + \gamma_c\gamma_s)$$

$$+F'(z_o)(k_{cn} + k_s - m_c\omega_c^2)]^2 + [\omega_c(\gamma_s k_{cn} + \gamma_c k_s) \tag{38}$$

$$-\omega_c^3\gamma_s m_c + F'(z_o)\omega_c(\gamma_s+\gamma_c)]^2\}^{-1/2},$$

and

$$Q_{cs} \approx -F_s F'(z_o)\{[k_{cn}k_s - \omega_s^2(m_c k_s + \gamma_c\gamma_s) + F'(z_o)(k_{cn}+k_s-m_c\omega_s^2)]^2 \tag{39}$$

$$+[\omega_s(\gamma_s k_{cn}+\gamma_c k_s) - \omega_s^3\gamma_s m_c + F'(z_o)\omega_s(\gamma_s+\gamma_c)]^2\}^{-1/2}.$$

2.3.2 Second iterative solution

The second iterative solution ζ_{cn} for each mode n of the cantilever is considerably more complicated, since it contains not only sum-frequency, difference-frequency, and generated harmonic-frequency components, but linear and static components as well. The second iterative solution ζ_{cn} is thus written as

$$\zeta_{cn} = \zeta_{cn,stat} + \zeta_{cn,lin} + \zeta_{cn,diff} + \zeta_{cn,sum} + \zeta_{cn,harm} \tag{40}$$

where $\zeta_{cn,stat}$ is a static or "dc" contribution generated by the nonlinear tip-surface interaction, $\zeta_{cn,lin}$ is a generated linear oscillatory contribution, $\zeta_{cn,diff}$ is a generated difference-frequency contribution resulting from the nonlinear mixing of the cantilever and sample oscillations, $\zeta_{cn,sum}$ is a generated sum-frequency contribution resulting from the nonlinear mixing of the cantilever and sample oscillations, and $\zeta_{cn,harm}$ are generated harmonic contributions.

Generally, the cantilever responds with decreasing displacement amplitudes as the drive frequency is increased above the fundamental resonance (for some cantilevers the second resonance mode has the largest amplitude), even when driven at higher modal frequencies. Thus, acoustic-atomic force microscopy methods do not generally utilize harmonic or sum-frequency signals. For expediency, such signals from the second iteration will not be considered here. Only the static, linear, and difference-frequency terms from the second iteration solution are relevant to the most commonly used A-AFM modalities.

The static contribution generated by the nonlinear interaction force is obtained to be

$$\zeta_{cn,stat} = \frac{1}{4}\frac{k_s F''(z_o)}{[k_{cn}k_s + F'(z_o)(k_{cn}+k_s)]}[2\varepsilon_o^2 + Q_{cc}^2 + Q_{cs}^2 + Q_{sc}^2 + Q_{ss}^2 \tag{41}$$

$$+2Q_{cc}Q_{sc}\cos(\alpha_{cc}-2\phi_{cc}) + 2Q_{cs}Q_{ss}\cos\alpha_{ss}]$$

where

$$\varepsilon_o = \frac{(k_{cn}+k_s)F(z_o)}{k_{cn}k_s + F'(z_o)(k_{cn}+k_s)}, \tag{42}$$

$$Q_{sc} \approx -F_c F'(z_o)\{[k_{cn}k_s - \omega_c^2(m_c k_s + \gamma_c\gamma_s) + F'(z_o)(k_{cn}+k_s-m_c\omega_c^2)]^2 \tag{43}$$

$$+[\omega_c(\gamma_s k_{cn}+\gamma_c k_s) - \omega_c^3\gamma_s m_c + F'(z_o)\omega_c(\gamma_s+\gamma_c)]^2\}^{-1/2},$$

$$Q_{ss} \approx F_s\{[k_s + F'(z_o)]^2 + \gamma_s^2\omega_s^2\}^{1/2}\{[k_{cn}k_s - \omega_s^2(m_c k_s + \gamma_c\gamma_s) \tag{44}$$

$$+F'(z_o)(k_{cn}+k_s-m_c\omega_s^2)]^2+[\omega_s(\gamma_s k_{cn}+\gamma_c k_s)$$

$$-\omega_s^3\gamma_s m_c+F'(z_o)\omega_c(\gamma_s+\gamma_c)]^2\}^{-1/2},$$

$$\alpha_{cc}=\tan^{-1}\frac{\gamma_s\omega_c}{k_s+F'(z_o)}, \tag{45}$$

$$\alpha_{ss}=\tan^{-1}\frac{\gamma_c\omega_s}{k_{cn}+F'(z_o)-m_c\omega_s^2} \tag{46}$$

and ϕ_{cc} is given by Eq. (36), Q_{cc} by Eq. (38) and Q_{cs} by Eq.(39).

The linear oscillatory contribution $\zeta_{cn,lin}$ generated by the nonlinear interaction force in the second iteration is obtained to be

$$\zeta_{cn,lin}=\frac{D_c}{R_{cc}}\varepsilon_o F''(z_o)[Q_{cc}^2+Q_{sc}^2+2Q_{cc}Q_{sc}\cos\alpha_{cc}]^{1/2}\cos(\omega_c t-2\phi_{cc}+\beta_c+\mu_{cc}) \tag{47}$$

$$+\frac{D_s}{R_{ss}}\varepsilon_o F''(z_o)[Q_{ss}^2+Q_{cs}^2+2Q_{ss}Q_{cs}\cos\alpha_{ss}]^{1/2}\cos(\omega_s t-2\phi_{ss}+\beta_s+\mu_{ss}+\theta)$$

where

$$\mu_{cc}=\tan^{-1}\frac{Q_{cc}\sin\alpha_{cc}}{Q_{cc}\cos\alpha_{cc}+Q_{sc}}, \tag{48}$$

$$\mu_{ss}=\tan^{-1}\frac{Q_{ss}\sin\alpha_{ss}}{Q_{ss}\cos\alpha_{ss}+Q_{cs}}, \tag{49}$$

$$\beta_c=\tan^{-1}\frac{\gamma_s\omega_c}{k_s}, \tag{50}$$

$$\beta_s=\tan^{-1}\frac{\gamma_s\omega_s}{k_s}, \tag{51}$$

$$D_c=[k_s^2+\gamma_s^2\omega_c^2]^{1/2}, \tag{52}$$

$$D_s=[k_s^2+\gamma_s^2\omega_s^2]^{1/2}, \tag{53}$$

$$R_{ss}=\{[k_{cn}k_s-\omega_s^2(m_c k_s+\gamma_c\gamma_s)+F'(z_o)(k_{cn}+k_s-m_c\omega_s^2)]^2 \tag{54}$$

$$+[\omega_s(\gamma_s k_{cn}+\gamma_c k_s)-\gamma_s m_c\omega_s^3+F'(z_o)\omega_s(\gamma_s+\gamma_c)]^2\}^{1/2},$$

and

$$R_{cc}=\{[k_{cn}k_s-\omega_c^2(m_c k_s+\gamma_c\gamma_s)+F'(z_o)(k_{cn}+k_s-m_c\omega_c^2)]^2 \tag{55}$$

$$+[\omega_c(\gamma_s k_{cn} + \gamma_c k_s) - \gamma_s m_c \omega_c^3 + F'(z_o)\omega_c(\gamma_s + \gamma_c)]^2\}^{1/2}.$$

The difference-frequency contribution $\zeta_{cn,diff}$ generated by the nonlinear interaction force in the second iteration is obtained to be

$$\zeta_{cn,diff} = G_n \cos[(\omega_c - \omega_s)t - \phi_{cc} + \phi_{ss} + \beta_{cs} - \phi_{cs} + \Gamma - \theta] \tag{56}$$

where

$$G_n = \frac{1}{2}\frac{D_{cs}}{R_{cs}} F''(z_o)\{Q_{cc}^2 Q_{cs}^2 + Q_{sc}^2 Q_{ss}^2 + Q_{cc}^2 Q_{ss}^2 + Q_{cs}^2 Q_{sc}^2 \tag{57}$$

$$+ 2Q_{cc}Q_{cs}Q_{sc}Q_{ss}\cos(\alpha_{cc} + \alpha_{ss}) + 2Q_{cc}^2 Q_{cs}Q_{ss}\cos\alpha_{ss} + 2Q_{cc}Q_{cs}^2 Q_{sc}\cos\alpha_{cc}$$

$$+ 2Q_{sc}^2 Q_{ss}Q_{cs}\cos\alpha_{ss} + 2Q_{cc}Q_{ss}Q_{cs}Q_{sc}\cos(\alpha_{cc} - \alpha_{ss})\}^{1/2},$$

$$D_{cs} = \sqrt{k_s^2 + \gamma_s^2(\omega_c - \omega_s)^2}\ , \tag{58}$$

$$R_{cs} = \sqrt{R_{cs1}^2 + R_{cs2}^2}\ , \tag{59}$$

$$R_{cs1} = k_{cn}k_s - m_c k_s(\omega_c - \omega_s)^2 - \gamma_c\gamma_s(\omega_c - \omega_s)^2 + F'(z_o)[k_{cn} + k_s - m_c(\omega_c - \omega_s)^2], \tag{60}$$

$$R_{cs2} = (\omega_c - \omega_s)(\gamma_s k_c + \gamma_c k_s) - \gamma_s m_c(\omega_c - \omega_s)^3 + F'(z_o)(\omega_c - \omega_s)(\gamma_s + \gamma_c), \tag{61}$$

$$\phi_{cs} = \tan^{-1}\frac{R_{cs2}}{R_{cs1}} \tag{62}$$

$$\approx \tan^{-1}\frac{(\gamma_c k_s + \gamma_s k_{cn})(\omega_c - \omega_s) - \gamma_s m_c(\omega_c - \omega_s)^3 + F'(z_0)(\gamma_c + \gamma_s)(\omega_c - \omega_s)}{k_{cn}k_s - (m_c k_s + \gamma_c\gamma_s)(\omega_c - \omega_s)^2 + F'(z_0)[k_{cn} + k_s - m_c(\omega_c - \omega_s)^2]},$$

$$\beta_{cs} = \tan^{-1}\frac{\gamma_s(\omega_c - \omega_s)}{k_s}, \tag{63}$$

and

$$\Gamma = \tan^{-1}\frac{Q_{cc}Q_{cs}\sin\alpha_{cc} - Q_{sc}Q_{ss}\sin\alpha_{ss} + Q_{cc}Q_{ss}\sin(\alpha_{cc} - \alpha_{ss})}{Q_{cc}Q_{cs}\cos\alpha_{cc} + Q_{sc}Q_{ss}\cos\alpha_{ss} + Q_{cc}Q_{ss}\cos(\alpha_{cc} - \alpha_{ss}) + Q_{cs}Q_{sc}}. \tag{64}$$

The phase term Γ given by Eq.(64) is quite complicated. However, advantage can be taken of the fact that k_s is generally large compared to other terms in the numerators of Q_{cc}, Q_{ss}, Q_{cs}, and Q_{sc}; the denominators of these terms are very roughly all equal. Hence, the magnitudes of Q_{cc} and Q_{ss} are usually large compared to those of Q_{cs} and Q_{sc}. The terms involving the product $Q_{cc}Q_{ss}$ thus dominate in Eq. (64) and we may approximate Γ as

$$\Gamma \approx \alpha_{cc} - \alpha_{ss} = \tan^{-1}\frac{\gamma_s \omega_c}{k_s + F'(z_0)} - \tan^{-1}\frac{\gamma_c \omega_s}{k_{cn} + F'(z_0) - m_c \omega_s^2} \tag{65}$$

where α_{cc} and α_{ss} are obtained from Eqs. (45) and (46), respectively. To the same extent that Γ may be approximated by Eq.(65) we may also approximate G_n, given by Eq. (57), as

$$G_n \approx \frac{F''(z_0)}{2}\frac{D_{cs}}{R_{cs}} Q_{cc} Q_{ss} \,. \tag{66}$$

2.3.3 Important features of the solution set

The present derivation is based on the assumption that the cantilever tip-sample surface interaction force is a multiply differentiable, nonlinear function of the tip-surface separation distance as indicated in Fig. 2. Points on the curve below the separation distance z_A in Fig. 2 correspond to a repulsive interaction force, while points above z_A in Fig. 2 correspond to an attractive interaction force. The force-separation curve has a minimum at a separation distance z_B corresponding to the maximum nonlinearity of the curve and that point lies in the attractive force portion of the curve. Cantilever oscillations result in continuous oscillatory changes in the tip-surface separation distance about the quiescent tip-surface separation distance z_0 (see Fig. 4). Since the cantilever oscillations are constrained to follow the force-separation curve, the fractions of the cantilever oscillation cycle in the repulsive and attractive portions of the force-separation curve depend on the quiescent tip-surface separation distance and the amplitude of the oscillations.

The cantilever oscillations are known to be bi-stable with the particular mode of oscillation being determined by the initial conditions that includes the tip-surface separation distance (Garcia & Perez, 2002). Unless some extraneous perturbation changes the mode of oscillation, the cantilever continues to oscillate in a given bi-stable mode for a given set of initial conditions. For large oscillation amplitudes the bi-stability coalesces to a single stable mode. In the present model the bi-stable mode of cantilever oscillation is set by the value of the "effective" sample stiffness constant k_s that has one of two values – one associated with the dominantly repulsive portion of the force-separation curve and one associated with the dominantly attractive portion (see Section 4.3). The value of the "effective" sample stiffness constant, hence cantilever oscillation mode, must be determined experimentally in the present model.

The total static solution to the coupled nonlinear equations (14) and (15) for the cantilever $\eta_{cn,stat}$ is the sum of the contribution ε_{cn}, given by Eq. (34), from the first iterative solution and the contribution $\zeta_{cn,stat}$, given by Eq. (41), from the second iteration as

$$\eta_{cn,stat} = \varepsilon_{cn} + \zeta_{cn,stat} \,. \tag{67}$$

The total linear solution $\eta_{cn,lin}$ to Eqs. (14) and (15) is the sum of the contribution ξ_{cn} given by Eq. (35) and the contribution $\zeta_{cn,lin}$ given by Eq. (47) as

$$\eta_{cn,lin} = \xi_{cn} + \zeta_{cn,lin} \,. \tag{68}$$

The total difference-frequency solution $\eta_{cn,diff}$ to Eqs. (14) and (15) is simply the contribution $\zeta_{cn,diff}$ given by Eq. (56).

It is interesting to note that ε_{cn} and the component ε_o in $\eta_{cn,stat}$ do not explicitly involve the cantilever drive amplitude F_c and the sample surface drive amplitude F_s, although other terms involving the Q factors, given by Eqs. (38), (39), (43), and (44), in $\zeta_{cn,stat}$ do involve these drive amplitudes. This means that only the contributions stemming from the nonlinearity in the cantilever tip-sample surface interaction force respond directly to variations in the drive amplitudes and in particular to the physical features of the material giving rise to variations in F_s. Further, the magnitudes of all second iteration (i.e. nonlinear) contributions, $\zeta_{cn,stat}$, $\zeta_{cn,lin}$, and $\zeta_{cn,diff}$ are strongly dependent on the cantilever tip-sample surface quiescent separation z_o, since the value of the nonlinear stiffness constant $F''(z_o)$ that dominates these contributions is highly sensitive to z_o. Indeed, $F''(z_o)$ attains a maximum value near the bottom of the force-separation curve of Fig. 2.

For large deflections of the cantilever that may occur for sufficiently hard contact, large bending moments may be introduced that produce frequency shifts in the cantilever resonance frequencies quite apart from those introduced by the interaction force stiffness constant $F'(z_0)$. For the assessment of $F'(z_0)$ near the bottom of the force-separation curve where the nonlinearity $F''(z_0)$ is maximum (maximum nonlinearity regime) and $F'(z_0)$ is relatively small, the bending moments are generally negligible and a reasonable estimate of $F'(z_0)$ can be obtained directly from differences in the engaged and non-engaged resonance (free space) frequencies of the cantilever.

For large driving force amplitudes, nonlinear modes of oscillation may be generated in the cantilever. Nonlinear tip-surface interactions are also known to excite nonlinear (anharmonic) cantilever modes (Stark & Heckl, 2003; Garcia & Perez, 2002). It is assumed that the nonlinear modes can be described in terms of a set of orthogonal eigenfunctions $Z_n(x)$ describing the nonlinearities of the unforced cantilever that are generally different from but orthogonal to $Y_n(x)$. In such case the nonlinear vibrational characteristics of the cantilever may also be included in the general cantilever response in a manner similar to that given above for the linear modes. The nonlinear modes are thus formally included in the present model by extending the set of eigenvalues k_{cn}, hence eigenvectors spanning the function space, to allow for nonlinear eigenmodes. This requires no additional formal analysis in the present model. All eigenvalues (including those from nonlinear modes) are ascertained in the present model from experimental measurements.

3. Signal generation for representative A-AFM modalities

Generally, there are two working modes in A-AFM - the contact mode and the non-contact mode. The contact mode is viewed as a modality for which the oscillating cantilever tip makes periodic contact with the sample surface irrespective of the distance of separation (setpont distance) between the non-oscillating (quiescent) cantilever tip and the sample surface. When the setpoint distance z_0 lies close to the sample surface, the cantilever operates near the dominantly repulsive portion of the cantilever tip-sample surface interaction force-separation curve and experiences a dominantly repulsive force over some appreciable fraction of an oscillation period (contact time). The oscillation amplitude is usually small for this contact mode of operation and the tip-surface interaction force may be approximated by a linear dependence of the tip-surface interaction force on the tip-surface separation distance. A-AFM modalities that operate in the contact mode include force modulation microscopy, atomic force acoustic microscopy, and a modality of amplitude modulation-atomic force microscopy (AM-AFM) that may be descriptively called 'small amplitude contact tapping mode.'

Various other A-AFM modalities operate in the non-contact mode where the cantilever tip-sample surface setpoint distance z_0 is sufficiently large that the cantilever tip, oscillating with small amplitude, does not contact with the sample surface. In such cases the modalities optimally operate in that portion of the force-separation curve that yields the maximum force-separation nonlinearity, appropriately called the 'maximum nonlinearity regime' of A-AFM operation. Ultrasonic force microscopy, heterodyne force microscopy, and resonant difference-frequency atomic force ultrasonic microscopy (RDF-AFUM) are examples of non-contact A-AFM modalities. Non-contact amplitude modulation-atomic force microscopy (noncontact tapping mode) also operates in this portion of the force-separation curve.

The equations derived in Section 2, describing the dynamical response of the cantilever resulting from the cantilever tip-sample surface interaction forces, have been used to quantify the signal generation and image contrast for all A-AFM modalities mentioned in the introduction (Cantrell & Cantrell, 2008). We consider here, however, only resonant difference-frequency atomic force ultrasonic microscopy (RDF-AFUM), and the commonly used amplitude modulation-atomic force microscopy (AM-AFM), a modality that includes the intermittent contact mode as well as contact and non-contact tapping modes. RDF-AFUM and AM-AFM represent opposite extremes in complexity, both in instrumentation and in the analytical expressions used to assess signal generation and image contrast.

RDF-AFUM uses input drive oscillations both to the cantilever and to the sample surface to interrogate the sample. It is the most complex of the A-AFM modalities and the assessment of signal generation and image contrast for RDF-AFUM requires application of the largest number of equations from Section 2. The AM-AFM modality uses only an input drive oscillation to the cantilever and is among the simplest of A-AFM modalities. The calculation of the AM-AFM output signal thus requires relatively few equations from Section 2. The AM-AFM modality may be viewed operationally and analytically as a subset of the RDF-AFUM modality.

3.1 Resonant difference-frequency atomic force ultrasonic microscopy

Resonant difference-frequency atomic force ultrasonic microscopy (RDF-AFUM) employs an ultrasonic wave launched from the bottom of a sample, while the AFM cantilever tip engages the sample top surface. The cantilever is driven at a frequency differing from the ultrasonic frequency by one of the resonance frequencies of the engaged cantilever. It is important to note that at high drive amplitudes of the ultrasonic wave or engaged cantilever (or both) the resonance frequency generating the difference-frequency signal may correspond to one of the nonlinear oscillation modes of the cantilever. The engaged cantilever resonance frequency for the (linear or nonlinear) mode n, neglecting dissipation, is given by $m_c\omega_{cn}^2 = k_{cn} + F'(z_0)k_{cn}[k_s + F'(z_0)]^{-1}$, where k_{cn} is the cantilever stiffness constant corresponding to the nth (linear or nonlinear) non-engaged (free space) resonance mode. Since $F'(z_0)$ may be positive or negative, depending on the shape of the force separation curve, at the separation distance z_0 corresponding to maximum $F''(z_0)$, the resonance frequency of the cantilever, when engaged at this value of z_0, may be larger or smaller, respectively, than the resonance frequency when not engaged. The nonlinear mixing of the oscillating cantilever and the ultrasonic wave in the region defined by the cantilever tip-sample surface interaction force generates difference-frequency oscillations at the engaged cantilever resonance. Variations in the amplitude and phase of the bulk wave due to the presence of subsurface nano/microstructures, as well as variations in near-surface material parameters, affect the amplitude and phase of the difference-frequency signal. These variations are used to create spatial mappings generated by subsurface and near-surface structures.

In RDF-AFUM the cantilever difference-frequency response is obtained from the nonlinear mixing in the region defined by the tip-surface interaction force. The interaction force varies nonlinearly with the tip-surface separation distance. The deflection of the cantilever obtained in calibration plots is related to this force. For small slopes of the deflection versus separation distance, the interaction force and cantilever deflection curves are approximately related via a constant of proportionality. The maximum difference-frequency signal amplitude occurs when the quiescent deflection of the cantilever is near the bottom of the force-separation curve (z_B in Fig. 2). There the maximum change in the slope of the force versus separation (hence maximum interaction force nonlinearity) occurs. We call this region of operation the maximum nonlinearity regime.

The dominant term or terms for the cantilever difference-frequency displacement in Eqs. (56) and (57) depend on the values of k_{cn} for the free modes of cantilever oscillation, the difference-frequency (ω_c – ω_s), and the value of $F'(z_0)$ obtained at the quiescent separation distance z_0 = $(z_0)B$ at which the maximum difference-frequency signal occurs. We designate the non-engaged linear or nonlinear mode n for which the difference-frequency engaged resonance occurs as n = p. The dominant difference-frequency component in Eqs.(56) and (57) is thus $\eta_{cp} = \eta_{cp,diff} = \zeta_{cp,diff}$ and is given by Eq.(56) for n = p as

$$\zeta_{cp,diff} = G_p \cos[(\omega_c - \omega_s)t - \phi_{cc} + \phi_{ss} + \beta_{cs} - \phi_{cs} + \Gamma - \theta] \tag{69}$$

where Gp is given by Eq.(57) and in approximation by Eq.(66). The phase terms in Eq.(69) are obtained from Eqs. (36), (37), (45), (46), and (62)-(64) where Γ may be approximated by Eq. (66).

It is important to point out in considering these equations that while the difference-frequency resonance frequency $(\omega_c - \omega_s)$ in RDF-AFUM is usually set to correspond to the lowest resonance mode of the engaged cantilever (although a higher modal resonance could be used), the cantilever driving frequency ω_c and ultrasonic frequency ω_s generally are set near (but not necessary equal to) higher resonance modes n = q and n = r, respectively, of the engaged cantilever. Thus, the cantilever stiffness constant k_{cn} is appropriately given as k_{cp} when involving the difference-frequency terms in Eqs. (36)-(39), (42)-(46), and (58)-(64), given as k_{cq} when involving the cantilever drive frequency ω_c at or near the frequency of the qth cantilever resonance mode, and given as k_{cr} when involving the ultrasonic frequency ω_s at or near the frequency of the rth cantilever resonance mode. If ω_c and ω_s are not set at or near a resonance modal frequency of the engaged cantilever, then it may be necessary to include more than one term in Eqs. (12) and (32) corresponding to various values of q and r.

It is seen from Eq. (57) that for a given value of $(\omega_c - \omega_s)$ the maximum value of $\zeta_{cp,diff}$ ideally occurs for a value of z_0 such that $F''(z_0)$ is maximized. It is important to note, however, that $F'(z_0)$, while relatively small in magnitude compared to that of the hard contact regime, is generally not equal to zero at that point. Strictly, the values of $F''(z_0)$ and $F'(z_0)$ for a given z_0 are each dependent on the exact functional form of $F(z_0)$. A functional form for $F(z_0)$ sufficiently quantitative to quantify $F''(z_0)$ and $F'(z_0)$ is not typically available. However, experimental curves for $F(z_0)$ can be obtained and compared to the experimental curves of $\zeta_{cp,diff}$ plotted as a function of z_0. An examination of Eq. (57) suggests that a more exact approach to maximizing $\zeta_{cp,diff}$ would be not only to vary z_0 but also to vary slightly the difference-frequency from the free space resonance condition until an optimal setting for both z_0 and the difference-frequency is achieved.

3.2 Amplitude modulation-atomic force microscopy

The amplitude modulation-atomic force microscopy (AM-AFM) mode (also called intermittent contact mode or tapping mode) is a standard feature on many atomic force microscopes for which the cantilever is driven in oscillation, but no surface oscillations resulting from bulk ultrasonic waves are generated (i.e., Fs and ω_s are zero). Thus, AM-AFM cannot be used to image subsurface features, but interesting surface properties and features can be imaged. Since AM-AFM can be used in both the hard contact and maximum nonlinearity regimes (i.e. the linear and maximally nonlinear regimes, respectively, of the force-separation curve), the cantilever displacement $\eta_{cn,lin}$ for mode n is given most generally as

$$\eta_{cn.lin} = \xi_{cn} + \zeta_{cn,lin} \tag{70}$$

where ξ_{cn} is given by Eq.(35) with the term involving Q_{cs} set equal to zero and $\zeta_{cn,lin}$ is given by Eq.(47) with all terms involving Q_{cs} and Q_{ss} set equal to zero.

3.2.1 Maximum nonlinearity regime

For the maximum nonlinearity regime the expression for $\eta_{cn,lin}$ is

$$\eta_{cn,lin} = H\cos(\omega_c t - \phi_{cc} + \Lambda) \tag{71}$$

where

$$\Lambda = \tan^{-1}\frac{\sin(\beta_c + \mu_{cc} - \phi_{cc} - \alpha_{cc})}{\cos(\beta_c + \mu_{cc} - \phi_{cc} - \alpha_{cc}) + (Q_{cc}/W)}, \tag{72}$$

$$W = \frac{D_c}{R_{cc}}\varepsilon_0 F''(z_0)(Q_{cc}^2 + Q_{sc}^2 + 2Q_{cc}Q_{sc}\cos\alpha_{cc})^{1/2}, \tag{73}$$

and

$$H = [Q_{cc}^2 + W^2 + 2Q_{cc}W\cos(\beta_c + \mu_{cc} - \phi_{cc} - \alpha_{cc})]^{1/2} \tag{74}$$

where Q_{cc} is given by Eq.(38), Q_{sc} by Eq.(43), ϕ_{cc} by Eq.(36), μ_{cc} by Eq.(48), ε_0 by Eq.(42); a_{cc}, β_c, D_c, and R_{cc}, are given by Eqs.(45), (50), (52), and (53), respectively.

3.2.2 Hard contact regime

The complexity of the cantilever response $\eta_{cn,lin}$ for AM-AFM is greatly reduced for the hard contact regime, where $F''(z_0)$ is negligibly small and $F'(z_0)$ is very large and negative. For sufficiently hard contact Λ and α_{cc} are approximately zero and we obtain from Eq. (71) that

$$\eta_{cn,lin} \approx Q_{cc} \cos(\omega_c t - \phi_{cc}) \tag{75}$$

where

$$Q_{cc} = F_c[(k_{cn} + k_s - m_c\omega_c^2)^2 + (\gamma_c + \gamma_s)^2 \omega_c^2]^{-1/2} \tag{76}$$

and

$$\phi_{cc} = \tan^{-1} \frac{(\gamma_c + \gamma_s)\omega_c}{k_{cn} + k_s - m_c\omega_c^2} . \tag{77}$$

The dependence of $\eta_{cn,lin}$ on the material damping coefficient γ_s and the sample stiffness constant k_s, both for the hard contact and the maximum nonlinearity regimes, means that AM-AFM can be used to assess the viscoelastic properties of the material irrespective of the regime of operation.

4. Image contrast for representative A-AFM modalities

All the above equations, except for Eqs.(26) - (31), were derived for constant values of the cantilever and material parameters. If, in an area scan of the sample, the parameters remain constant from point to point, the image generated from the scan would be flat and featureless. We consider here that the sample stiffness constant k_s may vary from point to point on the sample surface. Since k_s is dependent on the Young modulus E (see Section 4.3), this means that E also varies from point to point. We assume that the value of the sample stiffness constant k'_s at a given point on the surface differs from the value k_s at another position as $k'_s = k_s + \Delta k_s$. For any function $f(k_s)$ having a functional dependence on k_s, a variation in k_s generates a variation in $f(k_s)$ given by $\Delta f = (df / dk_s)_0 \Delta k_s$, where the subscripted zero indicates evaluation at k_s. A similar expression can be obtained for the material damping parameter γ_s, but we shall not consider such variations here.

A variation in k_s produces a variation in both amplitude and phase of the signal generated by the cantilever tip-sample surface interactions. The variations in amplitude and phase can be used to generate amplitude and phase images, respectively, in a surface scan of the sample. We consider here only images generated by the phase variations in the signal. The equations for amplitude-generated images are given elsewhere (Cantrell & Cantrell, 2008). The phase factors involved in RDF-AFUM are given from Eq.(69), (29), and (30) to be ϕ_{cc}, ϕ_{ss}, β_{cs}, ϕ_{cs}, Γ, and χ; the phase factors involved in the AM-AFM mode are, from Eq.(71), ϕ_{cc}, and Λ. Each of these phase factors is dependent on k_s and the variations in the phase factors resulting from variations in k_s are responsible for image generation when using phase detection of the A-AFM signal. The exact dependence of the phase on k_s, however, is different for hard contact and maximum nonlinearity regimes.

4.1 Resonant difference-frequency atomic force ultrasonic microscopy

RDF-AFUM operates only in the maximum nonlinearity regime where the total variation in phase is given as ($\Delta\beta_{cs}$ + $\Delta\phi_{cc}$ + $\Delta\phi_{ss}$ – $\Delta\phi_{cs}$ +$\Delta\Gamma$ – $\Delta\chi$). The phase factors relevant to RDF-AFUM are given as

$$\Delta\beta_{cs} = \left(\frac{d\beta_{cs}}{dk_s}\right)_0 \Delta k_s = -\frac{\gamma_s \Delta\omega}{[k_s + F'(z_0)]^2 + \gamma_s^2(\Delta\omega)^2}\Delta k_s \tag{78}$$

and

$$\Delta\phi_{cc} = -\frac{A_{cc}}{B_{cc}}\Delta k_s \tag{79}$$

where

$$A_{cc} = [\gamma_s k_{cq}^2 + 2F'(z_0)\gamma_s k_{cq} + F'(z_0)^2(\gamma_c + \gamma_s)]\omega_c \tag{80}$$

$$+ [\gamma_c^2\gamma_s - 2\gamma_s m_c(k_{cq} + F'(z_0))]\omega_c^3 + m_c^2\gamma_s\omega_c^5$$

and

$$B_{cc} = \{[\gamma_c k_s + \gamma_s k_{cq} + F'(z_0)(\gamma_c + \gamma_s)]\omega_c - \gamma_s m_c \omega_c^3\}^2 \tag{81}$$

$$+ \{[k_{cq} - m_c\omega_c^2 + F'(z_0)]k_s + F'(z_0)(k_{cq} - m_c\omega_c^2) - \gamma_c\gamma_s\omega_c^2\}^2,$$

$$\Delta\phi_{ss} = -\frac{A_{ss}}{B_{ss}}\Delta k_s \tag{82}$$

where

$$A_{ss} = [\gamma_s k_{cr}^2 + 2F'(z_0)\gamma_s k_{cr} + F'(z_0)^2(\gamma_c + \gamma_s)]\omega_s \tag{83}$$

$$+ [\gamma_c^2\gamma_s - 2\gamma_s m_c(k_{cr} + F'(z_0))]\omega_s^3 + m_c^2\gamma_s\omega_s^5$$

and

$$B_{ss} = \{[\gamma_c k_s + \gamma_s k_{cr} + F'(z_0)(\gamma_c + \gamma_s)]\omega_s - \gamma_s m_c \omega_s^3\}^2 \tag{84}$$

$$+ \{[k_{cr} - m_c\omega_s^2 + F'(z_0)]k_s + F'(z_0)(k_{cr} - m_c\omega_s^2) - \gamma_c\gamma_s\omega_s^2\}^2,$$

and

$$\Delta\phi_{cs} = -\frac{A_{cs}}{B_{cs}}\Delta k_s \tag{85}$$

where

$$A_{cs} = [\gamma_s k_{cp}^2 + 2F'(z_0)\gamma_s k_{cp} + F'(z_0)^2(\gamma_c + \gamma_s)](\Delta\omega) \tag{86}$$

$$+[\gamma_c^2\gamma_s - 2\gamma_s m_c(k_{cp} + F'(z_0))](\Delta\omega)^3 + m_c^2\gamma_s(\Delta\omega)^5$$

and

$$B_{cs} = \{[\gamma_c k_s + \gamma_s k_{cp} + F'(z_0)(\gamma_c + \gamma_s)](\Delta\omega) - \gamma_s m_c(\Delta\omega)^3\}^2 \tag{87}$$

$$+\{[k_{cp} - m_c(\Delta\omega)^2 + F'(z_0)]k_s + F'(z_0)[k_{cp} - m_c(\Delta\omega)^2] - \gamma_c\gamma_s(\Delta\omega)^2\}^2 .$$

To the extent that $\Gamma = \alpha_{cc} - \alpha_{ss}$, as given by Eq.(65), we may write

$$\Delta\Gamma = \Delta\alpha_{cc} = -\frac{\gamma_s\omega_c}{[k_s + F'(z_0)]^2 + \gamma_s^2\omega_c^2}\Delta k_s . \tag{88}$$

The phase term $\Delta\chi$ is given by Eqs. (22) and (31).

4.2 Amplitude modulation-atomic force microscopy

The appropriate variations in the phase factors relevant to the AM-AFM or tapping mode maximum nonlinearity regime are $\Delta\alpha_{cc}$, $\Delta\phi_{cc}$, and $\Delta\Lambda$. The factor $\Delta\Lambda$ is obtained from Eq.(72) as

$$\Delta\Lambda = \frac{1 + (Q_{cc}/W)\cos(\beta_c + \mu_{cc} - \phi_{cc} - \alpha_{cc})}{[\cos(\beta_c + \mu_{cc} - \phi_{cc} - \alpha_{cc}) + (Q_{cc}/W)]^2 + \sin^2(\beta_c + \mu_{cc} - \phi_{cc} - \alpha_{cc})} \tag{89}$$

$$\times(\Delta\beta_c + \Delta\mu_{cc} - \Delta\phi_{cc} - \Delta\alpha_{cc})$$

where

$$\Delta\beta_c = -\frac{\gamma_s\omega_c}{k_s^2 + \gamma_s^2\omega_c^2}\Delta k_s , \tag{90}$$

$\Delta\phi_{cc}$ is given by Eq. (79), and $\Delta\mu_{cc}$ is obtained from Eq.(48). To the extend that Q_{sc} is much smaller than Q_{cc}, we get from Eq. (48) that $\Delta\mu_{cc} = \Delta\alpha_{cc}$ where $\Delta\alpha_{cc}$ is given by Eq. (88).
For the hard contact regime where $F'(z_0)$ is very large and negative, the relevant phase variation is obtained from Eq. (77) as

$$\Delta\phi_{cc} = -\frac{(\gamma_c + \gamma_s)\omega_c}{(k_s + k_{cq} - m_c\omega_c^2)^2 + (\gamma_c + \gamma_s)^2\omega_c^2}\Delta k_s . \tag{91}$$

As a word of caution, the extent to which the hard contact equation applies depends on how well the approximation $F'(z_0) \to -\infty$ holds. In those cases where such an assumption is suspect, the equations for the maximum nonlinearity regime should be used.

4.3 Dependence on the Young modulus

Hertzian contact theory provides that the sample stiffness constant k_s is related to the Young modulus E of the sample as (Yaralioglu et al., 2000)

$$k_S = 2r_C\left(\frac{1-\upsilon_T^2}{E_T} + \frac{1-\upsilon^2}{E}\right)^{-1} \tag{92}$$

where υ is the Poisson ratio of the sample material, E_T and υ_T are the Young modulus and Poisson ratio, respectively, of the cantilever tip, and r_C is the cantilever tip-sample surface contact radius. Hence,

$$\Delta k_S = \frac{2r_C\left(1-\upsilon^2\right)}{E^2}\left(\frac{1-\upsilon_T^2}{E_T} + \frac{1-\upsilon^2}{E}\right)^{-2}\Delta E = \frac{k_S}{E}(1-\nu^2)\left(\frac{1-\upsilon_T^2}{E_T} + \frac{1-\upsilon^2}{E}\right)^{-1}\frac{\Delta E}{E}\ . \tag{93}$$

Strictly, Eq. (93) was derived for the case of repulsive interaction forces leading to a concave elastic deformation of a flat sample surface from a contacting hard spherical object. However, we consider here that to a crude approximation Eqs. (92) and (93) also hold for attractive interactive forces providing that the elastic deformation of the sample surface is viewed as a convex deformation (asperity) subtending an effective contact radius r_c with the cantilever tip that is appropriately different in magnitude from that of the repulsive force case. As pointed out in Section 2.3.3, the cantilever oscillations are known to be bi-stable with the particular mode of oscillation being determined by the initial conditions that includes the tip-surface separation distance. In the present model the bi-stable mode of cantilever oscillation is set by the value of the "effective" sample stiffness constant k_s corresponding either to the dominantly repulsive region or dominantly attractive region of the force-separation curve.

Eq. (93) can be used with Eqs. (78)-(91) to ascertain the fractional variation in the Young modulus $\Delta E / E$ from measurements of the phase variation in the signal from an appropriate A-AFM modality. For the case where $E_T >> E$, e.g. for polymeric or soft biological materials, Eq. (92) reduces to $k_s = 2r_cE$ and Eq. (93) reduces to $\Delta k_s = k_s(\Delta E/E)$.

5. Assessment of model validity

We assess the validity of the above analytical model by comparing variations in the Young modulus of a specimen as calculated from the model with independent experimental measurements of the same specimen material. The choice of material is influenced by a recent focus to develop high performance polymers having low density, high strength, optical transparency, and high radiation resistance for a variety of applications in hostile space environments. One such polymer is LaRC™-CP2 polyimide. We consider here the application of RDF-AFUM and AM-AFM to assess variations in the Young modulus of nancomposites composed of nanoparticles embedded in a LaRC™-CP2 polyimide matrix. We consider two nanocomposites – one embedded with gold nanoparticles and the other embedded with single wall carbon (SWCNT) nanotube bundles.

We first consider a specimen of LaRC™-CP2 polyimide polymer roughly 12.7 μm thick containing a monolayer of randomly distributed gold particles, roughly 10-15 nm in diameter and embedded roughly 7 μm beneath the specimen surface. Fig. 5a is an AM-AFM

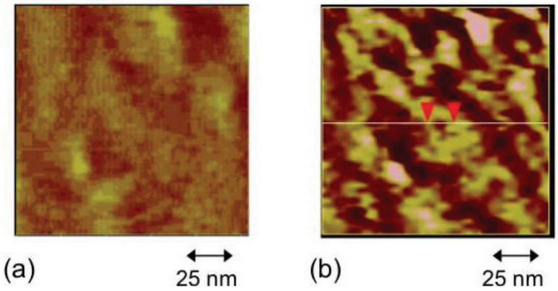


Fig. 5. Micrographs of LaRC™-CP2 polyimide polymer embedded with gold nanoparticles. (a) Noncontact tapping mode (AM-AFM) phase-generated micrograph. (b) RDF-AFUM phase-generated image over the same scan area as (a). (from Cantrell et al., 2007)

phase-generated image obtained in the maximum nonlinearity regime (noncontact tapping mode). A commercial cantilever having a stiffness constant of 14 N m^{-1}, a lowest-mode resonance frequency of 302 kHz, and a cantilever damping coefficient of roughly 10^{-8} kg s^{-1} is driven at 2.1 MHz to obtain the micrograph of Fig.5a (Cantrell et al., 2007). The values of the relevant model parameters for LaRC™-CP2 polyimide polymer are 1.4 x 10^3 kg m^{-3} for the mass density ρ, 2.4 GPa for the Young modulus E, 0.37 for the Poisson ratio υ, k_s = 96.1 N m^{-1}, and γ_s = 4.8 x 10^{-5} kg s^{-1} (Park et al., 2002; Fay et al., 1999; Cantrell et al. 2007). Since no bulk ultrasonic wave is involved, the image contrast results only from variations in the specimen near-surface sample stiffness constant k_s. The darker areas in the image correspond to larger values of the sample stiffness constant, hence Young modulus, relative to that of the brighter areas. The maximum phase difference between the bright and dark areas in the image is approximately 1.5 degrees. Using the value 1.5 degrees, we obtain from the model that the variation in the Young modulus ΔE/E ≈ 18%. This value is consistent with the value ΔE/E ≈ 21%obtained from independent mechanical stretching experiments on pure LaRC™-CP2 polymer sheets (Fay et al., 1999).

An RDF-AFUM phase image of the same scan area as that of Fig. 5a is shown in Fig. 5b. The RDF-AFUM image reveals bright and dark regions over the scan area that broadly correspond to the bright and dark regions in the surface image of Fig. 5a, although the image contrast and local detail appears to differ in the two images. F′(z) is assessed to be roughly –53 N m^{-1} at the tip-surface separation corresponding to the maximum difference-frequency signal. The acoustic wave has a frequency of 1.8 MHz. The maximum variation in phase shown in Fig.5b is approximately 13.2 degrees. Using the value 13.2 degrees, we obtain from the model that the variation in the Young modulus ΔE/E ≈ 24%. This value is also consistent with the value ΔE/E ≈ 21%obtained from independent mechanical stretching experiments on pure LaRC™-CP2 polymer sheets.

The existence of contiguous material with differing elastic constants suggests that the LaRC™-CP2 material is not homogeneous. The broad coincidence of dark (bright) regions in the images of Fig.5a and 5b suggests that the polymer structure giving rise to a larger (smaller) elastic modulus in the bulk material occurs in varying amounts through the bulk to the surface, the degree of darkness (brightness) in Fig. 5b being somewhat reflective of the structural homogeneity of the material along the propagation path of the ultrasonic wave. It is assumed that the appearance of contiguous material with different elastic coefficients may result from the growth of a strain-nucleated harder material phase resulting from the difference in the coefficients of thermal expansion between the polymer matrix and the embedded gold particles.

To test the assumption of a strain-nucleated harder phase, micrographs were obtained of a specimen formed from bundles of single-wall carbon nanotubes (SWCNTs) distributed randomly through the bulk of a 50μm-thick film of LaRC™-CP2 polymer. Figure 6a shows a conventional atomic force microscope (AFM) topographical image of the specimen showing only surface features. A RDF-AFUM phase-image of the specimen, taken in the same scan area as that of Fig 6a, is shown in Fig. 6b. Comparison of the two images reveals the appearance of subsurface bundles of SWCNTs (dark contrast filamentary features) lying in the plane of the RDF-AFUM image that do not appear in the AFM topographical scan. Dramatic variations from dark to bright to slightly bright contrast occur in image plane along portions of the boundary between the bundles of SWCNTs and the matrix material. The variations follow the contour of the nanotube bundles and suggest the occurrence of an interphase region (bright contrast feature) at the nanotube bundle-polymer interface. The interphase consists of polymer material having dramatically different mechanical properties from that of the matrix material. We note, however, that aside from the local interphase regions in Fig. 6b there are no broad, contiguous regions of material with differing elastic constants as observed in Fig. 5. Since the difference between the coefficients of thermal expansion of LaRC™-CP2 polymer and SWCNT bundles is considerable less than that for LaRC™-CP2 polymer and gold particles, we infer that the thermal strains in SWCNT bundle-embedded polymer material are not sufficiently large to generate the larger contiguous features observed in material embedded with gold particles.

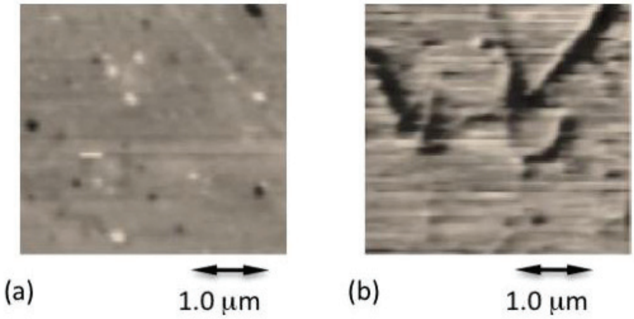


Fig. 6. Micrographs of LaRC™-CP2 polyimide polymer embedded with single wall carbon nanotube bundles. (a) AFM topographical image. (b) RDF-AFUM phase-generated image over the same scan area as (a).

6. Conclusion

The various dynamical implementations of the atomic force microscope have become important nanoscale characterization tools for the development of novel materials and devices. One of the most significant factors affecting all dynamical AFM modalities is the cantilever tip-sample surface interaction force. We have developed a detailed mathematical model of this interaction that includes a quantitative consideration of the nonlinearity of the interaction force as a function of the cantilever tip-sample surface separation distance. The model makes full use of cantilever beam dynamics and the multiply differentiability of the continuous force-separation curve that results in a set of coupled differential equations, Eqs.(14) and (15), for the displacement amplitudes of both the cantilever and the sample surface. The coupled dynamical equations are recast in matrix form and solved by a standard iteration procedure, but space limitations allow only a presentation of the salient features of the procedure. Although the mathematical form of the coupled equations are valid for any vibrational mode, only flexural vibrations of the cantilever and out-of-plane oscillations of the sample surface are considered.

We emphasize that Eqs.(14) and (15) are obtained assuming that the cantilever is a rectangular beam of constant cross-section, the dynamics of which are characterized by a set of eigenfunctions that form an orthogonal basis for the solution set. For some other cantilever shape a different orthogonal basis set of eigenfunctions would be appropriate. However, the mathematical procedure used here would lead again to Eqs.(14) and (15) with values of the coefficients appropriate to the different cantilever geometry. Practicably, this means that the shape of the cantilever is not as important in the solution set as knowing the cantilever modal resonant frequencies, obtained experimentally. The modal frequencies and solution set are expanded to include nonlinear modes generated by nonlinear interaction forces or large cantilever drive amplitudes.

A general steady state solution of the coupled dynamical equations is found that accounts for the positions of the excitation force (e.g., a piezo-transducer) and the cantilever tip along the length of the cantilever and for the position of the laser probe on the cantilever surface. The solution is applied to two dynamical AFM modalities - resonant difference-frequency atomic force ultrasonic microscopy, and the commonly used amplitude modulation-atomic force microscopy. Image generation and contrast equations are obtained for each of the two A-AFM modalities assuming for expediency that the contrast results only from variations in the sample stiffness constant. Since the sample stiffness constant is related directly to the Young modulus of the sample, the contrast can be expressed in terms of the variation in the Young modulus from point to point as the sample is scanned. We note further the existence of two values of the sample stiffness constant, corresponding to the dominantly attractive and dominantly repulsive regimes of the force-separation curve. The two values allow for a bi-stability in the cantilever oscillations that is experimentally observed.

Equations for both the maximum nonlinearity regime and the hard contact (linear) regime of cantilever engagement with the sample surface are obtained. For dynamical AFM operation outside these regimes, it is necessary to use all terms in the solution set given in Section 2 to describe the signal output of a given A-AFM modality. The extent to which the hard contact (linear regime) equations apply depends on how well the approximation $F'(z_0) \rightarrow -\infty$ holds.

In those cases where such an assumption is suspect, all terms in the equations for a given modality should be used.

In order to test the validity of the present model, comparative measurements of the fractional variation of the Young modulus ΔE/E in a film of LaRC™-CP2 polyimide polymer were obtained from phase-generated images obtained over the same scan area of the specimen using the RDF-AFUM and AM-AFM maximum nonlinearity modalities. The two modalities represent opposite extremes in measurement complexity, both in instrumentation and in the analytical expressions used to calculate ΔE/E. The values 24 percent calculated for RDF-AFUM and 18 percent calculated for the AM-AFM maximum nonlinearity mode are in remarkably close agreement for such disparate techniques. The agreement of both calculations with the value of 21 percent obtained from independent mechanical stretching experiments of LaRC™-CP2 polymer sheet material offers strong evidence for the validity of the present model.

The present model can also be used to quantify the image contrast from variations in the sample damping coefficient γ_S in the material. Space limitations prohibit the inclusion of such contrast mechanisms here, but the effects can be derived straightforwardly by the reader from the equations derived in Section 2. Although the present model is developed for flexural oscillations of the cantilever and out-of-plane vibrations of the sample surface, the model can be extended to include other modes of cantilever oscillation and sample surface response as well.

7. References

Binnig, G; Quate, C. F. & Gerber, Ch. (1986). Atomic force microscope. *Physical Review Letters*, 56, 930-933.

Bolef, D. I. & and J. G. Miller, J. G. (1971). High-frequency continuous wave ultrasonics. In: *Physical Acoustics, Vol. VIII*, W. P. Mason and R. N. Thurston, Ed., Academic, New York, 95-201.

Cantrell, J. H. (2004). Determination of absolute bond strength from hydroxyl groups at oxidized aluminum-epoxy interfaces by angle beam ultrasonic spectroscopy. *Journal of Applied Physics*, 96, 3775-3781.

Cantrell, S. A.; Cantrell, J. H. & Lillehei, P. T. (2007). Nanoscale subsurface imaging via resonant difference-frequency atomic force ultrasonic microscopy. *Journal of Applied Physics*, 101, 114324.

Cantrell, J.H. & Cantrell, S. A. (2008). Analytical model of the nonlinear dynamics of cantilever tip-sample surface interactions for various acoustic atomic force microscopies. *Physical Review B*, 77, 165409.

Chan, H. B.; Aksyuk, V. A.; Kleiman, R. N.; Bishop, D. J. & Capasso, F. (2001). Nonlinear micromechanical Casimir oscillator. *Physical Review Letters*, 97, 211801.

Cuberes, M. T.; Alexander, H. E.; Briggs, G. A. D. & and Kolosov, O. V. (2000). Heterodyne force microscopy of PMMA/rubber nanocomposites: nanomapping of viscoelastic response at ultrasonic frequencies. *Journal of Physics D: Applied Physics*, 33, 2347-2355.

Cuberes, M. T. (2009). Intermittent-contact heterodyne force microscopy. *Journal of Nanomaterials*, 2009, 762016.

Eguchi, T. & Hasegawa, Y. (2002). High resolution atomic force microscopic imaging of the Si(111)-(7x7) surface: contribution of short-range force to the images. *Physical Review Letters*, 89, 266105.

Fay, C. C.; Stoakley, D. M. & St. Clair, A. K. (1999). Molecularly oriented films for space applications. *High Performance Polymers*, 11, 145-156.

Garcia, R & Perez, R. (2002). Dynamic atomic force microscopy methods. *Surface Science Reports*, 47, 1-79.

Geer, R. E.; Kolosov, O. V.; Briggs, G. A. D. & Shekhawat, G. S. (2002). Nanometer-scale mechanical imaging of aluminum damascene interconnect structures in a low-dielectric-constant polymer. *Journal of Applied Physics*, 91, 9549-4555.

Hölscher, H.; Schwarz, U. D. & Wiesendanger, R. (1999). Calculation of the frequency shift in dynamic force microscopy. *Applied Surface Science*, 140, 344-351.

Hurley, D. C.; Shen, K.; Jennett, N. M. & Turner, J. A. (2003). Atomic force acoustic microscopy methods to determine thin-film elastic properties. *Journal of Applied Physics*, 94, 2347-2354.

Kokavecz, J.; Marti, O.; Heszler, P. & and Mechler, A. (2006). Imaging bandwidth of the tapping mode atomic force microscope probe. *Physical Review B*, 73, 155403.

Kolosov O. & Yamanaka, K. (1993). Nonlinear detection of ultrasonic vibrations in an atomic force microscope. *Japanese Journal of Applied Physics*, 32, L1095-L1098.

Kolosov, O. V.; Castell, M. R.; Marsh, C. D.; Briggs, G. A. D.; Kamins, T. I. & Williams, R. S. (1998). Imaging the elastic nanostructure of Ge islands by ultrasonic force microscopy. *Physical Review Letters*, 81, 1046-1049.

Kopycinska-Müller, M.; Geiss, R. H. & Hurley, D. C. (2006). Contact mechanics and tip shape in AFM-based nanomechanical measurements. *Ultramicroscopy* 106, 466-474.

Lantz, M. A.; Hug, H. J.; Hoffmann, R.; van Schendel, P. J. A.; Kappenberger, P.; Martin, S.; Baratoff, A. & Güntherodt, H.-J. (2001). Quantitative measurtement of short-range chemical bonding forces. *Science*, 291, 2580-2583.

Law, B. M. & Rieutord, F. (2002). Electrostatic forces in atomic force microscopy. *Physical Review B*, 66, 035402.

Lee, H.-L.; Yang, Y.-C.; Chang, W.-J. & Chu, S.-S. (2006). Effect of interactive damping on vibration sensitivities of V-shaped atomic force microscope cantilevers. *Japanese Journal of Applied Physics,* 45, 6017-6021.

Maivald, P.; Butt, H. J.; Gould, S. A.; Prater, C. B.; Drake, B.; Gurley, J. A.; Elings, V. B. & Hansma, P. K. (1991). Using force modulation to image surface elasticities with the atomic force microscope. *Nanotechnology*, 2, 103-106.

Meirovitch, L. (1967). *Analytical Methods in Vibrations*, Macmillan, New York.

Muthuswami, L. & Geer, R. E. (2004). Nanomechanical defect imaging in premetal dielectrics for integrating circuits. *Applied Physics Letters*, 84, 5082-5084.

Nony, L.; Boisgard, R. & Aime, J. P. (1999). Nonlinear dynamical properties of an oscillating tip-cantilever system in the tapping mode. *Journal of Chemical Physics*, 111, 1615-1627.

Park, C.; Ounaies, Z.; Watson, K. A.; Crooks, R. E.; Smith, Jr., J.; Lowther, S. E.; J. Connell, W.; Siochi, E. J.; Harrison, J. S. & St. Clair, T. L. (2002). Dispersion of single wall carbon nanotubes by in situ polymerization under sonication. *Chemical Physics Letters*, 364, 303-308.

Polesel-Maris, J; Piednoir, A.; Zambelli, T.; Bouju, X. & Gauthier, S. (2003). Experimental investigation of resonance curves in dynamic force microscopy. *Nanotechnology*, 14, 1036-1042.

Rabe U. & Arnold, W. (1994). Acoustic microscopy by atomic force microscopy. *Applied Physics Letters*, 64, 1493-1495.

Rabe, U.; Amelio, S.; Kopychinska, M.; Hirsekorn, S.; Kempf, M.; Goken, M. & Arnold, W. (2002). Imaging and measurement of local mechanical properties by atomic force micrscopy. *Surface and Interface Analysis*, 33, 65-70.

Saint Jean, M.; Hudlet, S.; Guthmann, C. & Berger, J. (1994). Van der Waals and capacitive forces in atomic force microscopies. *Journal of Applied Physics*, 86, 5245-5248.

Schiff, L. I. (1968). *Quantum Mechanics*, McGraw-Hill, New York.

Shekhawat, G. S. & Dravid V. P. (2005). Nanoscale imaging of buried structures via scanning near-field ultrasonic holography. *Science*, 310, 89-92.

Sokolnikoff, I. S. & Redheffer, R. M. (1958). *Mathematics of Physics and Modern Engineering,* McGraw-Hill, New York.

Stark, R. W. & Heckl, W. M. (2003). Higher harmonics imaging in tapping-mode atomic-force microscopy. *Review of Scientific Instruments*, 74, 5111-5114.

Stark, R. W.; Schitter, G.; Stark, M.; Guckenberger, R. & Stemmer, A. (2004). State-space model of freely vibrating surface-coupled cantilever dynamics in atomic force microscopy. *Physical Review B*, 69, 085412.

Turner, J. A. (2004). Nonlinear vibrations of a beam with cantilever-Hertzian contact boundary conditions. *Journal of Sound and Vibration*, 275, 177-191.

Überall, H. (1997). Interference and steady-state scattering of sound waves. In: *Encyclopedia of Acoustics*, *Vol. 1*, Malcohm J. Crocker, (Ed.), 55-68, Wiley, ISBN 0-471-17767-9, New York.

Wolf, K. & Gottlieb, O. (2002). Nonlinear dynamics of a noncontacting atomic force microscope cantilever actuated by a piezoelectric layer. *Journal of Applied Physics*, 91, 4701-4709.

Yagasaki, K. (2004). Nonlinear dynamics of vibrating microcantilevers in tapping-mode atomic force micrscopy. *Physical Review B*, 70, 245419.

Yamanaka, K.; Ogiso, H. & Kolosov, O. (1994). Ultrasonic force microscopy for nanometer resolution subsurface imaging. *Applied Physics Letters*, 64, 178-180.

Yaralioglu, G. G.; Degertekin, F. L.; Crozier, K. B. & Quate, C. F. (2000). Contact stiffness of layered materials for ultrasonic atomic force microscopy. *Journal of Applied Physics*, 87, 7491-7496.

Zheng, Y.; Geer, R. E.; Dovidenko, K.; Kopycinska-Müller, M. & Hurley, D. C. (2006). Quantitative nanoscale modulus measurements and elastic imaging of SnO_2 nanobelts. *Journal of Applied Physics*, 100, 124308.

Zhong , Q.; Inniss, D.; Kjoller, K. & Elings, V. B. (1993). Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. *Surface Science Letters*, 290, L688 – L692.

5

Nonlinear Phenomena during the Oxidation and Bromination of Pyrocatechol

Takashi Amemiya[1] and Jichang Wang[1,2]
[1]*Graduate School of Environment and Information Sciences,*
Yokohama National University, Yokohama, 240-8501
[2]*Department of Chemistry and Biochemistry, University of Windsor,*
Windsor, Ontario, N9B 3P4
[1]*Japan*
[2]*Canada*

1. Introduction

Many complex and interesting phenomena in nature are due to nonlinear interactions of the constituents (Nicolis & Prigogine, 1977; Ball, 1999; Morowitz, 2002). The study of nonlinear dynamical systems has achieved significant progress over the last four decades, which allows scientists to understand various rather complicated behaviors such as self-organization and pattern formation in the neuronal networks of brain (Scott Kelso, 1995). Unusual properties of reagents in far-from-equilibrium conditions and the prevalence of instability where small changes in initial conditions may lead to amplified effects have been documented more than a century ago, but those nonlinear chemical phenomena did not get much attention until late 1960s after the discovery of oscillatory behavior in a homogeneous solution reaction between acidic bromate and malonic acid in the presence of metal catalyst cerium (Field & Burgur, 1985; Scott, 1994; Epstein & Pojman 1998; Sagues & Epstein, 2003). The system is now commonly known as the Belousov-Zhabotinsky (BZ) reaction (Zaikin & Zhabotinsky, 1970; Field & Burger, 1985). Since then, the study of chemical oscillations and wave formation has blossomed, which led to the observation of various nonlinear spatiotemporal behaviours such as both simple and complex oscillations in a stirred system (Smoes, 1979; Györgi & Field, 1992; Wang et al., 1995 & 1996; Zhao et al., 2005), Turing pattern (Horváth et al., 2009), target and spiral waves in a two-dimensional reaction-diffusion medium (Zaikin & Zhabotinsky, 1970; Winfree, 1972; Yamaguchi et al., 1991; Steinbock et al., 1995; Kádár et al., 1998), and scroll waves in a 3-dimensional system (Welsh et al., 1983; Winfree, 1987; Jahnke et al., 1988; Amemiya et al., 1996). Understanding the onset of those exotic phenomena in chemical systems has provided important insight into the formation of similar behaviour in nature (Goldbeter, 1996; Dutt & Menzinger, 1999; Dhanarajan et al., 2002; Carlsson et al., 2006; Chiu et al., 2006).

As opposed to nonlinear systems in physical and biological areas, in which dynamic control parameters are often inaccessible or difficult to adjust, chemical reactions can be conveniently manipulated through adjusting the initial concentration of each reagent, temperature, or flow rate in a continuously flow stirred tank reactor (CSTR) (Epstein, 1989;

Mori et al., 1993; Amemiya et al., 2002). As a result, chemical media have played a very important role in gaining insights into various nonlinear behaviors encountered in nature (Nicolis & Prigogine, 1989; Sørensen et al., 1990; Kumli et al., 2003; Kurin-Csörgei et al., 2004; McIlwaine et al., 2006). Among existing chemical oscillators, the vast majority relies on a few elements that possess multiple oxidation states, such as halogens, sulfur and some transition metals. In 1978, Orbán and Körös carried out an extensive search to explore chemical oscillations in the oxidations of aromatic compounds by acidic bromate (Körös & Orbán, 1978; Orbán & Körös, 1978a; 1978b). Because of the absence of metal catalysts, systems reported by Orbàn and Körös in 1978 and discovered more recently by other groups have been frequently referred as uncatalyzed bromate oscillators (UBO) (Farage & Janjic, 1982; Szalai & Körös, 1998; Adamcikova et al., 2001). In general, reactions of UBOs represent the parallel running of oxidation and bromination of an organic substrate.

This chapter described nonlinear chemical kinetics in the bromate-pyrocatechol reaction with or without the presence of metal catalysts (Harati & Wang, 2008a; 2008b). The bromate-pyrocatechol reaction system was initially investigated by Orbàn and Körös in 1978 (Orbán & Körös, 1978a; 1978b). Unfortunately, no oscillatory behavior could be observed. The absence of spontaneous oscillations in the earlier attempt has been attributed to two major factors: First, the reaction between acidic bromate and pyrocatechol results in the production of bromine, which inhibits autocatalytic reactions; secondly, the oxidation product of pyrocatechol is a stable benzoquinone. As is shown in our recent reports, upon extensive search in the concentration phase space the bromate-pyrocatechol reaction was found to be capable of exhibiting spontaneous oscillations in a stirred batch system (Harati & Wang, 2008b). A phase diagram established in the bromate and pyrocatechol concentration space sheds light on why finding chemical oscillations in this chemical system is such a challenging task. Same as reported in other UBOs (Wang et al., 2001; Zhao & Wang, 2006 & 2007), the bromate-pyrocatechol reaction exhibits subtle responses to illumination, where, depending on the reaction conditions, either light-induced or light-quenched oscillatory phenomena could be observed. The influence of metal catalysts on the nonlinear dynamics of the bromate-pyrocatechol reaction was also discussed here.

2. Experimental observation of spontaneous oscillations

2.1 The uncatalyzed bromate-pyrocatechol reaction

Figure 1 presents three time series of the bromate-pyrocatechol (H_2Q) reaction performed under different initial concentrations of $NaBrO_3$: (a) 0.085 M, (b) 0.093 M, and (c) 0.095 M. Other reaction conditions are $[H_2Q]$ = 0.057 M and $[H_2SO_4]$ = 1.4 M. Details of the experimental procedure can be found in the original reports (Harati & Wang, 2007b). Shortly after mixing all chemicals together, Pt potential as seen in Fig. 1a exhibited clock reaction phenomenon, which was followed by gradual decrease for several hours. Phenomenologically, the excursion of the Pt potential was accompanied by a dramatic color change of the reaction solution from transparent to deep red. After the rapid color change, which has been observed in all of the following experiments, the red color gradually turned into yellow within the next two hours. Our experiments showed that for low bromate concentrations (<0.09 M), Pt potential of the system decreased monotonically after the initial excursion. Chemical oscillations were obtained when bromate concentration was increased to 0.093 M. Further increase of bromate concentration led to slightly irregular oscillations in Fig. 1c, where not only the amplitude but also the frequency of oscillation fluctuated. To

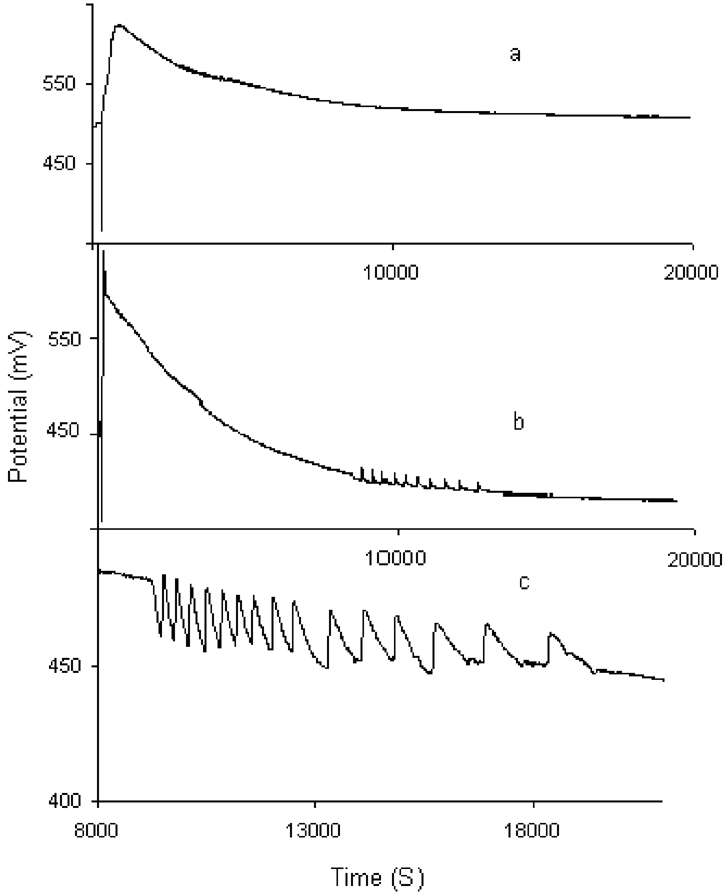


Fig. 1. Time series of the pyrocatechol – bromate reaction at different initial concentrations of bromate: (a) 0.085 M, (b) 0.093 M and (c) 0.095M. Other reaction conditions are [H_2Q] = 0.057 M, and [H_2SO_4] = 1.4 M.

show modulations in the oscillation frequency clearly, only the oscillation window is plotted in Fig. 1c, in which the long induction time period, similar to the ones plotted in Figs. 1a and 1b, is omitted. As bromate concentration was increased continuously, the system underwent reverse bifurcations leading the system back to non-oscillatory progress in time where the evolution of Pt potential was the same as that in Fig. 1a. For conditions employed in Fig. 1, spontaneous oscillations have been obtained when bromate concentration was between 0.09 and 0.11 M (Harati & Wang, 2007b).

Figure 2a plots the number of oscillation peak as a function of bromate concentration, where it increases with bromate concentration and then drops sharply to 0 as the system moves out of the oscillation window at the high bromate concentration. Fig. 2b illustrates that the induction time (IP) of these spontaneous oscillations grows monotonically with the increase

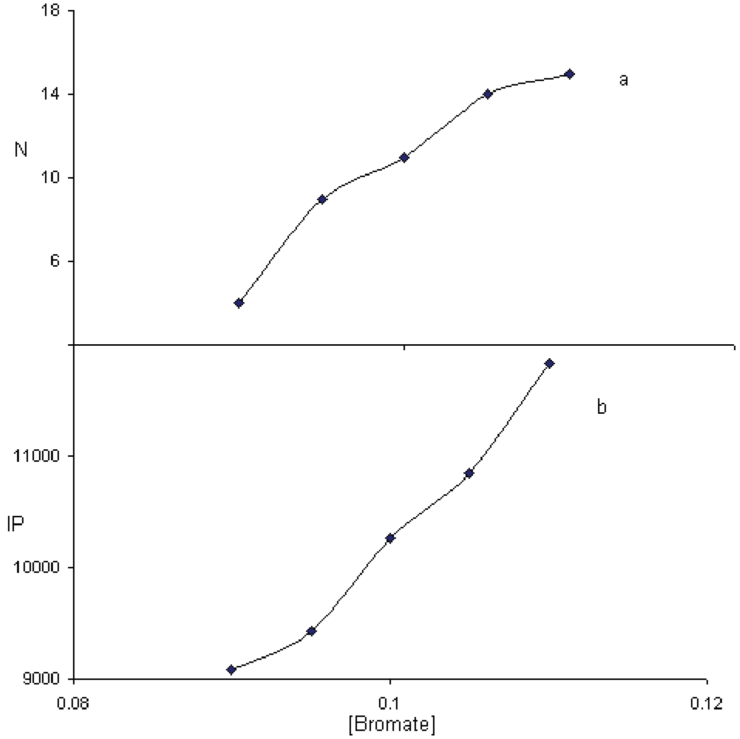


Fig. 2. Dependence of the number of oscillations (N) and induction period (IP) on the initial concentration of bromate. Other reaction conditions are [H_2Q] = 0.057 M and [H_2SO_4] = 1.2 M.

of bromate concentration. The extremely long induction seen here is similar to that reported in other uncatalyzed bromate oscillators (Farage & Janjic, 1982; Szalai & Körös, 1998; Adamcikova et al., 2001)

Figure 3 presents temporal evolutions of the bromate-H_2Q reaction under different initial concentrations of H_2Q: (a) 0.038 M, (b) 0.044 M, and (c) 0.047 M. In Fig. 3a Pt potential exhibits a clock reaction phenomenon, followed by a gradual decrease. This behavior is the same as seen at a low bromate concentration, where the clock variation of Pt potential is accompanied by a dramatic color change of the reaction solution. When H_2Q concentration was increased to 0.044 M in Fig. 3b, spontaneous oscillations took place at about 2 hours after the solution has turned into yellow. Further increase of H_2Q concentration also resulted in some irregularity in those transient oscillations such as the one shown in Fig. 3c. Again, to show details of the chemical oscillations time scale in Fig. 3c is different from that used in Figs. 3a and 3b. Within the oscillation window the induction time decreased monotonically with the increase of H_2Q concentration. On the other hand, the total number of oscillations increased rapidly as H_2Q concentration became larger than the lower bifurcation threshold and then decreased gradually as H_2Q concentration was increased further. The above results indicate that bromate and H_2Q have opposite effects on the oscillatory behavior.

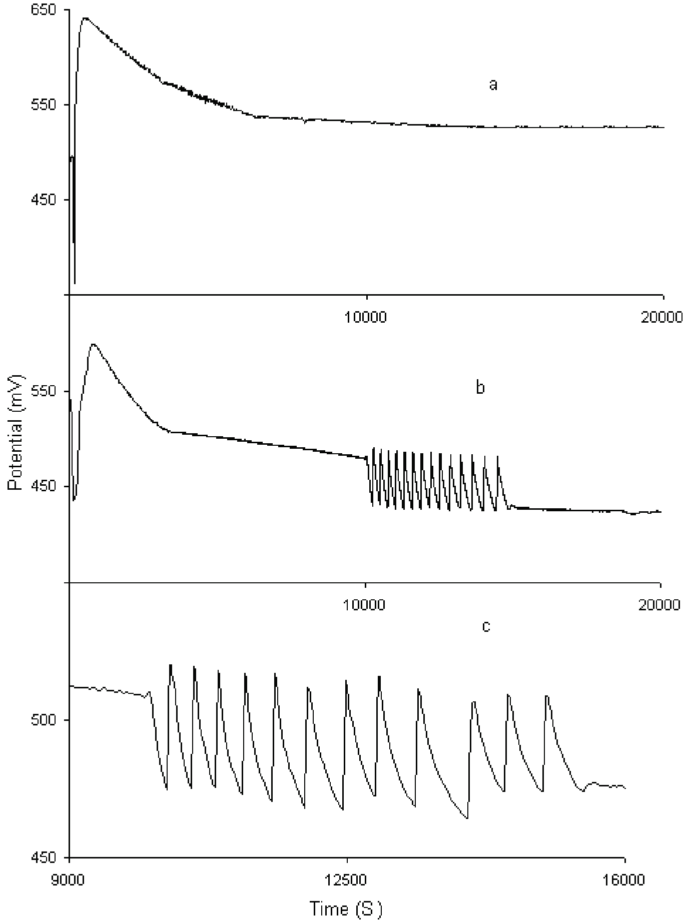


Fig. 3. Time series of the pyrocatechol – bromate reaction at different initial concentrations of pyrocatechol: (a) 0.038 M, (b) 0.044 M and (c) 0.047M. Other reaction conditions are [$NaBrO_3$] = 0.085 M, and [H_2SO_4] = 1.4 M

Figure 4 is a phase diagram in the pyrocatechol - bromate concentration plane, where filled triangles denote the conditions at which the system exhibits spontaneous oscillations. Here, the concentration of H_2SO_4 is fixed at 1.4 M. First glance of this phase diagram indicates that the oscillatory behavior exists over broad concentrations of pyrocatechol and bromate. However, at each given concentration of bromate (or pyrocatechol) there is only a narrow range of pyrocatechol (or bromate) concentration within that the system oscillates. This diagonal narrow band window sheds light on the difficulty of landing the initial conditions within such a window, when starting the experiments without existing information of this system.

Dependence of the above chemical oscillations on H_2SO_4 and bromate concentrations is summarized in Fig. 5, in which H_2Q concentration was fixed at 0.057 M. Filled triangles are

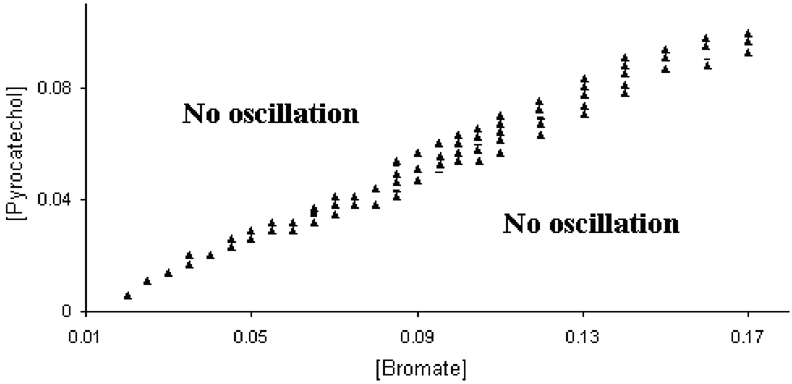


Fig. 4. Phase diagram of the bromate-pyrocatechol reaction in the pyrocatechol – bromate concentration plane. (▲) denotes where the system exhibits transient oscillations. Sulfuric acid concentration was fixed at 1.4 M.

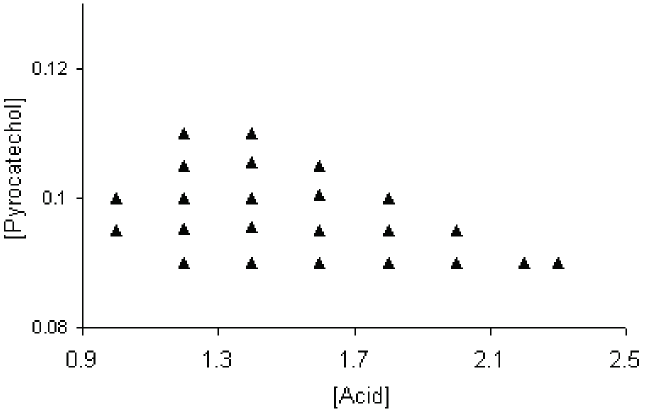


Fig. 5. Phase diagram of the bromate-pyrocatechol reaction in the bromate – H_2SO_4 concentration plane. (▲) denotes the conditions under which the system exhibits oscillations. The concentration of pyrocatechol is 0.057 M.

the conditions under which the system exhibits spontaneous oscillations. This phase diagram shows that when the concentration of H_2SO_4 is larger than 2.5 M or smaller than 0.9 M, no oscillations can be obtained regardless bromate concentration. On the other hand, the range of H_2SO_4 concentration over which the system exhibits spontaneous oscillations is broadened by lowering bromate concentration.

2.2 The ferroin-bromate-pyrocatechol reaction

Figure 6 presents time series of (a) the uncatalyzed and (b) ferroin-catalyzed bromate-pyrocatechol reactions. In the uncatalyzed system, the Pt potential decreased gradually after the initial excursion and then reached a plateau. In general, one might have considered that this closed reaction is over. However, the Pt potential suddenly started oscillating after another two hours, and the oscillatory process lasted for longer than an hour with about 14 peaks. This result illustrates that under the conditions investigated here the uncatalyzed bromate-pyrocatechol is capable of exhibiting spontaneous oscillations. There is no periodic color change during the oscillation and thus the uncatalyzed system is deemed unsuitable for studying chemical waves in spatially extended media.

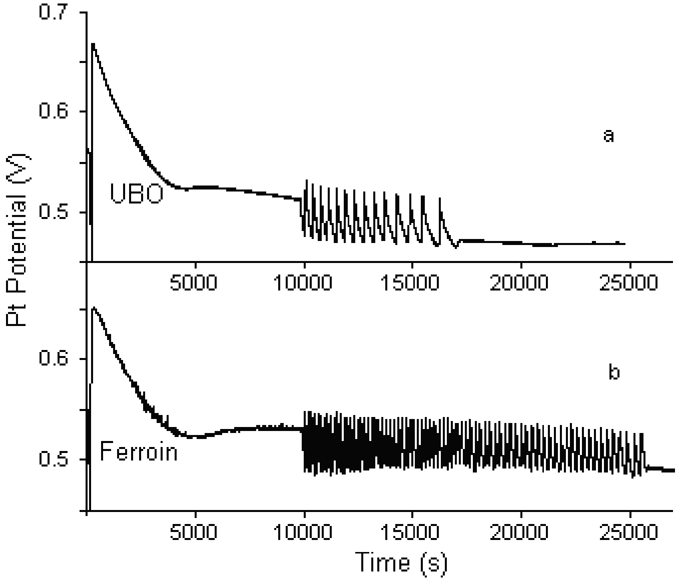


Fig. 6. Time series of the (a) uncatalyzed, and (b) ferroin-catalyzed bromate-pyrocatechol reaction. Other reaction conditions are: $[H_2SO_4]$ = 1.30 M, $[H_2Q]$ = 0.043 M and $[BrO_3^-]$ = 0.078 M. The concentration of ferroin is equal to 1.0 x 10^{-4} M in (b).

In Fig. 6b, when 1.0 x 10^{-4} M ferroin was added to the bromate-pyracatechol reaction, spontaneous oscillations commenced at about the same time as in the uncatalyzed system. However, there are significant changes in the frequency of oscillation and the total number of oscillations and both have been increased greatly. Notably, in this catalyzed system the oscillation lasted for longer than 4 hours. Our experiments illustrate that this system exhibits observable periodic color changes from yellowish to faint pink during the oscillatory window when the concentration of ferroin is above 1.0 x 10^{-4} M. Further increase of the concentration of ferroin results in a better contrast, but reduces the lifetime of the oscillatory period. Furthermore, when ferroin concentration is higher than 1.0 x 10^{-3} M no obvious color change could be seen in the stirred system. After oscillations in the ferroin-bromate-pyracatechol system stopped, the solution has a blue color if the concentration of ferroin added is above 1.0 x 10^{-3} M, or a pink color when the ferroin concentration is less than 5 × 10^{-4} M.

Figure 7 summarizes the dependence of the number of oscillations (N) and the induction time (IP) on the concentration of ferroin. There is a sharp increase in the number of oscillations at a very low concentration of ferroin (10^{-5} M), suggesting that the presence of small amounts of metal catalyst favours the oscillatory behaviour. As the amount of ferroin is increased, however, the number of oscillations decreases, which may be due to the increased consumption of the reactants. Notably, ferroin shows a little effect on the induction time (IP), where increasing ferroin concentration to 0.002 M only reduces the IP by about 10 percent (Harati & Wang, 2008a).

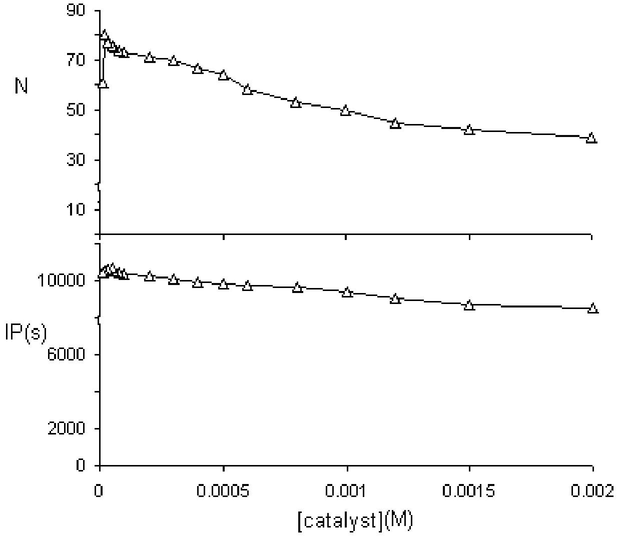


Fig. 7. Dependence of the number of oscillations (N) and induction period (IP) on the concentration of ferroin. Other reaction conditions are: [H_2SO_4] = 1.30 M, [BrO_3^-] = 0.078 M, and [pyrocatechol] = 0.043 M.

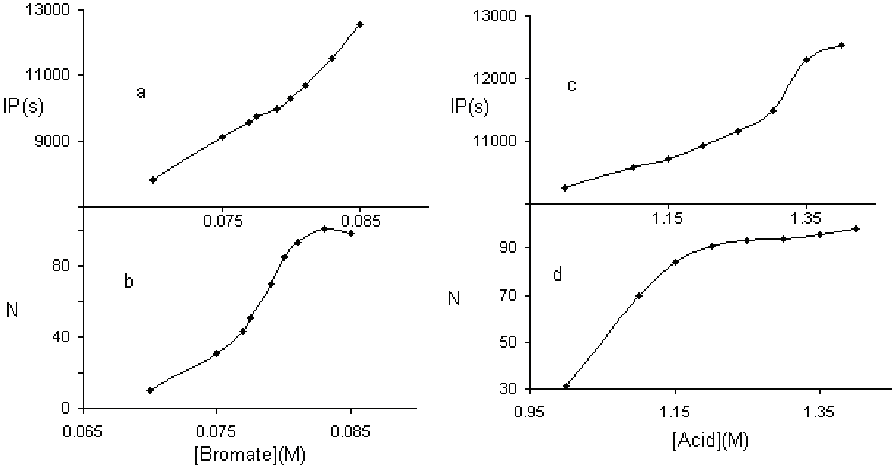


Fig. 8. Dependence of the number of oscillations (N) and induction time (IP) of the ferroin-catalyzed system on the concentration of bromate and sulfuric acid. Other reaction conditions are: [H_2Q] = 0.044 M, [ferroin] = 1.0 x 10^{-4} M, and (a &b) [H_2SO_4] = 1.40 M; (c&d) [$NaBrO_3$] = 0.085 M.

Figure 8 plots, respectively, the number of oscillations (N) and induction time (IP) as a function of concentrations of bromate and sulfuric acid in the ferroin-bromate-pyrocatechol system, where the concentration of ferroin was fixed at 1.0 x 10^{-4} M. Figs. 8a and 8b show that increasing bromate concentration prolongs the induction period, which may arise from the production of larger amounts of bromine in the reaction solution. The number of peaks ascends first and then declines slightly with increasing bromate concentration. Under the conditions studied here, the concentration of bromate must be between 0.070 and 0.085 M for the system to show spontaneous oscillations. As shown in Figs. 8c and 8d, both N and IP increase monotonically with the increase of H_2SO_4 concentration. The system does not oscillate when the concentration of H_2SO_4 is higher than 1.4 M or lower than 1.0 M under the conditions studied.

Figure 9 is a phase diagram of the ferroin-catalyzed system in the pyrocatechol and bromate concentration plane, where (♦) indicates the conditions under which the system exhibits spontaneous oscillations. Similar to the situation of the uncatalyzed bromate-pyrocatechol reaction, the first glance of this figure suggests that the system is able to exhibit oscillatory dynamics over a broad range of bromate and pyrocatechol concentrations. However, at each given concentration of pyrocatechol (or bromate), the proper concentration of bromate (or pyrocatechol) is quite narrow. This narrow band shaped phase diagram suggests that nonlinear behavior of this catalyzed system is more sensitive to the ratio of $[H_2Q]/[BrO_3^-]$ than their absolute concentrations. In comparison to the uncatalyzed bromate-pyrocatechol system, the presence of ferroin does not change the shape of this phase diagram, but makes the area of the parameter window slightly larger, implicating that ferroin favors the oscillations.

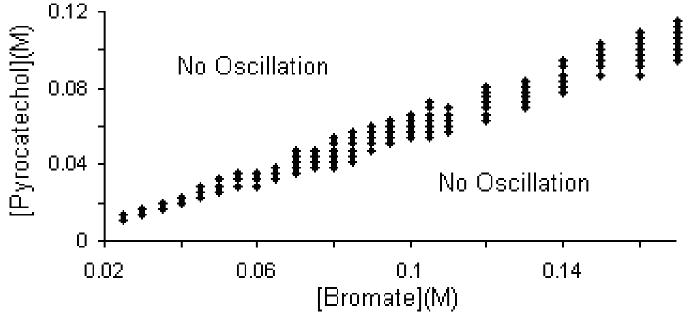


Fig. 9. Phase diagram of the ferroin-catalyzed reaction in the bromate–pyrocatechol concentration plane. (♦) denotes where the system exhibits simple periodic oscillations. The concentration of ferroin is 1.0 x 10^{-4} M.

Time series measured with a bromide selective electrode show that bromide concentration increases slowly during the long induction time and then starts oscillating (Harati & Wang, 2008b). It is similar to the behavior reported in earlier studies of the uncatalyzed bromate-1,4-cyclohexanedione and bromate-1,4-benzoquinone reactions (Szalai & Körös, 1998; Zhao & Wang, 2006), in which the accumulation of bromide precursors has been suggested to be responsible for the induction time. In this system, however, the initial addition of bromide, which leads to the rapid production of bromine and then causes the bromination of pyrocatechol, evidenced by mass spectrometry study (Harati & Wang, 2008b), does not shorten the induction time. The slight decrease in the induction time observed at a very high bromide concentration may result from decreases in H_2Q and BrO_3^- concentrations due to reactions with bromine. The insensitivity of the induction time to the initial presence of brominated substrates suggests that the governing mechanism of this oscillator may be different from UBOs reported earlier.

2.3 The influence of Ce^{4+}/Ce^{3+} and Mn^{3+}/Mn^{2+}

It is well known that metal catalysts such as ferroin participate the autocatalytic reactions with bromine dioxide radicals (BrO_2*) and therefore redox potential of the metal catalyst in relative to the redox potential of $HBrO_2/BrO_2$* couple is an important parameter in determining the rate of the autocatalytic cycle, which in turn has significant effects on the overall reaction behavior. In the BZ reaction, four metal catalysts including ferroin, ruthenium, cerium and manganese can be oxidized by bromine dioxide radicals, in which the redox potential of $HBrO_2/BrO_2$* couple is larger than that of ferroin and ruthenium, but smaller than that of Ce^{4+}/Ce^{3+} and Mn^{3+}/Mn^{2+}. Therefore, it is anticipated that when cerium or manganese ions are introduced into the bromate-pyrocatechol reaction, behavior different from that achieved in the ferroin-bromate-pyrocatechol system may emerge. Figure 10 plots the number of oscilllations (N) and induction time (IP) of the catalyzed bromate-pyrocatechol reaction as a function of catalyst (i.e. Ce^{4+} and Mn^{2+}) concentration.

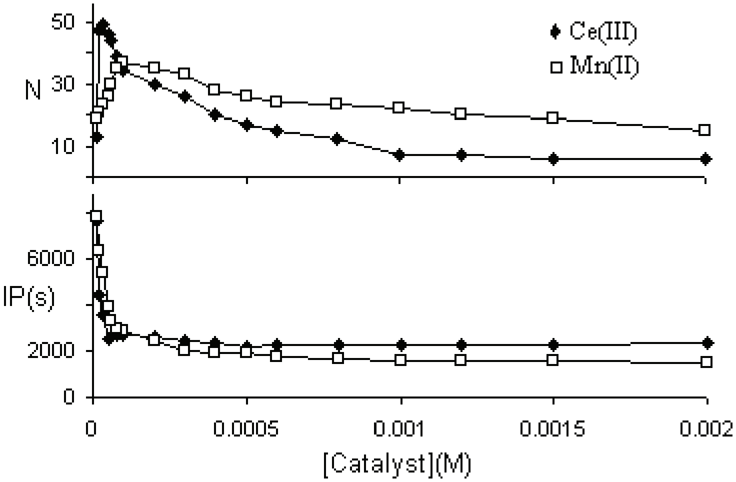


Fig. 10. Dependence of the number of oscillations (N) and induction time (IP) on the initial concentrations of cerium and menganese. Other reaction conditions are $[H_2SO_4] = 1.3$ M, $[NaBrO_3] = 0.078$ M, and $[H_2Q] = 0.043$ M.

The sharp increase in the number of oscillations at the low concentration of cerium and manganese illustrates that the presence of a small amount of metal catalyst favours the oscillatory behaviour, similar to the case of ferroin. As the amount of catalyst (i.e. Ce^{4+} or Mn^{2+}) increases, however, the number of oscillations decreases rapidly. It could be due to the increased consumption of major reactants, in particular bromate. Overall, the effect of Mn^{2+} or Ce^{4+} on the number of oscillations was not as significant as ferroin, although they doubled the number of peaks at an optimized condition. In contrast, the presence of a small amount of cerium or manganese dramatically reduced the induction time, where the induction time was shortened from about 3 hours in the uncatalyzed system to approximately half an hour when the concentration of manganese and cerium reached, respectively, 2.0×10^{-4} and 5.0×10^{-5} M. The IP became relative stable when the concentration of manganese or cerium was increased further.

When comparing with the time series of the ferroin system presented in Fig. 6b, for the cerium-catalyzed bromate-pyrocatechol reaction the Pt potential stayed flat after the initial excursion. The amplitude of oscillation became significantly larger than that of the uncatalyzed as well as the ferroin-catalyzed systems; but, there was no significant increase in the total number of oscillations when compared with the uncatalyzed system. Unlike the ferroin-catalyzed system, no periodic color change was achieved and thus is unfit for studying waves. A short induction time and large oscillation amplitude (> 300 mV), however, make the cerium-catalyzed system suitable for exploring temporal dynamics in a stirred system. In particular, oscillations in the cerium system have a broad shoulder which may potentially develop into complex oscillations. Times series of the Mn^{2+}-catalyzed bromate-pyrocatechol reaction is very similar to that of the cerium-catalyzed one, in which the Pt potential stayed flat after the initial excursion and the oscillation commenced much earlier than in the uncatalyzed system. The number of oscillations in the manganese system is also slightly larger than that of the uncatalyzed system. Overall, cerium and manganese, both have a redox potential above the redox potential of $HBrO_2/BrO_2^*$, exhibit almost the same influence on the reaction behavior.

2.4 Photochemical behavior

Ferroin-catalyzed BZ reaction is insensitive to the illumination of visible light. As a result, the vast majority of existing studies on photosensitive chemical oscillators have been performed with ruthenium as the metal catalyst, despite that ruthenium complex is expensive and difficult to prepare. In Figure 11, the photosensitivity of the ferroin-catalyzed bromate-pyrocatechol reaction was examined, in which the concentration of ferroin was adjusted. As shown in Fig. 11a, when the system was exposed to light from the beginning of the reaction, spontaneous oscillations emerged earlier, where the induction time was shortened to about 6000 s, but the oscillatory process lasted for a shorter period of time. The system then evolved into non-oscillatory evolution. Interestingly, after turning off the illumination the Pt potential jumped to a higher value immediately and, more significantly, another batch of oscillations developed after a long induction time. The above result indicates that the ferroin-bromate-pyrocatechol reaction is photosensitive and influence of light in this ferroin-catalyzed system is subtle. On one hand, illumination seems to favor the oscillatory behavior by shortening the induction time, but it later quenches the oscillations.

In Fig. 11b the concentration of ferroin was doubled. When illuminated with the same light as in Fig. 11a from the beginning, no oscillation was achieved, except there was a sharp drop in the Pt potential at about the same time as that when oscillations occurred in Fig. 11a. After turning off the light, the un-illuminated system exhibited oscillatory behaviour with a long induction time. We have also applied illumination in the middle of the oscillatory window, in which a strong illumination such as 100 mW/cm^2 immediately quenched the oscillatory behaviour and oscillations revived shortly after reducing light intensity to a lower level such as 30 mW/cm^2. Interestingly, although ferroin itself is not a photosensitive

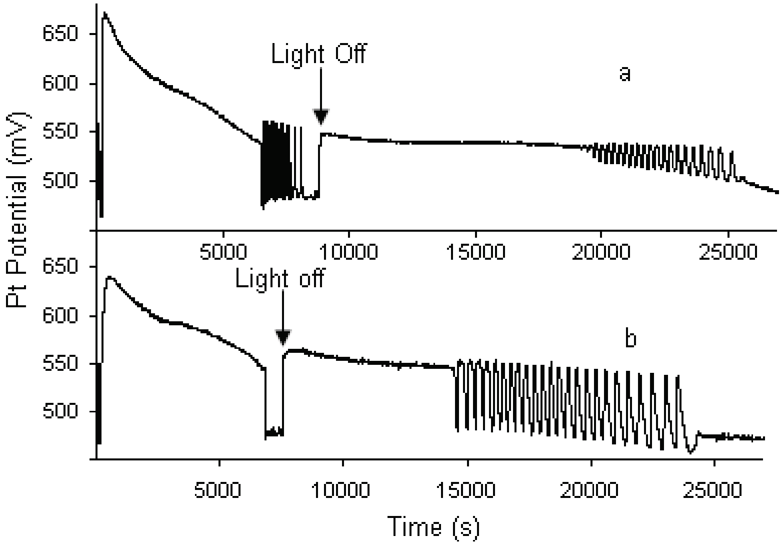


Fig. 11. Light effect on the bromate – pyrocatechol – ferroin reaction (a) and (b) light illuminating from the beginning with intensity equal to 70 mW/cm^2, under conditions [$NaBrO_3$] = 0.10 M, [H_2SO_4] = 1.40 M, [H_2Q] = 0.057 M, (a) [Ferroin] = 5.0×10^{-4} M, and (b) [Ferroin] = 1.0×10^{-3} M.

reagent, here its concentration nevertheless exhibits strong influence on the photoreaction behaviour of the bromate-pyrocatechol system. Carrying out similar experiments with the cerium- and manganese-catalyzed system under the otherwise the same reaction conditions showed little photosensitivity, in which no quenching behaviour could be obtained, although light did cause a visible decrease in the amplitude of oscillation.

3. Modelling

3.1 The model

To simulate the present experimental results, we employed the Orbán, Körös, and Noyes (OKN) mechanism (Orbán et al., 1979) proposed for uncatalyzed reaction of aromatic compounds with acidic bromate. The original OKN mechanism is composed of sixteen reaction steps, i.e., ten steps K1 – K10 in Scheme I and six steps K11 – K16 in Scheme II as listed in Table 1. We selected all ten reaction steps K1 – K10 from Scheme I and the first four reaction steps K11 – K14 in Scheme II. The reason behind such a selection is that all reaction steps in Scheme I as well as the first four reaction steps in Scheme II are suitable for an aromatic compound containing at least two phenolic groups such as pyrocatechol used in the present study.

Reaction steps K15 and K16 in Scheme II, on the other hand, suggest how phenol and its derivatives could be involved in the oscillatory reactions. There is no experimental evidence that pyrocatechol can be transformed into a substance of phenol type, we thus did not take into account reactions involving phenol and its derivatives. The model used in our simulation consists of fourteen reaction steps K1 – K14, and eleven variables, BrO_3^-, Br^-, BrO_2*, $HBrO_2$, HOBr, Br*, Br_2, $HAr(OH)_2$, HAr(OH)O*, Q, and BrHQ, where $HAr(OH)_2$ is pyrocatechol abbreviated as H_2Q in the experimental section, HAr(OH)O* is pyrocatechol radical, $HArO_2$ is 1,2-benzoquinone and $BrAr(OH)_2$ is brominted pyrocatechol.

The simulation was carried out by numerical integration of the set of differential equations resulting from the application of the law of mass action to reactions K1 – K14 with the rate constants as listed in Table 1. The values of the rate constants for reactions K1 – K3, K5, K8 have already been determined in the studies of the BZ reaction, and those of all other reactions were either chosen from related work on the modified OKN mechanism by Herbine and Field (Herbine & Field, 1980) or adjusted to give good agreement between experimental results and simulations.

no.	reaction	rate constant		ref.
		forward	reverse	
Scheme I				
(K1)	$BrO_3^- + Br^- + 2H^+ \leftrightarrow HBrO_2 + HOBr$	$k_1 = 2.1\,[H^+]^2\,M^{-1}\,s^{-1}$	$k_{r1} = 1 \times 10^4\,M^{-1}\,s^{-1}$	*a*
(K2)	$HBrO_2 + Br^- + H^+ \rightarrow 2HOBr$	$k_2 = 2 \times 10^9\,[H^+]\,M^{-1}\,s^{-1}$		*a*
(K3)	$BrO_3^- + HBrO_2 + H^+ \leftrightarrow 2BrO_2^* + H_2O$	$k_3 = 1 \times 10^4\,[H^+]\,M^{-1}\,s^{-1}$	$k_{r3} = 2 \times 10^7\,M^{-1}\,s^{-1}$	*a*
(K4)	$BrO_2^* + HAr(OH)_2 \rightarrow HBrO_2 + HAr(OH)O^*$	$k_4 = 3 \times 10^2\,M^{-1}\,s^{-1}$		*a, b*
(K5)	$2HBrO_2 \rightarrow BrO_3^- + HOBr + H^+$	$k_5 = 4 \times 10^7\,M^{-1}\,s^{-1}$		*a*
(K6)	$HOBr + HAr(OH)O^* \leftrightarrow Br^* + HArO_2 + H_2O$	$k_6 = 2 \times 10^5\,M^{-1}\,s^{-1}$	$k_{r6} = 2 \times 10^2\,M^{-1}\,s^{-1}$	*a, b*
(K7)	$Br^* + HAr(OH)O^* \rightarrow Br^- + HArO_2 + H^+$	$k_7 = 4 \times 10^4\,M^{-1}\,s^{-1}$		*a, b*
(K8)	$HOBr + Br^- + H^+ \leftrightarrow Br_2 + H_2O$	$k_8 = 8 \times 10^9\,[H^+]\,M^{-1}\,s^{-1}$	$k_{r8} = 1.1 \times 10^2\,s^{-1}$	*a*
(K9)	$Br_2 + HAr(OH)_2 \rightarrow BrAr(OH)_2 + Br^- + H^+$	$k_9 = 7 \times 10^2\,M^{-1}\,s^{-1}$		*a, b*
(K10)	$HOBr + HAr(OH)_2 \rightarrow BrAr(OH)_2 + H_2O$	$k_{10} = 1\,M^{-1}\,s^{-1}$		*a, b*
Scheme II				
(K11)	$BrO_2^* + HAr(OH)O^* \rightarrow HBrO_2 + HArO_2$	$k_{11} = 5 \times 10^3\,M^{-1}\,s^{-1}$		*b*
(K12)	$Br^* + HAr(OH)_2 \rightarrow Br^- + HAr(OH)O^* + H^+$	$k_{12} = 1 \times 10^3\,M^{-1}\,s^{-1}$		*a, b*
(K13)	$HAr(OH)_2 + HArO_2 \leftrightarrow 2HAr(OH)O^*$	$k_{13} = 1 \times 10^{-2}\,M^{-1}\,s^{-1}$,	$k_{r13} = 1 \times 10^4\,M^{-1}\,s^{-1}$	*a, b*
(K14)	$2HAr(OH)O^* \rightarrow Ar_2(OH)_4$	$k_{14} = 1 \times 10^4\,M^{-1}\,s^{-1}$		*a, b*
(K15)	$BrO_2^* + HAr(OH) \rightarrow HArO^* + HBrO_2$			*c*
(K16)	$HArO^* + BrO_3^- + H^+ \rightarrow Ar(OH)_2 + BrO_2^*$			*c*

[a] Herbine and Field 1980. [b] Adustable parameter chosen to give a good fit to data. [c] Not used in the present model.

In this scheme, $HAr(OH)_2$ represents pyrocatechol compound containing two phenolic groups, HAr(OH)O* is the radical obtained by hydrogen atom abstraction, $HArO_2$ is the related quinone, $BrAr(OH)_2$ is the brominated derivative, and $Ar_2(OH)_4$ is the coupling product; HAr(OH) is phenol, HArO* is the hydrogen-atom abstracted radical, and $Ar(OH)_2$ is the product.

Table 1. OKN mechanism and rate constants used in the present simulation

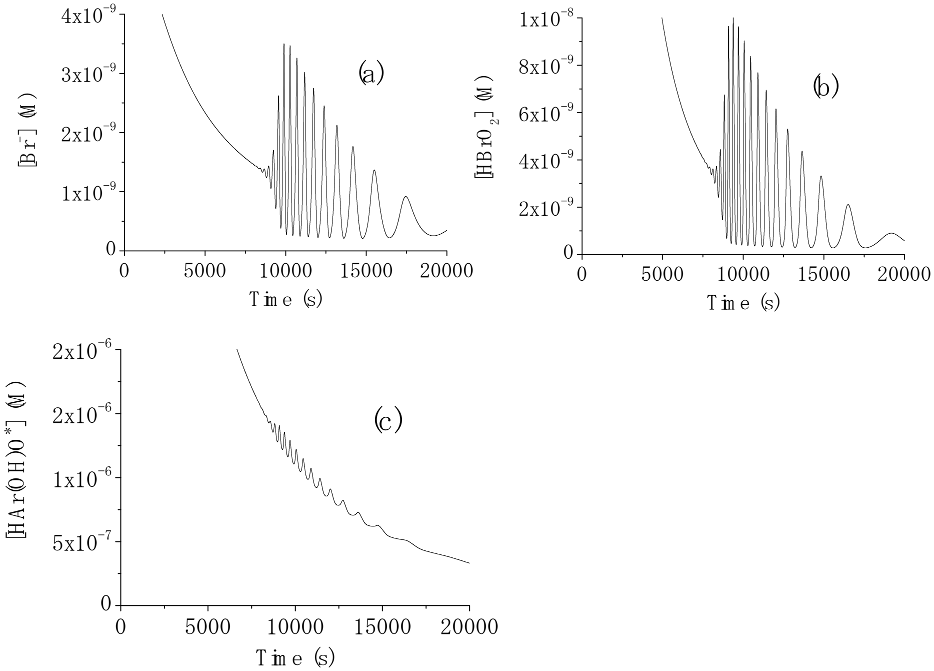


Fig. 12. Numerical simulations of oscillations in (a) Br^- (b) $HBrO_2$, and (c) pyrocatechol radical, HAr(OH)O*obtained from the present model K1 – K14 by using the rate constants listed in Table 1. The initilal concentraions were $[BrO_3^-]$=0.08 M, $[HAr(OH)_2]$=0.057 M, $[H_2SO_4]$=1.4 M, and $[Br^-]$=1.0 x 10^{-10} M; the other initial concentrations were zero.

3.2 Numerical results

Figure 12 shows oscillations in three (Br^-, $HBrO_2$, and HAr(OH)O*) of the eleven variables obtained in a simulation based on reactions K1 – K14 and the rate constant values listed in Table 1. The initial concentraions used in the simulation were $[NaBrO_3]$ = 0.08 M, $[HAr(OH)_2]$ = 0.057 M, $[H_2SO_4]$ = 1.4 M, and $[Br^-]$ = 1.0 x 10^{-10} M with the other initial concentrations to be zero with reference to those in the expreimental conditions as shown in Fig. 1. Other four variables, BrO_2*, Br*, HOBr, and Br_2, exhibited oscillations, whereas the rest variables, namely, BrO_3^-, $HAr(OH)_2$, $HArO_2$, and $BrAr(OH)_2$, did not exhibt oscillations in the present simulation.

Figure 13 shows oscillations in $[Br^-]$ at different initial concentrations of bromate: (a) 0.08 M, (b) 0.09 M, and (c) 0.1 M, with the same initial concentrations of $[HAr(OH)_2]$ = 0.057 M, $[H_2SO_4]$ = 1.4 M, and $[Br^-]$ = 1.0 x 10^{-10} M with reference to the experimental conditions as shown in Fig. 1. Although the concentration of bromate in the simulation is slightly smaller than that in the experiments, the agreement between experimentally obtained redox potential (Fig. 1) and simulated oscillations as shown in Figs. 12 and 13 is good. In particular, the induction period and the period of oscillations are similar in magnitude, as well as the degree of damping. The number of oscillations, and the prolonged period of

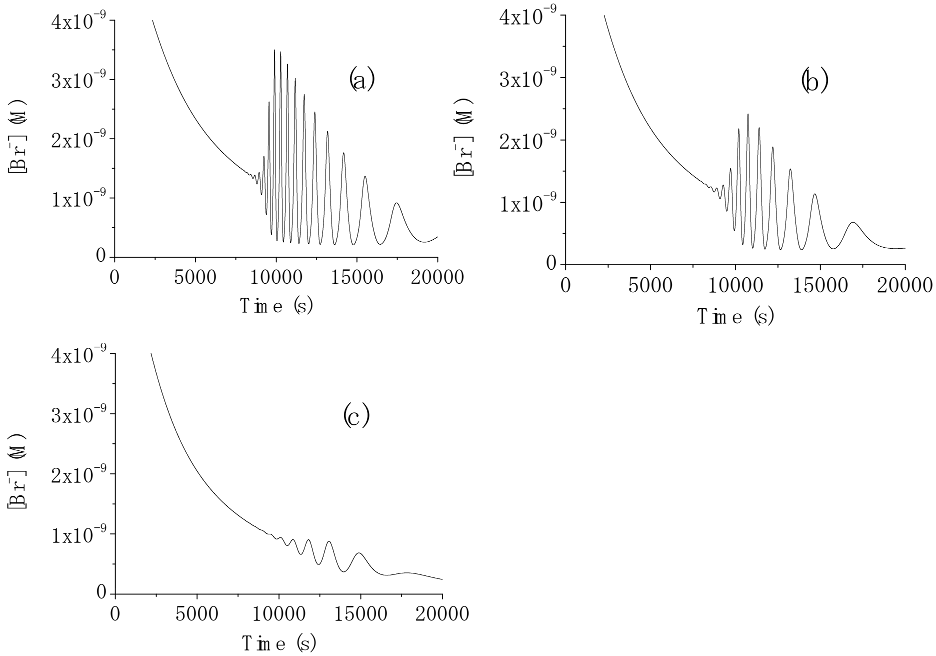


Fig. 13. Numerical simulations of the present model K1 – K14 at different initial concentrations of bromate: (a) 0.08 M, (b) 0.09 M, and (c) 0.1M. Other reaction conditions are [HAr(OH)$_2$] = 0.057 M, [H_2SO_4] = 1.4 M, and [Br$^-$]=1.0 x 10^{-10} M.

oscillations near the end of oscillations are also similar between experimental and simulated results as shown in Fig. 1 (c), Fig.3 (c), Fig.12, and Fig.13. The above simulation not only supports that the oscillatory phenomena seen in the batch system arises from intrinsic dynamics, but also provides a tempelate for further understanding the mechanism of this uncatalyzed bromate-pyrocatechol system.

While the above model is adequte in reproducing these spontaneous oscillations seen in experiments, the concentration range over which oscillations could be achieved is somehow different from what was determined in experiments. In the simulation, oscillatins were obtained in the range of 0.02 M < [BrO_3^-] < 0.1 M with [HAr(OH)$_2$] = 0.057 M and [H_2SO_4] = 1.4 M in the present simuations, whereas no oscillation could be seen in experiments for the condition of [BrO_3^-] < 0.085 M. This discrepancy of range of the reactant concentrations for exhibiting oscillations between experiments and simulations was also discerned for the concentration of HAr(OH)$_2$ under the conditions [BrO_3^-] = 0.085 M and [H_2SO_4] = 1.4 M: Oscillations were exhibited in the range of 3× 10^{-4} M < [HAr(OH)$_2$] < 0.3 M in the simulation, whereas no oscillation could be observed in experiments under [HAr(OH)$_2$] = 0.038 M as shown in Fig. 3 (a). The discrepancy in the suitable concentration range between experiment and simulation may arise from two sources: (1) the currently employed model may have skipped some of the unknown, but important reaction processes; (2) the rate constants used in the calculation are too far away from their actual value. Note that those values were original proposed for the phenol system (Herbine & Field, 1980). To shed light on this issue, we have carefully adjusted the values of the adjustable rate constants in K4, K6, K7, K9 – K14, but so far no significant improvment was achhieved.

Two other sensitive properties that can help improve the modelling are the dependence of the number of oscillations (N) and induction period (IP) on the reaction conditions. In experiments, the N value increased monotonically from 4 to 15 as bromate concentration was increased and then oscillatory behavior suddenly disappeared with the further increase of bromate concentration. In contrast, in the simulation the number of oscillations decreased gradually from 17 to 9 and then oscillatory behavior disappeared as the result of increasing bromate concentration. On the positive side, IP values increased in both experiments and simulations with respect to the increase of bromate concentration, i.e., from 9100 s to 11700 s in the experiments, and from 8000 s to 9700 s in the simulations, respectively. We would like to note that the simulated IP values firstly decreased from 12600 s to 7500 s with increase in the initial concentration of bromate from 0.03 M to 0.06 M, then increased from 7600 s to 9700 s with increase in the bromate concentration from 0.07 M to 0.11 M.

3.3 Simplification of the model

In an attempt to catch the core of the above proposed model, we have examined the influence of each individual step on the oscillatory behavior and found that reaction step K12 in Scheme II is indispensable for oscillations under the present simulated conditions as shown in Fig. 12. Such an observation is different from what has been suggested earlier steps K1 to K10 would be sufficient to account for oscillations in the uncatalyzed bromate-aromatic compounds oscillators (Orbán et al., 1979). For the Scheme II, our calculations show that while setting one of the four rate constants k_{11} to k_{14} to zero; only when k_{12} was set to zero, no oscillation could be achieved. We further tested which reaction steps could be eliminated by setting the rate constants to zero under the condition of $k_{12} \neq 0$. The results are as follows: (i) when three rate constants k_{11}, k_{13}, k_{14} were simultaneously set to zero, no oscillation was exhibited, (ii) when only one of the three rate constants was set to zero, oscillation was observed in each case, and (iii) when two of the three rate constants were set to zero, oscillations were exhibited under the condition of either $k_{13} \neq 0$ ($k_{11}=k_{14}=0$) or $k_{14} \neq 0$ ($k_{11}=k_{13}=0$) with the range of the rate constants as $3.0 \times 10^3 < k_{13}$ (M^{-1} s^{-1}) $< 2.9 \times 10^4$ and $2.2 \times 10^3 < k_{14}$ (M^{-1} s^{-1}) $< 6.0 \times 10^4$, respectively. Thus our numerical investigation has concluded that oscillations can be exhibited with minimal reaction steps as ten reaction steps in Scheme I together with a combination of two reaction steps either K12 and K13 or K12 and K14 in Scheme II.

Fig. 14 presents time series calculated under different combinations of reaction steps from scheme II. This calculation result clearly illustrates that the oscillatory behavior is nearly identical when the reaction step K11 was eliminated. Meanwhile, eliminating K13 or K14 seems to have the same influence on total number of oscillations (Fig.14 (c) ,(d)). However, chemistry of the present reaction of aromatic compounds suggests that both reaction K13 and K14 are equally important (Orbán et al., 1979). The equilibrium of step K13 is well precedented, and equimolar mixtures of quinone and dihydroxybenzene are intensely colored, and the radical HAr(OH)O* may be responsible for the color changes observed during oscillations (Orbán et al., 1979). In addition, step K14 is said to explain the observed coupling products and to prevent the buildup of quinone for further oscillations (Orbán et al., 1979).

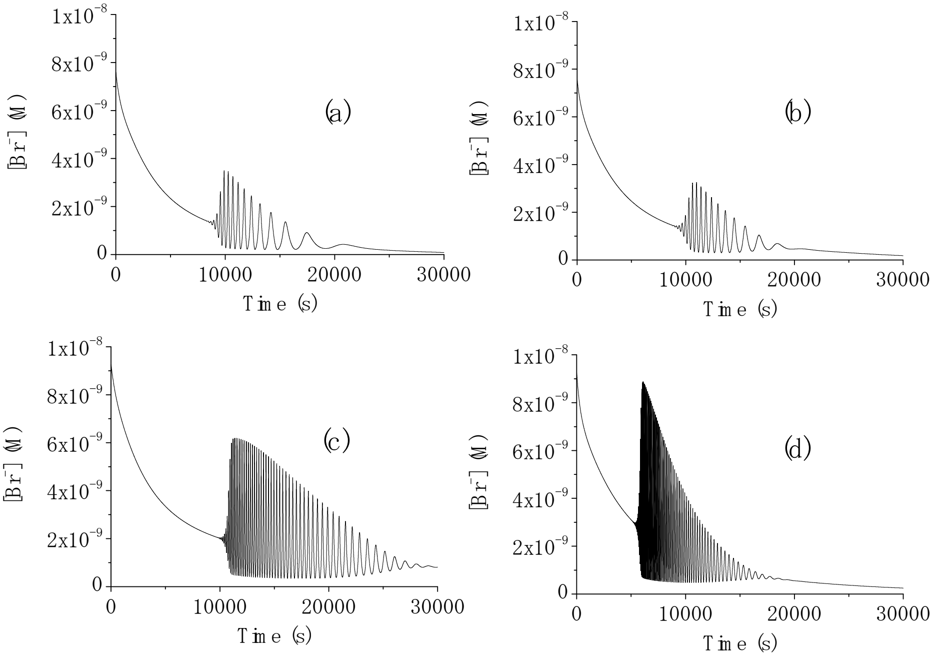


Fig. 14. Numerical simulations of the present model of K1 – K10 with different reaction steps in Scheme II: (a) K11 – K14, (b) K12 – K14, (c) K12 and K13, and (d) K12 and K14. The initilal concentraions were $[BrO_3^-] = 0.08$ M, $[HAr(OH)_2] = 0.057$ M, $[H_2SO_4] = 1.4$ M, and $[Br^-] = 1.0 \times 10^{-10}$ M as shown in Fig. 10. Note that the scales of x and y axes are different from those in Fig. 12.

In our numerical simulation, when we eliminated either step K13 or step K14, the simulated numerical results such as (i) the time series of oscillations, (ii) the initial concentration range of BrO_3^-, H_2SO_4, and $HAr(OH)_2$ for oscillations, and (iii) the dependence of the number of oscillations and induction period on the initial concentration of BrO_3^- became significantly different from those in experiments. In particular, the number of oscillations are too large under the above conditions as shown in Figs.14 (c) and (d). Such observation suggests that both K13 and K14 are important in the system studied here.

Consequently, we have concluded that the simplified model should include reaction steps K1 to K10 in Scheme I, and K12, K13, and K14 in Scheme II to reproduce the experimental results qualitatively.

3.4 Influence of reaction step K11 on the equilibrium of step K13

The numerical investigation presented in Fig. 14b suggests that reaction step K11 is not necessary for qualitatively reproducing the experimental oscillations. Besides, more positive reason for eliminatiing step K11 from the present model is that step K11 affects the range of rate constant of the equilibrium step K13 significantly. The equilibrium must lie well to the left (Orbán et al., 1979), i.e., the rate constant k_{r13} to the left must be much larger than that k_{13}

to the right. However, when we included step K11 in the model, we found no upper limit of the rate constant to the right; for instance, the rate constant can be more than 1.0×10^9 for the system to exhibit scillations under the conditions as shown in Fig.10. This value is already too large for the rate constant to the right, because we set the rate constant to the left to be 3.0×10^4 in the present simulations.

On the other hand, if we eliminated step K11 from the modelling, the range of the rate constant to the right was $0.007 < k_{13}$ (M^{-1} s^{-1}) < 0.03 for the system to exhibit oscillations, which seems to be reasonable for the equilibrium reaction step K13 to lie well to the left. Thus, this numerical analysis suggests that reaction step K11 should be eliminated from the present model.

4. Conclusions

This chapter reviewed recent studies on the nonlinear dynamics in the bromate-pyrocatechol reaction (Harati & Wang, 2008a and 2008b), which showed that spontaneous oscillations could be obtained under broad range of reaction conditions. However, when the concentration of bromate, the oxidant in this chemical oscillator, is fixed, the concentration of pyrocatechol within which the system could exhibit spontaneous oscillations is quite narrow. This accounts for the reason why earlier attempt of finding spontaneous oscillations in the bromate-pyrocatechol system had failed. As illustrated by phase diagrams in the concentration space, it is critical to keep the ratio of bromate/pyrocatechol within a proper range. From the viewpoint of nonlinear dynamics, bromate is a parameter which has a positive impact on the nonlinear feedback loop, where increasing bromate concentration enhances the autocatalytic cycle (i.e. nonlinear feedback). On the other hand, pyrocatechol involves in the production of bromide ions, a reagent which inhibits the autocatalytic process, where an increase of pyrocatechol concentration accelerates the production of bromide ions through reacting with such reagents as bromine molecules. The requirement of having a proper ratio of bromate/pyrocatechol reflects the need of having a balanced interaction between the activation cycle and inhibition process for the onset of oscillatory behaviour in this chemical system. If the above conclusion is rational, one can expect that the role that pyrocatechol reacts with bromine dioxide radicals to accomplish the autocatalytic cycle is less important than its involvement in bromide production in this uncatalyzed bromate oscillator, and therefore when a reagent such as metal catalyst is used to replace pyrocatechol to react with bromine dioxide radicals for completing the autocatalytic cycle, oscillations are still expected to be achievable. This is indeed the case. Experiments have shown spontaneous oscillations when cerium, ferroin or manganese ions were introduced into the bromate-pyrocatechol system.

Numerical simulations performed in this research show that the observed oscillatory phenomena could be qualitatively reproduced with a generic model proposed for non-catalyzed bromate oscillators. The simulation further indicates that while either two reaction steps K12 and K13 or K12 and K14 together with ten steps K1 – K10 in Scheme I in the OKN mechanism are sufficient to qualitatively reproduce oscillations, three steps K12, K13, and K14 with ten steps K1 – K10 are more realistic for representing the chemistry involving the oscillatory reactions, and also for reproducing oscillatory behaviors observed experimentally. The ratio of the rate constants for the equilibrium reaction K13 was a key reference to eliminate reaction step K11 from the original model. Although the present model still needs to be improved to reproduce the experimental results quantitatively, it has given us a glimpse that the autocatalytic production of bromous acid could be modulated periodically even in the absence of a bromide ion precursor such as bromomalonic acid in the BZ reaction. Understanding the reproduction of bromide ion appears to be a key for deciphering the oscillatory mechanism for the family of uncatalyzed oscillatory reactions of substituted-aromatic compounds with bromate and should be given particularly attention in the future research.

5. Acknowledgements

This material is based on work supported by Natural Science and Engineering Research Council (NSERC), Canada, and Canada Foundation for Innovation (CFI). JW is grateful for an invitation fellowship from Japan Society for the Promotion of Science (JSPS).

6. References

Adamčíková, L.; Farbulová, Z. & Ševčík, P. (2001) *New J. Chem.* Vol. 25, 487-490.

Amemiya, T.; Kádár, S.; Kettunen, P. & Showalter K. (1996). *Phys. Rev. Lett.* Vol. 77, 3244-3247.

Amemiya, T.; Yamamoto, T.; Ohmori, T. & Yamaguchi, T. (2002) *J. Phys. Chem. A* Vol. 106, 612-620.

Ball P. (2001) *The Self-Made Tapestry: Pattern Formation in Nature*, Oxford University Press, ISBN-10: 0198502435.

Carlsson, P.; Zhdanov, V. P. & Skoglundh, M. (2006) *Phys. Chem. Chem. Phys.* Vol. 8, 2703–2706.

Chiu, A. W. L.; Jahromi, S. S.; Khosravani, H.; Carlen, L. P. & Bardakjian, L. B. (2006) *J. Neural Eng.* Vol. 3, 9-20.

Dhanarajan, A. P.; Misra, G. P. & Siegel, R. A. (2002) *J. Phys. Chem. A* Vol. 106, 8835-8838.

Dutt, A. K. & Menzinger, M. (1999) *J. Chem. Phys.* Vol. 110, 7591-7593.

Epstein, I. R. (1989). *J. Chem. Edu.* (1989) Vol. 66, 191-195.

Epstein, I. R. & Pojman, J. A. (1998) *An Introduction to Nonlinear Chemical Dynamics*, Oxford University Press, ISBN10: 0-19-509670-3, Oxford.

Farage, V. J. & Janjic, D. (1982) *Chem. Phys. Lett.* Vol. 88. 301-304.

Field, R. J. & Burger, M. (1985) (Eds.), *Oscillations and Traveling Waves in Chemical Systems*, Wiley-Interscience, ISBN-10: 0471893846, New York.

Goldbeter, A. (1996). *Biochemical Oscillations and Cellular Rhythms*, Cambridge University Press, ISBN 0-521-59946-6, Cambridge.

Györgi, L. & Field, R. J. (1992) *Nature* Vol. 355, 808-810.

Harati, M. & Wang, J. (2008a) *J. Phys. Chem. A* Vol. 112, 4241-4245.

Harati, M. & Wang, J. (2008b) *Z. Phys. Chem. A* Vol. 222, 997-1011.

Herbine, P. & Field, R. J. (1980) *J. Phys. Chem.* Vol. 84, 1330-1333.

Horváth, J.; Szalai, I. & De Kepper, P. (2009) *Science* Vol. 324, 772-775.

Jahnke, W.; Henze C. & Winfree, A. T. (1988) *Nature* Vol. 336, 662-665.

Kádár, S.; Wang, J. & Showalter, K. (1998) *Nature* Vol. 391, 770-743.

Körös, E. & Orbán, M. (1978) *Nature* Vol. 273, 371-372.

Kumli, P. I.; Burger, M.; Hauser, M. J. B.; Müller, S. C. & Nagy-Ungvarai, Z. (2003) *Phys. Chem. Chem. Phys.* Vol. 5, 5454-5458.

Kurin-Csörgei, K.; Epstein, I. R. & Orbán, M. (2004) *J. Phys. Chem. B* Vol. 108, 7352-7358.

McIlwaine, M.; Kovacs, K.; Scott, S. K. & Taylor, A. F. (2006) *Chem. Phys. Lett.* Vol. 417, 39-42.

Mori, Y.; Nakamichi Y.; Sekiguchi, T.; Okazaki, N.; Matsumura T. & Hanazaki, I. (1993) *Chem. Phys. Lett.* Vol. 211, 421-424.

Morowitz, H. J. (2002), *The Emergence of Everything: How the World Became Complex*, Oxford University Press, ISBN-13 978-0195135138, Oxford.

Nicolis, G. & Prigogine, I. (1977) *Self-Organization in Non-Equilibrium Systems*, Wiley, ISBN 10 - 0471024015.

Nicolis, G. & Prigogine, I. (1989) *Exploring Complexity*, FREEMAN, ISBN 0-7167-1859-6, New York.

Orbán, M. & Körös, E. (1978a) *J. Phys. Chem.* Vol. 82, 1672-1674.

Orbán, M. & Körös, E. (1978b) *React. Kinet. Catal. Lett.* Vol. 8, 273-276.

Orbán, M.; Körös, E. & Noyes, R. M. (1979) *J. Phys. Chem.* Vol. 83, 3056-3057.

Sagues, F. & Epstein, I. R. (2003) Nonlinear Chemical Dynamics, *Dalton Trans.*, 1201-1217.

Scott Kelso J. A. (1995), *Dynamic Patterns: The self-organization of brain and behavior,* The MIT Press, ISBN-10: 0262611317, Cambridge, MA.

Scott, S. K. (1994) *Chemical Chaos*, Oxford University Press, ISBN 0-19-8556658-6, Oxford.

Smoes, M-L. *J. Chem. Phys.* (1979) Vol. 71, 4669-4679.

Sørensen, P. G.; Hynne, F. & Nielsen, K. (1990) *React. Kinet. Catal. Lett.* Vol. 42, 309-315.

Steinbock, O.; Kettunen, P. & Showalter K. (1995) *Science* Vol. 269, 1857-1860.

Straube, R.; Flockerzi, D.; Müller, S. C. & Hauser, M. J. B. (2005) *Phys. Rev. E.* Vol. 72, 066205-1 - 12.

Straube, R.; Müller, S. C. & Hauser, M. J. B. (2003) *Z. Phys. Chem.* Vol. 217, 1427-1442.

Szalai, I. & Körös, E. (1998) *J. Phys. Chem. A* Vol. 102, 6892-6897.

Yamaguchi, T.; Kuhnert, L.; Nagy-Ungvarai, Zs.; Müller, S. C. & Hess, B. (1991) *J. Phys. Chem.* Vol. 95, 5831-5837.

Vanag, V. K.; Míguez, D. G. & Epstein, I. R. (2006) *J. Chem. Phys.* Vol. 125, 194515:1-12.

Wang, J.; Hynne, F.; Sørensen, P. G. & Nielsen K. (1996) *J. Phys. Chem.* Vol. 100, 17593-17598.

Wang, J.; Sørensen, P. G. & Hynne, F. (1995) *Z. Phys. Chem.* Vol. 192, 63-76.

Wang, J.; Yadav, Y.; Zhao, B.; Gao, Q. & Huh, D. (2004) *J. Chem. Phys.* Vol. 121, 10138-10144.

Welsh, B. J.; Gomatam, J. & Burgess, A. E. *Nature* Vol. 304, 611-614.

Winfree, A. T. (1972) Science Vol. 175, 634-636.

Winfree, A. T. (1987) *When Time Breaks Down*, Princeton University Press, ISBN 0-691-02402-2, Princeton.

Witkowski, F. X.; Leon, L. J.; Penkoske, P. A.; Giles, W. R.; Spanol, M. L.; Ditto, W. L. & Winfree, A. T. (1998) *Nature* Vol 392, 78-82.

Zaikin, A. N. & Zhabotinsky, A. M. (1970) *Nature* Vol. 225, 535-537.

Zhao, J.; Chen, Y. & Wang, J. (2005) *J. Chem. Phys.* Vol. 122, 114514:1-7.

Zhao, B. & Wang, J. (2006) *Chem. Phys. Lett.* Vol. 430, 41-44.

Zhao, B. & Wang, J. (2007) *J. Photochem. Photobiol: Chemistry*, Vol. 192, 204-210.

6

Dynamics and Control of Nonlinear Variable Order Oscillators

Gerardo Diaz and Carlos F.M. Coimbra
University of California, Merced
U.S.A.

1. Introduction

The denomination Fractional Order Calculus has been widely used to describe the mathematical analysis of differentiation and integration to an arbitrary non-integer order, including irrational and complex orders. First proposed around three hundred years ago, it has attracted much interest during the past three decades (Oldham & Spanier (1974), Miller & Ross (1993), Podlubni (1999)). The increased interest in fractional systems in the past few decades is due mainly to a large body of physical evidence describing fractional order behavior in diverse areas such as fluid mechanics, mechanical systems, rheology, electromagnetism, quantitative finances, electrochemistry, and biology. Fractional order modeling provides exceptional capabilities for analysing memory-intense and delay systems and it has been associated with the exact description of complex transport phenomena such as fractional history effects in the unsteady viscous motion of small particles in suspension (Coimbra et al. 2004, L'Esperance et al. 2005). Although fractional order dynamical and control systems were studied only marginally until a few decades ago, the recent development of effective mathematical methods of integration of non-integer order differential equations (Charef et al. (1992); Coimbra & Kobayashi (2002), Diethelm et al. (2002); Momany (2006), Diethelm et al. (2005)) has resulted in a number of control schemes and algorithms, many of which have shown better performance and disturbance rejection compared to other traditional integer-order controllers (Podlubni (1999); Hartly & Lorenzo (2002), Ladaci & Charef (2006), among others).

Variable order (VO) systems constitute a generalization of fractional order representations to functional order. In VO systems the order of the derivative changes with respect to either the dependent or the independent variables (or both), or parametrically with respect to an external functional behavior (Samko & Ross, 1993). Compared to fractional order applications, VO systems have not received much attention, although the potential to characterize complex behavior by the functional order of differentiation or integration is clear. Variable order formulations have been utilized, among other applications, to describe the mechanics of an oscillating mass subjected to a variable viscoelasticity damper and a linear spring (Coimbra, 2003), to analyze elastoplastic indentation problems (Ingman & Suzdalnitsky (2004)), to interpolate the behavior of systems with multiple fractional terms (Soon et al., 2005), and to develop a statistical mechanics model that yields a macroscopic constitutive relation for a viscoelastic composite material undergoing compression at varying strain rates (Ramirez & Coimbra, 2007). Concerning the dynamics and control of VO systems, the authors of this chapter have previously analyzed the dynamics and linear control of a variable viscoelasticity oscillator and have presented a generalization of the van der Pol equation using the VO differential equation formulation (Diaz & Coimbra, 2009).

In the present work, we utilize the Coimbra Variable Order Differential Operator (VODOs) to analyze the dynamics of the Duffing equation with a VO damping term. Coimbra's VODO returns the correct value of the p-th derivative for p < 2, as can be generalized to any order, positive or negative.The behavior of the variable order differintegrals are shown in variable phase space for different parameters that constitute a pictorial representation of the dynamics of the variable order system, and help understand the transitional regimes between the extreme values of the derivatives. Also, a tracking controller is developed and applied to the oscillator for different expressions of the variable order q(x(t)). Finally, a variable order controller is used to eliminate chaotic oscillations of Lorenz-type systems.

2. Fractional and variable order operators

Over the past few centuries, different definitions of a fractional operator have been proposed. For instance the Riemann-Liouville integral is defined as

$$D_{0,\,t}^{-\alpha}x(t)=\frac{1}{\Gamma(\alpha)}\int_{0}^{t}(t-\tau)^{\alpha\,-\,1}x(\tau)d\tau \tag{1}$$

where $\alpha \in$ R$^+$ is the order of integration of the function *x(t)* when the lower limit of integration (initial condition) is chosen to be identically zero. The Riemann-Liouville derivative of order α is given as

$$D_{0,t}^{\alpha}x(t)=\frac{1}{\Gamma(m-\alpha)}\frac{d^m}{dt^m}\int_{0}^{t}(t-\tau)^{m-\alpha-1}x(\tau)d\tau\,, \tag{2}$$

and the Grundwald-Letnikov differential operation is defined as

$$D_{0,t}^{\alpha}x(t)=\lim_{h\to 0,nh=t}h^{-\alpha}\sum_{k=0}^{n}(-1)^k\left(\begin{smallmatrix}p\\k\end{smallmatrix}\right)x(t-kh)\,. \tag{3}$$

Finally, the Caputo derivative of fractional order α of *x(t)* is defined as

$$D_{0,t}^{\alpha}x(t)=\frac{1}{\Gamma(m-\alpha)}\int_{0}^{t}(t-\tau)^{m-\alpha-1}x^{(m)}(\tau)d\tau\,, \tag{4}$$

for which *m-1* < α <*m* $\in$ Z$^+$. More details about these operators can be found in Li & Deng (2007), Diethelm (2002), and Hartley & Lorenzo (2002).

For variable order systems, Coimbra (2003) defined the canonical differential operator as:

$$D^{q(x(t))}x(t)=\frac{1}{\Gamma(1-q(x(t)))}\int_{0+}^{t}(t-\sigma)^{-q(x(t))}D^{1}x(\sigma)d\sigma+\frac{(x(0^{+})-x(0^{-}))t^{-q(x(t))}}{\Gamma(1-q(x(t)))} \tag{5}$$

where *q(x(t))* < 1. The constraint on the upper limit of differentiation can be easily removed, and is adopted here only for convenience. One of the important characteristics of Coimbra's operator is that it is dynamically consistent with causal behavior in the initial conditions, i.e. the operator returns the appropriate Heaviside contribution to the integral value of $D^{q(x(t))}x(t)$ when $x(t)$ is not continuous between t=0- and t=0+ (Coimbra, 2003; Ramirez & Coimbra, 2007; Diaz & Coimbra (2009)). Also of relevance is that all integer and fractional order differentials are returned correctly by the operator, including the upper limit. In this work we used the extended version of this operator that covers the range of $q(x(t))$<2. The generalized order differential operator can thus be calculated by the following numerical algorithm:

$$D^q x_n = \frac{1}{\Gamma(4-q)} \sum_{i=0}^{n} a_{i,n} D^2 x_i + \frac{x(0^+)(1-q)(t_n)^{-q} + D^1 x(0^+) t_n^{\;1-q}}{\Gamma(2-q)}, \tag{6}$$

with quadrature weights given by

$$a_{i,n} = (3-q)n^{2-q} - n^{3-q} + (n-1)^{3-q} \qquad \text{, if } i=0$$

$$a_{i,n} = (n-i-1)n^{3-q} - 2(n-i)^{3-q} + (n-i+1)^{3-q} \qquad \text{, if } 0<i<n.$$

$$a_{i,n} = 1 \qquad \text{, if } i=n.$$

As stated earlier, one of the critical properties of this operator for generalized order modeling is that it returns the p-th derivative of $x(t)$ when $q(x(t)) = p$. This can be graphically demonstrated by considering an arbitrary function with known derivatives such as

$$y = t^2 \sin(t) \tag{7}$$

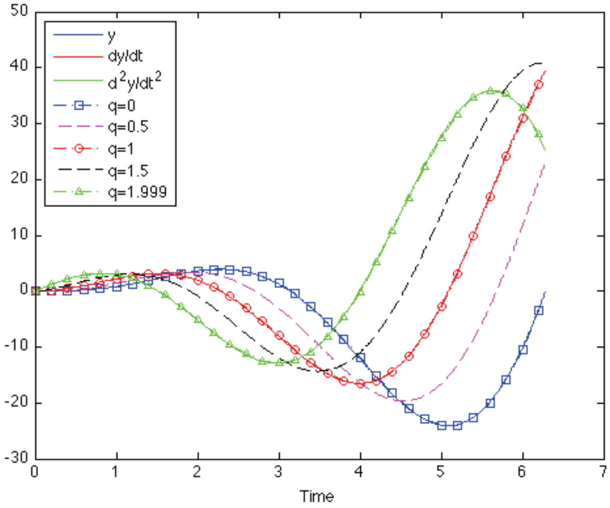


Fig. 1. Comparison of values of function y=t^2sin(t) and its derivatives with the results obtained with operator described by Eq. (6) for several values of the order q.

Figure 1 shows the values of function *y* (Eq. 7) and its derivatives *dy/dt*, and *d²y/dt²* calculated analytically. The figure also shows that the operator described by Eq. (6) returns values that match the functions *y* for *q=0*, *dy/dt* for *q=1*, and *d²y/dt²* for *q≈2*, respectively. The values of *q*=0.5 and *q*=1.5 are also shown to indicate the matching of the rational order derivatives with the values calculated using the VO operator.

3. Dynamics of the Duffing equation with variable order damping

Together with the van de Pol equation, the Duffing equation represents the behavior of one of the most studied oscillators in the field of nonlinear dynamics (Guckenheimer & Holmes (1983), Drazin (1994)). First introduced in 1918 by G. Duffing, different variations of the equation have been used to analyze its dynamics for the automomous and forced cases. Moon and Holmes (1979, 1980) considered a negative linear stiffness term to analyse the forced vibrations of a cantilever beam near two magnets. Vincent & Kenfack (2008) recently studied the bifurcation structure and synchronization of a double-well Duffing oscillator. They were able to show regions of chaos and quasiperiodicity and they found threshold parameters for which synchronization occured. With respect to fractional order systems, Sheu et al. (2007) analyzed the Duffing equation with negative linear stiffness and a fractional damping term. They reported a period doubling route to chaos in their study.

3.1 Forced oscillations

We generalize the concept of fractional damping to include a variable order term as:

$$D^2x + \delta D^q x - x + x^3 = \gamma \sin(\omega t)\,. \tag{8}$$

The main difference with respect to the work by Sheu et al. (2007) is that they studied the dynamics of Eq. (8) for a range of values of the fractional order *q* where this parameter was kept constant for every case analyzed. Here, the oscillator is generalized to include a damping term where the order of the derivative reacts to the effect of the forcing function over time, thus *q* = *q(t)*. In our analysis, we choose the value of parameters δ and ω to be 0.1 and 2, respectively.

Case γ = 1.5:

The first case considered in this work relates to the behavior of the oscillator given by Eq. (8) for γ = 1.5 for two different conditions, i.e. *q* = 1 and *q* = (99/100) + sin(ω *t*). We note that the operator described by Eq. 6 is valid for *q(t)* < 2, thus the expression used for the change in *q* with respect to time ensures that this condition is met.

Figure 2 shows the dynamics of the oscillator given by Eq. (8) for *q* = 1 as the order of the derivative in the damping term. The simulations cover the time range $t \in [0, 700]$ where only the results for *t* > 200 are plotted to exclude the initial transients. Chaotic behavior is observed and a strange attractor is depicted in Fig. 2(a). The Poincaré map is shown in Fig. 2(b).

The effect of the variable order derivative on the damping term of Eq. (8) significantly changes the dynamics of the oscillator. This can be observed in Figs. 3(a) and 3(b) where it is seen that after removing the intial transients, the dynamics of the oscillators are confined to a narrower region in the phase space.

The dynamics of the VO oscillators can also be analyzed utilizing a modified version of the phase diagram where the variable order derivative, $D^q x(t)$, is plotted on the ordinate axis and the position, $x(t)$, is plotted on the abcisa axis. Figure 4(a) shows the variable order phase space (a plot of the value of the VO derivative, $D^q x(t)$, as a function position), whereas Fig. 4(b) shows the behavior of $D^q x(t)$ as a function of the order of the derivative, *q(t)*. It is seen in Fig. 4(b) that *q(t)* < 2, thus meeting the upper limit of differentiation mandated by the numerical algorithm used here (Eq. 6).

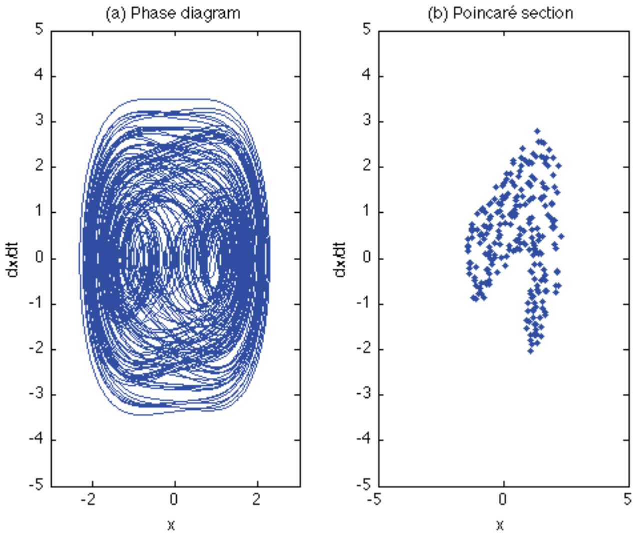


Fig. 2. Phase diagram and Poincare map for $\gamma = 1.5$ and $q = 1$.

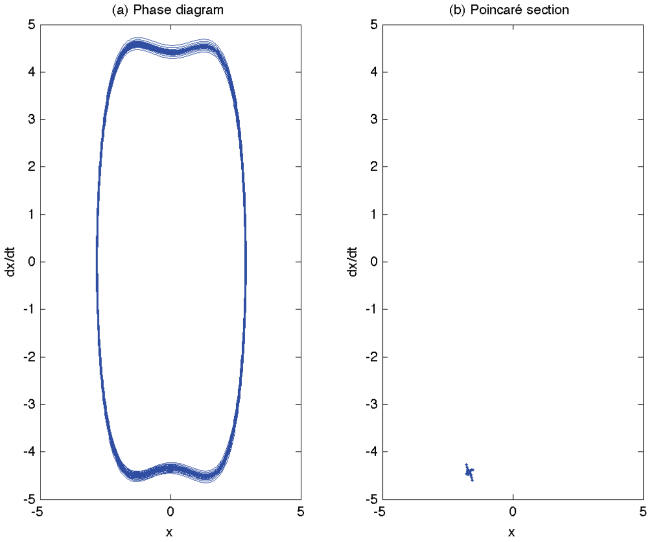


Fig. 3. Phase diagram and Poincare map for $\gamma = 1.5$ and $q = (99/100) + \sin(\omega t)$.

Figure 5(a) shows the change of *x(t)* and $D^q x(t)$ as a function of time. Figures 6(a) and 6(b) show that *q(t)* also has an oscillatory behavior with $D^q x(t)$ having a minimum value when *x(t)* and *q*(*t*) approach their maximum value. This is also depicted in the VO phase diagrams shown in Figs. 4(a) and 4(b).

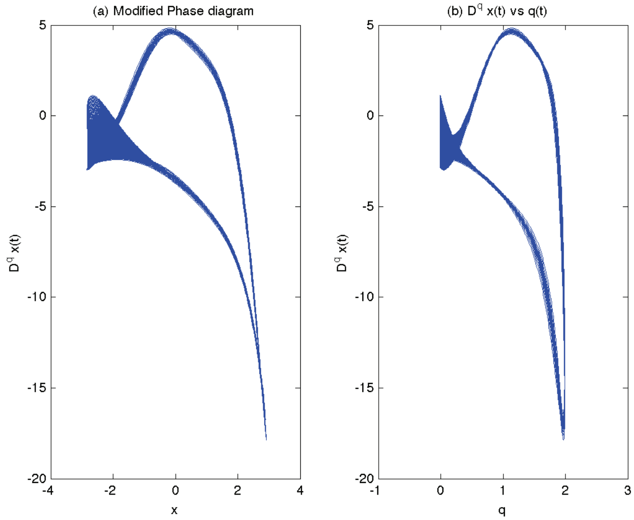


Fig. 4. Modified phase diagram and $D^q x(t)$ vs. *q(t)* plots for γ=1.5.

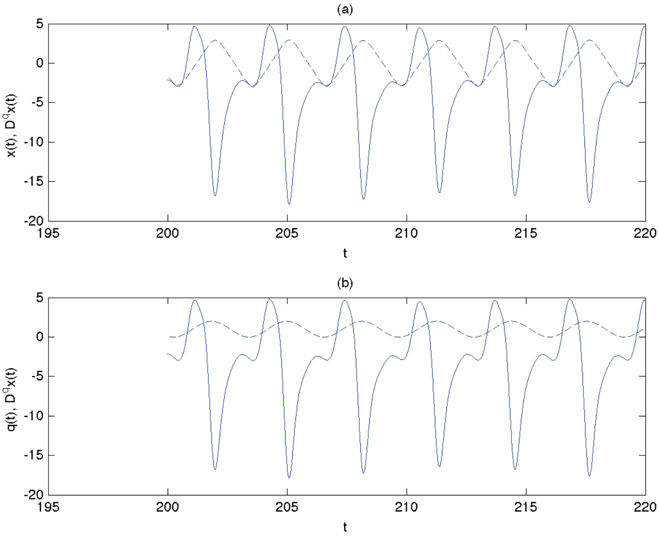


Fig. 5. Dynamics of VO Duffing equation with respect to time for γ=1.5. (a) - - - *x(t),* ____ = $D^q x(t)$; (b) - - - *q(t),* ____ = $D^q x(t)$;

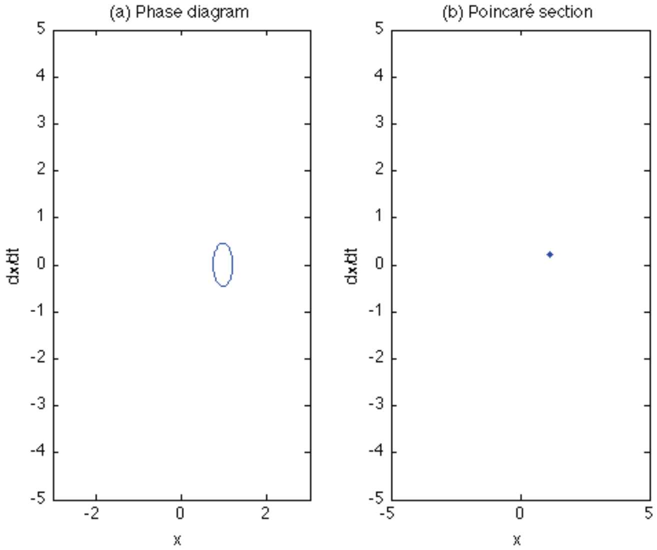


Fig. 6. Phase diagram and Poincare map for γ=0.5 and q=1.

Case γ = 0.5:

We now analyze the case where parameter γ = 0.5. After the initial transient, the standard configuration (q = 1) shows an oscillatory behavior as depicted in Fig. 6(a) with a single point appearing in the Poincare map, Fig. 6(b).

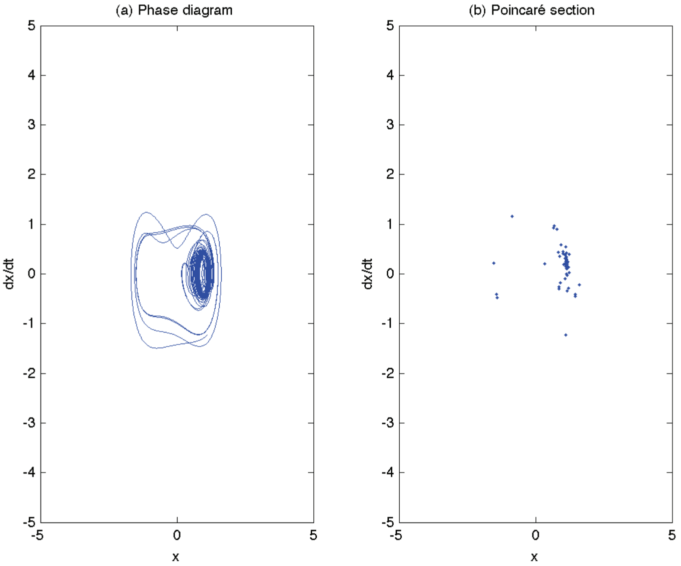


Fig. 7. Phase diagram and Poincare map for γ = 0.5 and q = (99/100)+ sin(ωt) for t > 200.

Figures 7(a) and 7(b) show the results of the simulations for $\gamma = 0.5$ and a variable order of the derivative given by $q(t) = (99/100) + \sin(\omega\, t)$. It is seen that the phase diagram and Poincare maps differ significantly from the case $q = 1$. However, plotting $x(t)$ as a function of time, as depicted in Fig. 8, shows the transient effects seem to last longer than for the case of $q = 1$. After $t \sim 400$, the system settles to an oscillatory behavior with a smaller amplitude.

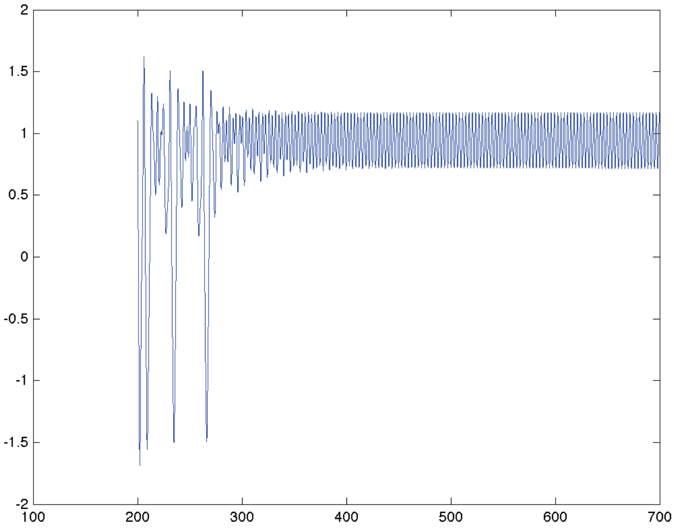


Fig. 8. Phase diagram and Poincare map for $\gamma = 0.5$ and $q = (99/100) + \sin(\omega\, t)$ for $t > 200$.

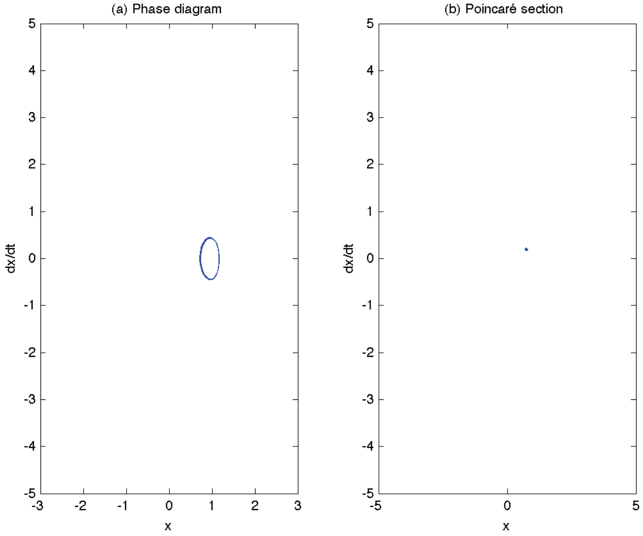


Fig. 9. Phase diagram and Poincare map for $\gamma = 0.5$ and $q = (99/100) + \sin(\omega\, t)$ for $t > 400$.

Plots of the phase diagram and the Poincare map for $t > 400$ are shown in Figs. 9(a) and 9(b), respectively. Similar dynamics compared to $q = 1$ are displayed by the system.

3.2 Control of the VO Duffing equation

The dynamics of the variable order Duffing equations were analyzed in the previous section for the cases $\delta = 0.1$, $\omega = 2$, with $\gamma = 1.5$ and $\gamma = 0.5$, respectively. In this section, we study controls aspects of this equation subject to a VO damping term. An exact feedback linearization is performed to obtain a tracking controller that drives the VO Duffing oscillator to follow a periodic reference function, r (Khalil, 1996). The forcing function in Eq. (8) can be replaced by a control action as shown by Eq. 9.

$$D^2 x = x - x^3 - \delta D^q x + u . \tag{9}$$

Exact feedback linearization is obtained by choosing the control action

$$u = x^3 + \delta D^q x + v . \tag{10}$$

Thus, Eq. 9 is converted to a linear equation of the form

$$D^2 x = x + v . \tag{11}$$

This second order differential equation is transformed to a system of first order differential equations

$$\begin{gathered} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = A\vec{x} + Bv = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v, \\ y = C\vec{x} = [1 \quad 0]\vec{x}. \end{gathered} \tag{12}$$

A control action of form $u = -K\vec{x} + Gr = -k_1 x_1 - k_2 x_2 + Gr$ is chosen where k_1 and k_2 are constants that are used to select the location of the closed-loop eigenvalues, G is the feedforward gain, and r is the reference. For the controllable system given by Eq. (12) we arbitrarily select closed-loop egivenvalues $\lambda_{1,2}$=-5 to obtain $k_1 = 24$ and $k_2 = 10$. The feedforward gain is obained with Eq. (13) (Williams & Lawrence, 2007).

$$G = -(C(A - BK)^{-1} B)^{-1} . \tag{13}$$

The tracking scheme is tested with the variable order derivative in the VO damping term having the expression $q = (99/100) + \sin(\omega\, t)$, where $\gamma = 1.5$ and $\omega = 2$. Figures 10(a) to 10(d) show the behavior of the tracking system for $r(t) = 2\cos(\omega/10) + \sin(3\omega/10)$. The ouput of the system, $y(t)$, follows the reference, r(t), consistently, as seen in Fig. 10(a). Figure 10(b) shows the control action, $u(t)$, and the sinusoidal behavior of the order of the VO derivative, $q(t)$, is shown in Fig. 10(c) where the value of the variable order derivative, $D^q y(t)$, is plotted in Fig. 10(d).

Exact feedback linearization can be used for different functions of $q(t)$. Figure 11(a) to 11(d) show the tracking of reference r for $q(t) = r(t)/3$. Scaling of $q(t)$ with respect to r is performed so that the value of $q(t)$ remains smaller than 2.

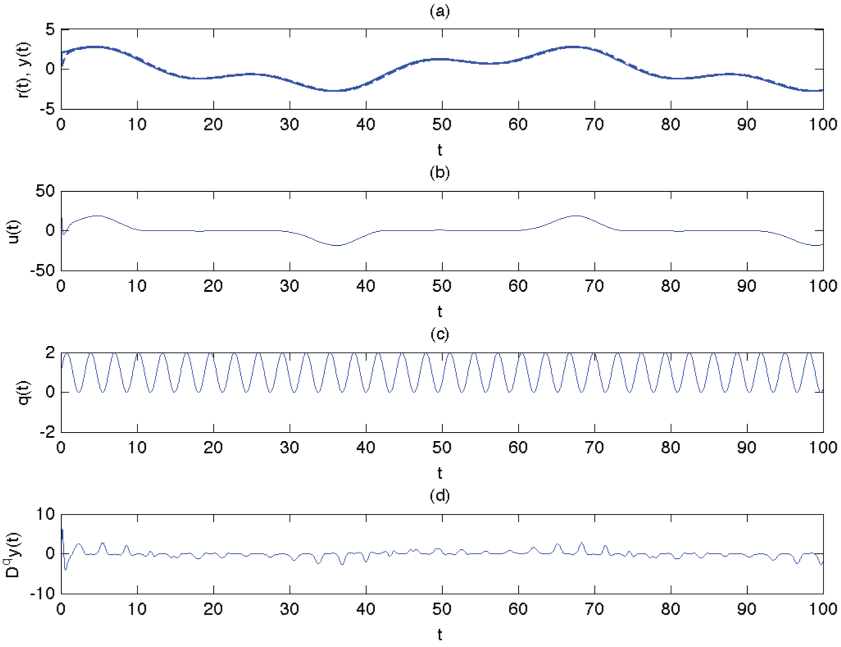


Fig. 10. Tracking control for the VO duffing equation for *q(t)*= (99/100)+ sin(ωt). (a) __ = *r*(*t*), . . .=*y*(*t*); (b) *u*(*t*), (c) *q(t)*, and (d) *D^q^y(t)*.

We note that if the value of the order of the VO derivative, q(t), is known to remain within the requirement of the operator (i.e. *q*(t)< 2) then an implicit form of the variation of q (i.e. *q*=*q(x)*) can also be utilized (Diaz & Coimbra, 2009). It is also mentioned that if the closed-loop eigenvalues are chosen to have positive real parts then the system becomes unstable.

4. VO control of the Lorenz system

So far, we have analyzed the dynamics and control of VO systems that have the term $D^q x(t)$ as part of the expression describing their dynamics. We now apply the variable order approach as the control action to stabilize a chaotic dynamical system. First proposed as a way to discribe the dynamics of weather systems, the Lorenz system of equations (Lorenz, 1963) has been intensively studied as a dynamical system that displays chaotic behavior where a strange attractor is encountered under certain values of its parameters. Control techniques have been proposed in the past (Vincent & Yu, 1991) but to the best knowledge of the authors, there is no study in the literature that has utilized a variable order controller to stabilize the chaotic dynamics of the Lorenz system.

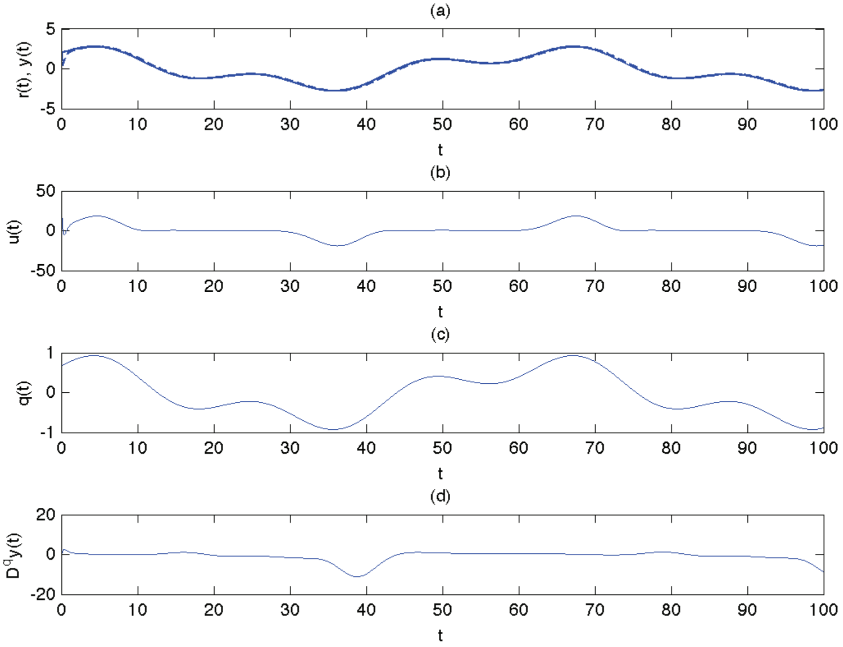


Fig. 11. Tracking control for the VO duffing equation for *q(t)*=(1/3) [2cos(ω/10)+sin(3ω/10)]. (a) __ = *r*(*t*), . . .= *y*(*t*); (b) *u*(*t*), (c) *q(t)*, and (d) *D^qy(t)*.

The Lorenz system is described by the folowing equations

$$\frac{dx_1}{dt} = -\sigma x_1 + \sigma x_2,$$
$$\frac{dx_2}{dt} = rx_1 - x_2 - x_2 x_3, \qquad (14)$$
$$\frac{dx_3}{dt} = x_1 x_2 - bx_3 + u.$$

For $r > 1$ there are two non-trivial equilibrium points, i.e. $\bar{x}_1 = \bar{x}_2 = \pm\ (b\ (r-1))^{1/2}$, $\bar{x}_3 = r-1$. Linearizing the system with respect to the first non-trivial equilibrium point, we obtain

$$\frac{dz_1}{dt} = -\sigma z_1 + \sigma z_2,$$
$$\frac{dz_2}{dt} = z_1 - z_2 - \sqrt{b(r-1)}z_3, \qquad (15)$$
$$\frac{dz_3}{dt} = \sqrt{b(r-1)}z_1 + \sqrt{b(r-1)}z_2 - bz_3 + u^*,$$

which can be written as $\frac{dz}{dt} = Az + Bu^*$, where

$$\begin{aligned} z_1 &= x_1 - \sqrt{b(r-1)}, \\ z_2 &= x_2 - \sqrt{b(r-1)}, \\ z_3 &= x_3 - (r-1). \end{aligned} \tag{16}$$

Tavazoei et al. (2009) developed a control strategy using a fractional order controller with three parameters that is used to suppress chaos. They showed that a chaotic system is stabilized using the single control input *u(t)=J^q^y(t)*, where J^q is a fractional integral operator and *y(t)* = *-(μT₁+νT₃)(x(t)-x*)*, and where T_1 and T_3 are the first and third row of a transformation matrix such that

$$\bar{A} = TAT^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a & -b & -c \end{bmatrix}, \qquad \bar{B} = TB = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{17}$$

where the parameters *a,b,c* are the coefficients of the characteristic polynomial of the Jacobian matrix *A*

$$s^3 + cs^2 + bs + a = 0. \tag{18}$$

Tavazoei et al. (2009) also showed that for the integral fractional operator with -1 < q < 0 the controller stabilizes the system when

$$0 < \mu < \frac{cb^{(1-q/2)}}{\cos(-q\pi/2)}; \quad \nu > \frac{ab^{(-1-q/2)}}{\cos(-q\pi/2)}. \tag{19}$$

We use the VO operator described by Eq. (5) with a negative value of *q* (i.e. integral variable order operator) to suppress chaos of the Lorenz system. Choosing $\sigma = 10$, $b = 8/3$, $r = 28$ and $q = -0.2$, we obtain $0 < \mu < 2310.9$ and $\nu > 23.7$. Arbitrary values of $\mu = 23.1$ and $\nu = 26.1$ are chosen that satisfy the constraints given by Eq. (19). Figure 12(a) depicts the chaotic behaviour displayed by the Lorenz system for $t < 25$. At $t = 25$, the controller is turned on and the system is stabilized around the selected equilibrium point. Figure 12(b) shows the values of the control action, *u*(t). In this case *q* has been considered constant for the VO operator.

The variable order capability of the controller can be verified by running a similar case where the parameters μ and ν are kept constant and the order of the VO derivative is changed. The controller works until the constraints given by Eq. (19) are no longer met. Fixed values for μ and ν are used. However, for $t > 25$ the order of the VO derivative *q(t)* is monotonically decreased starting from $q = -0.2$. Figure 13(a) shows the behaviour of the system subject to the control action u(t) shown in Fig. 13(b). It is observed that once the controller is turned on ($t > 25$) stabilization of the chaotic system is obtained for variable *q* until parameters μ and ν fall outside of the constraints. Figure 13(c) shows the variation of *q* over time. The controller reaches a point where it no longer stabilizes the chaotic behaviour of the system. This situation is resolved by re-calculating the values of μ and ν for the VO

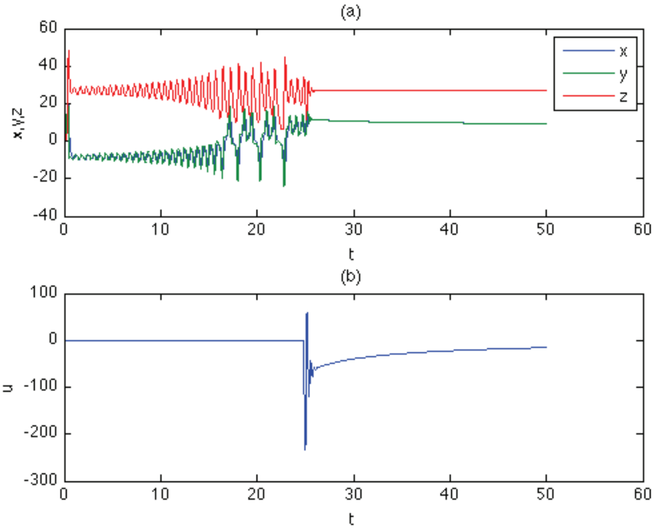


Fig. 12. Chaos suppression in the Lorenz system with $\sigma = 10$, $b = 8/3$, $r = 28$, q =-0.2, and fixed values of μ and ν in VO operator in Eq. (5). (a) *x, y, z* vs *t* (b) *u* vs *t*.

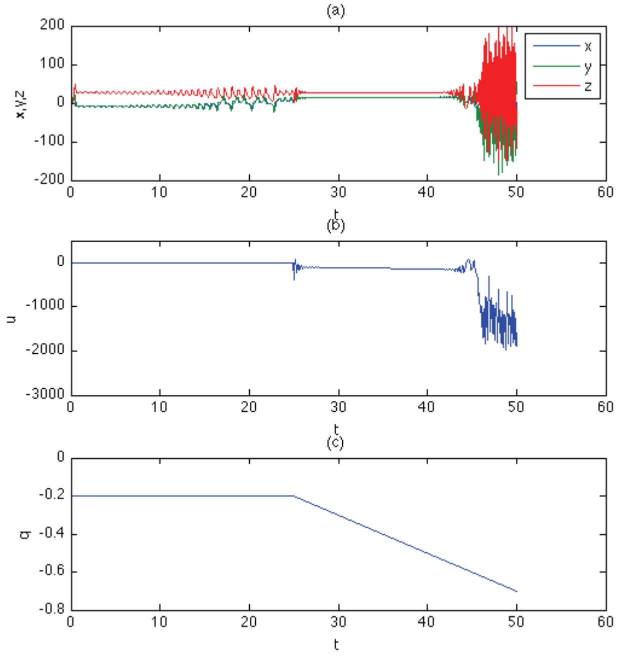


Fig. 13. Performance of controllers for fixed values of μ and ν and decreasing value of *q(t)*. (a) *x, y, z* vs *t* (b) *u* vs *t*, (c) *q* vs *t*.

value of q to remain within the required constraints. Figure 14(a) shows that the controller stabilizes the chaotic system under the variation of q with respect to time shown in Fig. 14(c) that generates the control action displayed in Fig. 14(b). The variation in the values of μ and ν is observed in Fig. 14(d) that shows that as q decreases the values of μ and ν also increase rapidly.

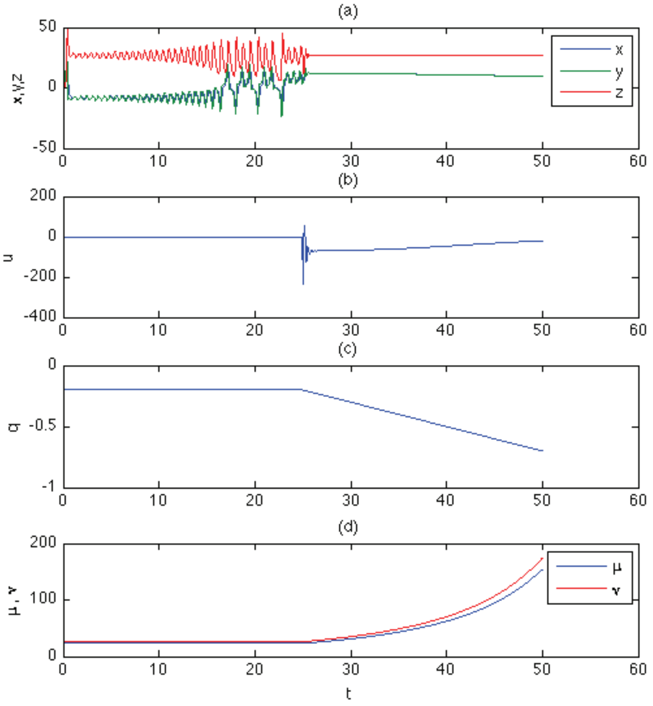


Fig. 14. Performance of controllers for variable values of μ and ν and decreasing value of *q(t)*. (a) *x, y, z* vs *t* (b) u vs t, (c) *q* vs *t*, (d) *μ, ν* vs *t*.

Grigorenko and Grigorenko (2003) have shown that the generalized fractional order Lorenz system also presents chaotic behaviour. Clearly, a VO controller technique as presented here can also be utilized to suppress chaos in such a system.

5. Conclusion

Variable order systems, i.e. systems where the order of the derivative changes with respect to either the dependent or the independent variables have not received as much attention as fractional order systems, despite of the ability of variable order formulations to model continuous spectral behavior in complex dynamics. We illustrate some of the characteristics of variable order systems and controllers through the numerical simulation of nonlinear dynamic oscillators and systems of equations. In this work, we analyze the dynamics of a modified Duffing equation, which includes a variable order derivative as the damping term, and illustrate its behavior as compared to the classical Duffing equation. Exact feedback linearization is used to derive a linear controller of the Duffing equation with variable order damping. Finally, a variable order controller is used to suppress chaos on the Lorenz system of equations. To the best knowledge of the authors, this is the first time a variable order controller is described.

6. References

Charef, A.; Sun, H.H.; Tsao, Y.Y. & Onaral, B. (1992) Fractal system as represented by singularity function, IEEE Transactions on Automatic Control, 37(9) 1465–1470.

Coimbra, C.F.M & Kobayashi, M.H. (2002). On the Viscous Motion of a Small Particle in a Rotating Cylinder. Journal of Fluid Mechanics (469) pp. 257-286.

Coimbra, C.F.M. (2003) Mechanics with variable-order differential operators, Annalen der Physik, 12(11-12) 692–703.

Coimbra, C.F.M.; L'Esperance, D.; Lambert, A., Trolinger, J.D. & Rangel, R.H., (2004) An experimental study on the history effects in high-frequency Stokes flows, Journal of Fluid Mechanics (504) 353–363.

Diaz, G. & Coimbra, C.F.M. (2009) Nonlinear dynamics and control of a variable order oscillator with application to the van der Pol equation, Nonlinear Dynamics, 56: 145-157.

Diethelm, K.; N. J. Ford, N.J. & Freed, A.D. (2002) A predictor-corrector approach for the numerical solution of fractional differential equations, Nonlinear Dynamics, 29(2002) 3–22.

Diethelm, K.; Ford, N.J.; Freed, A.D.& Luchko, Y. (2005) Algorithms for the fractional calculus: A selection of numerical methods, Computational Methods in Applied Mechanics and Engineering, 194 (6-8) 743-773.

Drazin, P.G. (1994) Nonlinear Systems, Cambridge Texts in Applied Mathematics, Cambridge University Press, UK.

Grigorenko, I & Grigorenko, E. (2003) Chaotic dynamics of the fractional Lorenz system. Physical Review Letters 91(3) 034101-1-0.4101-4.

Guckenheimer, J. & Holmes, P. (1983) Nonlinear Oscillators, Dynamical Systems, and Bifurcations of Vector Fields, Applied Mathematical Sciences 42, Spriner-Verlag, New York

Hartley, T.T. & Lorenzo, C.F. , Dynamics and control of initialized fractional-order systems, Nonlinear Dynamics, 29(2002) 201–233.

Ingman, D. & Suzdalnitsky, J., Control of damping oscillations by fractional differential operator with time-dependent order, Computer Methods in Applied Mechanics and Engineering, 193(2004), 5585–5595.

Khalil, H.K. (1996) Nonlinear Systems. Prentice Hall, 2nd Ed, USA. ISBN 0-13-228024-8.

Ladaci, S. and Charef, A. (2006), On fractional adaptive control, Nonlinear Dynamics, 43365–378.

L'Esperance, D.; Coimbra, C.F.M.; Trolinger, J.D. & Rangel, R.H. (2005) Experimental verification of fractional history effects on the viscous dynamics of small spherical particles, Experiments in Fluids (38) 112-116.

Li, C. & Deng, W., Remarks on fractional derivatives, Applied Mathematics and Computation, 187(2007) 777–784.

Miller, K.S. & Ross,B.(1993) An Introduction to the Fractional Calculus and Fractional Differential Equations, John Wiley and Sons, New York, NY.

Momani, S. (2006) A numerical scheme for the solution of multi-order fractional differential equations, Applied Mathematics and Computation, 182 761–770.

Moon, F.C. & Holmes, P.J. (1979) A magnetoelastic strange attractor. J Sound Vib, 65(2) 285-296.

Moon, F.C. & Holmes, P.J. (1980) Addendum: A magnetoelastic strange attractor, J Sound Vib, 69(2) 339.

Oldham, K.B. & Spanier, J. (1974) The Fractional Calculus, Academic Press, New York, NY.

Podlubni, I. (1999) Fractional Differential Equations, Academic Press, San Diego, CA.

Podlubni, I., Fractional-order systems and P I λ Dμ -Controllers, IEEE Transactions on Automatic Control, 44(1)(1999) 208–214.

Ramirez, L.E.S. & Coimbra, C.F.M. (2007) A Variable Order Constitutive Relation for Viscoelasticity Annalen der Physik (16), No. 7-8, pp. 543-552.

Samko, S.K. & Ross, B., Integration and differentiation to a variable fractional order, Integral Transforms and Special Functions, 1 (4) (1993) 277–300.

Sheu, L-J; Chen, H-K, Chen, J-H; Tam & L-M (2007) Chaotic dynamics of the fractionally damped Duffing equation, Chaos Solitons & Fractals, 32 1459-1468.

Soon, C. M., Coimbra, C.F.M., & Kobayashi, M. H. (2005). "The Variable Viscoelasticity Oscillator" Annalen der Physik, (14) N.6, pp. 378-389.

Vincent, U.E. & Kenfack, A. (2008) Synchronization and bifurcation structures in coupled periodically forced non-identical Duffing oscillators, Physica Scripta, 77 045005 (7pp).

Williams, R.L. & Lawrence, D.A. (2007) Linear State-Space Control Systems. John Wiley and Sons, USA. ISBN 978-0-471-73555-7.

7

Nonlinear Vibrations of Axially Moving Beams

Li-Qun Chen
Shanghai University
China

1. Introduction

Axially moving beams can represent many engineering devices, such as band saws, power transmission belts, aerial cable tramways, crane hoist cables, flexible robotic manipulators, and spacecraft deploying appendages. Despite usefulness and advantages of these devices, vibrations associated with the devices have limited their applications. Therefore, understanding transverse vibrations of axially moving beams is important for the design of the devices. The investigations on vibrations of axially moving beams have theoretical importance as well, because an axially moving beam is a typical representative of distributed gyroscopic systems. The term "gyroscopic" arose in recognition of an early problem in gyrodynamics. Actually, the Coriolis acceleration component experienced by axially moving materials imparts a skew-symmetric or gyroscopic term to their governing equations. Due to particular characteristics of the gyroscopic term, the approaches developed in analyzing vibrations of an axially moving string can be applied to other more complicated distributed gyroscopic systems. Because of the practical and theoretical significance, the investigation on nonlinear vibrations of axially moving beams is a challenging subject which has been studied for many years and is still of interest today.

The relevant researches on transverse vibrations of axially moving strings can be dated back to (Aiken, 1878). There are several excellent and comprehensive survey papers, notably (Mote, 1972), (Ulsoy and Mote, 1978), (Mote et al., 1982), (D'Angelo et al., 1985), (Wickert and Mote, 1988), (Wang and Liu, 1991), (Abrate, 1992), (Zhu, 2000), reviewing the state-of-the-art in different time phases of investigations related to vibrations of axially moving beams. The present chapter emphasizes on the recent achievements, although some early results are mentioned for the sake of completeness. Besides, the chapter focuses the nonlinear problem only. If the vibration amplitude is large, the nonlinearity should be taken into account. Hence the chapter, unlike (Chen, 2005a) for axially moving strings, is not a comprehensive survey with a complete and detailed representation of current researches. Instead, the chapter is a counterpart of (Chen et al., 2008) for axially moving beams. The author tries to put the some available results into a general framework, as well as to highlight the work of the author and his collaborators. It is hoped that the chapter serves as a collection of ideas, approaches, and main results in investigations on nonlinear vibration of axially moving beams.

The chapter is organized as follows: Section 2 focuses on the mathematical models of nonlinear transverse vibration. The special attentions are paid to the comparison of two different nonlinear models and the introduction of the material time derivative into the viscoelastic constitutive relations. Section 3 covers the developments and the applications of approximately analytical methods, including the asymptotic method, the Lindstedt-Poincaré method, the method of normal forms, the method of nonlinear, complex modes, the method of multiple scales, and the incremental harmonic balance method. Section 4 is devoted to the numerical approaches, including the Galerkin discretization, the finite difference, and the differential quadrature. Section 5 reveals the nonlinear behaviors such as bifurcation and chaos based on the numerical solutions. Section 6 discusses energetics, conserved quantity and the applications. The final section recommends future research directions.

2. Governing equations

2.1 Coupled vibration

The governing equation is the base of all analytical or numerical investigations. Generally, an axially moving beam undergoes both the longitudinal vibration and the transverse vibration, and they are coupled. (Thurman & Mote, 1969) obtained the governing equation of coupled longitudinal and transverse vibrations of an axially moving beam. (Koivurova & Salonen 1999) revisited the same modeling problem and clarified its kinematic aspects. Their nonlinear formulation for the moving beam problem has two limitations: the material of the beam is linear elastic constituted by Hooke's law, and the axial speed of the beam is a constant. As (Wickert & Mote 1988) pointed out, modeling of dissipative mechanisms is an important vibrations analysis topic of axially moving materials. An effective approach is to model the beam as a viscoelastic material. Therefore, it is necessary to deal with constitutive laws other than Hooke's law. Axial transport acceleration frequently appears in engineering systems. For example, if an axially moving beam models a belt on a pair of rotating pulleys, the rotation vibration of the pulleys will result in a small fluctuation in the axial speed of the belt. The nonlinear model in (Thurman & Mote, 1969) for coupled vibration can be generalized to an axially accelerating viscoelastic beam as follows.

Consider a uniform axially moving beam of density ρ, cross-sectional area A, moment of inertial I, and initial tension P_0, as shown in Figure 1. The beam travels at the uniform transport speed γ between two boundaries separated by distance l. Assume that the deformation of the beam is confined to the vertical plane. A mixed Eulerian-Lagrangian description is adopted. The distance from the left boundary is measured by fixed axial coordinate x. The beam is subjected to an external excitation $f_u(x,t)$ and $f_v(x,t)$ in longitudinal and transverse directions respectively, where t is the time. The in-plane motion of the beam is specified by the longitudinal displacement $u(x,t)$ related to coordinate translating at speed γ and the transverse displacement $v(x,t)$ related to the spatial frame.

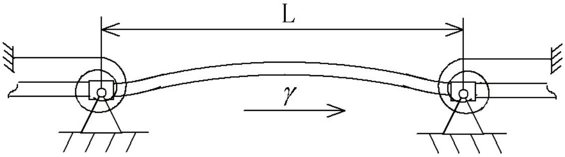


Fig. 1. The physical model of an axially accelerating beam

Study the motion of the beam in a reference frame moving in the axial direction and at speed γ. The reference system is not an inertial frame if γ is not a constant. The beam is a one-dimensional continuum undergoing an in-plane motion in the moving reference frame, the Eulerian equation of motion of a continuum gives

$$\begin{aligned}
\rho\frac{\mathrm{d}^2 u}{\mathrm{d}t^2} &= \frac{\partial}{\partial x}\left(\frac{1}{A}\frac{(P_0 + A\sigma)(1+u,_x)}{\sqrt{(1+u,_x)^2 + v,_x^2}}\right) - \rho\dot{\gamma} + \frac{f_u(x,t)}{A}, \\
\rho\frac{\mathrm{d}^2 v}{\mathrm{d}t^2} &= \frac{\partial}{\partial x}\left(\frac{1}{A}\frac{(P_0 + A\sigma)v,_x}{\sqrt{(1+u,_x)^2 + v,_x^2}}\right) - \frac{M,_{xx}(x,t)}{A} + \frac{f_v(x,t)}{A},
\end{aligned} \tag{1}$$

where a comma preceding *x* or *t* denotes partial differentiation with respect to *x* or *t*, $\sigma(x,t)$ is the axial disturbed stress, and *M*(*x*,*t*) is the bending moment. The viscoelastic material of the beam obeys the Kelvin model, with the constitution relation

$$\sigma(x,t) = \left(E + \eta\frac{\mathrm{d}}{\mathrm{d}t}\right)\varepsilon_{\mathrm{N}}(x,t), \tag{2}$$

where, *E* is the Young's modulus, η is the dynamic viscosity, and the disturbed strain $\varepsilon_{\mathrm{N}}(x, t)$ of the beam is given by the nonlinear geometric relation

$$\varepsilon_{\mathrm{N}} = \sqrt{(1+u,_x)^2 + v,_x^2} - 1 \tag{3}$$

For a slender beam (for example, with $I/(Al^2)<0.001$), the linear moment-curvature relationship of Euler-Bernoulli beams is sufficiently accurate,

$$M(x,t) = \left(E + \eta\frac{\mathrm{d}}{\mathrm{d}t}\right)Iv,_{xx} \tag{4}$$

In the moving reference frame, the beam itself is without any axial transportation, while the boundaries are moving at speed γ. The axially moving beam is constrained by rotating sleeves with rotational springs (Chen & Yang, 2006a). The stiffness constant of two springs is the same, denoted as *K*. Nullifying the transverse displacements and balancing the bending moment at both ends lead to the boundary conditions

$$u(s,t) = 0, u(l+s,t) = 0,; \tag{5}$$

$$v(s,t) = 0, EIv,_{xx}(s,t) - Kv,_x(s,t) = 0; v(l+s,t) = 0, EIv,_{xx}(l+s,t) + Kv,_x(l+s,t) = 0. \tag{6}$$

where $\dot{s} = \gamma$. To avoid the moving boundary conditions (5), which are difficult to tackle, the transformation of coordinates is introduced as follows

$$x \leftrightarrow x + s, t \leftrightarrow t. \tag{7}$$

Then, expressed in the new coordinates, the boundary conditions have a simpler form

$$u(0,t) = 0, u(l,t) = 0; \tag{8}$$

$$v(0,t) = 0, EIv,_{xx}(0,t) - Kv,_x(0,t) = 0; v(l,t) = 0, EIv,_{xx}(l,t) + Kv,_x(l,t) = 0. \tag{9}$$

Under the new coordinates, the partial derivatives with respect to *x* and *t* remain invariant, and the total time derivative changes as follows

$$\frac{\mathrm{d}}{\mathrm{d}t} \leftrightarrow \gamma\frac{\partial}{\partial x}+\frac{\partial}{\partial t} \tag{10}$$

Substitution of equations (2), (4), and (10) into equation (1) yields

$$\begin{aligned}
&\rho A\left(u_{,tt}+2\gamma u_{,xt}+\dot{\gamma}\left(1+u_{,x}\right)+\gamma^{2}u_{,xx}\right)= \\
&\frac{\partial}{\partial x}\left(\frac{\left[P_{0}+A\left(E\varepsilon_{\mathrm{N}}+\eta\varepsilon_{\mathrm{N}\,,t}+\eta\gamma\varepsilon_{\mathrm{N}\,,x}\right)\right]\left(1+u_{,x}\right)}{\sqrt{\left(1+u_{,x}\right)^{2}+v_{,x}^{2}}}\right)+f_{u}\left(x,t\right), \\
&\rho A\left(v_{,tt}+2\gamma v_{,xt}+\dot{\gamma}v_{,x}+\gamma^{2}v_{,xx}\right)= \\
&\frac{\partial}{\partial x}\left(\frac{\left[P_{0}+A\left(E\varepsilon_{\mathrm{N}}+\eta\varepsilon_{\mathrm{N}\,,t}+\eta\gamma\varepsilon_{\mathrm{N}\,,x}\right)\right]v_{,x}}{\sqrt{\left(1+u_{,x}\right)^{2}+v_{,x}^{2}}}\right)-\left[EIv_{,xxxx}+\eta I\left(v_{,xxxxt}+\gamma v_{,xxxxx}\right)\right]+f_{v}\left(x,t\right),
\end{aligned} \tag{11}$$

If other viscoelastic constitutive relations are used to describe the beam materials, they can be incorporated into the governing equation in the similar way. However, a controversial issue arises concerning the application of differential-type constitutive laws including the Kelvin relation in axially moving materials. Some investigators used the partial time derivative in the Kelvin model for axially moving strings (Zhang & Zu, 1998) (Zhang & Song, 2007), (Chen et al., 2007) and (Ghayesh, 2008), or beams (Chen & Yang, 2005, 2006a,b), (Ghayesh & Balar, 2008), (Ghayesh & Khadem, 2008), (Yang et al., 2009), and (Özhan & Pakdemirli, 2009). However, (Mochensturm & Guo, 2005) convincingly argued that the Kelvin model generalized to axially moving materials should contain the material time derivative to account for the added "steady state" dissipation of an axially moving viscoelastic string. Actually the material time derivative was also used in other works on axially moving viscoelastic beams (Marynowski, 2002, 2004, 2006), (Marynowski & Kapitaniak, 2002, 2007), (Yang & Chen, 2005), (Ding & Chen, 2008), (Chen & Ding, 2008, 2009), (Chen & Wang, 2009) and (Chen, et al., 2008, 2009, 2010). Here a coordinate transform will be proposed to develop the governing equations, which can introduce naturally the material time derivative in the viscoelastic constitutive relations.

In small but finite stretching problems in literatures of nonlinear oscillations, only the lowest order nonlinear terms need to be retained so that the governing equation of small-amplitude motion will be obtained. Such simplified coupled governing equations were used in analytical investigations on axially moving elastic beams (Thurman & Mote, 1969), (Riedel & Tan, 2002), and (Sze et al., 2005).

It should be remarked that there are different types of governing equations for axially moving beams (Tabarrok et al., 1974), (Wang & Mote, 1986, 1987), (Wang, 1991), (Hwang & Perkins, 1992a,b, 1994), (Vu-Quoc & Li 1995), (Behdinan, et al, 1997), (Hochlenert et al., 2007), (Pratiher & Dwivedy 2008), (Spelsbrg-Korspeter et al., 2008), and (Humer & Irschik, 2009). Actually, there are various beam theories such as Euler-Bernoulli theory, shear-deformable theories, and three-dimensional theories, and geometric nonlinearities may take different forms. Correspondingly, there are various governing equations of axially moving beams. Even if an axially stationary slender structure is prescribed by more sophisticated governing equations, the coordinate transform (7) is a still convenient approach to derive the governing equations of the slender structure undergoing an axial motion.

2.2 Transverse vibration

Although the transverse vibration is generally coupled with the longitudinal vibration, many researchers considered only the transverse vibration in order to derive a tractable equation. Inserting u=0 into equation (3) and then omitting higher order nonlinear terms yield a simplified strain-displacement relation termed as the Lagrange strain

$$\varepsilon_{\mathrm{L}} = v_{,x}^{2}\big/2 \tag{12}$$

Inserting u=0 into equation (11) and then retaining lower order nonlinear terms only yield a nonlinear partial-differential equation

$$\begin{aligned}&\rho A\left(v_{,tt}+2\gamma v_{,xt}+\dot{\gamma}v_{,x}+\gamma^{2}v_{,xx}\right)-P_{0}v_{,xx}+\left[EIv_{,xxxx}+\eta I\left(v_{,xxxxt}+\gamma v_{,xxxxx}\right)\right]\\&=\frac{\partial}{\partial x}\left[\left(AE\varepsilon_{\mathrm{L}}+A\eta\varepsilon_{\mathrm{L}}{}_{,t}+A\eta\gamma\varepsilon_{\mathrm{L}}{}_{,x}\right)v_{,x}\right]+f_{v}\left(x,t\right).\end{aligned} \tag{13}$$

The quasi-static stretch assumption means that one can use the averaged value of the disturbed tension $\int_0^1\left[AE\varepsilon_{\mathrm{L}}+A\eta\left(\varepsilon_{\mathrm{L}}{}_{,t}+\gamma\varepsilon_{\mathrm{L}}{}_{,x}\right)\right]\mathrm{d}x\big/l$ to replace the exact value $AE\varepsilon+A\eta(\varepsilon_{,t}+c\varepsilon_{,x})$. Thus equation (18) leads to nonlinear integro-partial-differential equation

$$\begin{aligned}&\rho A\left(v_{,tt}+2\gamma v_{,xt}+\dot{\gamma}v_{,x}+\gamma^{2}v_{,xx}\right)-P_{0}v_{,xx}+\left[EIv_{,xxxx}+\eta I\left(v_{,xxxxt}+\gamma v_{,xxxxx}\right)\right]\\&=\frac{v_{,xx}}{l}\int_0^l\left(AE\varepsilon_{\mathrm{L}}+A\eta\varepsilon_{\mathrm{L}}{}_{,t}+A\eta\gamma\varepsilon_{\mathrm{L}}{}_{,x}\right)\mathrm{d}x+f_{v}\left(x,t\right).\end{aligned} \tag{14}$$

Both equation (13) and equation (14) are governing equations of transverse nonlinear vibration.

Both the nonlinear partial-differential equation and the nonlinear integro-partial-differential equation have been applied to some special cases such as free vibration without external excitation (F_v=0), elastic beams without viscoelasticity (η=0), uniformly moving beams without axially acceleration ($\dot{\gamma}$=0). The applications of the nonlinear partial-differential equation include (Chen & Zu, 2004) for uniformly moving elastic beams without external excitation, (Marynowski, 2002, 2004) and (Marynowski & Kapitaniak, 2007) for axially moving viscoelastic beams without external excitation, (Yang & Chen, 2005) and (Chen & Yang, 2006) for axially accelerating viscoelastic beams, and (Chen et al., 2007) for uniformly moving elastic beams without external excitation. The applications of the nonlinear integro-partial-differential equation include (Wickert, 1991), (Pellicano & Zirilli, F., 1997), (Pellicano & Vestroni, 2000), (Chakraborty & Mallik, 2000a), (Pellicano, 2001), (Kong & Parker, 2004) and (Chen & Zhao, 2005) for uniformly moving elastic beams without external excitation, (Ghayesh, 2008) for uniformly moving viscoelastic beams without external excitation, (Pellicano & Vestroni, 2000), (Özhan & Pakdemirli, 2009) for uniformly moving elastic beams, (Chakraborty & Mallik, 1999), (Öz et al, 2001) and (Ravindra & Zhu, 1998) for axially accelerating elastic beams without external excitation, (Chakraborty & Mallik, 1998) (Chakraborty et al., 1999), (Chakraborty & Mallik, 2000b) for axially moving elastic beams, (Parker & Lin, 2001) for axially accelerating elastic beams, and (Yang et al., 2009), (Chen et al., 2009) for axially accelerating viscoelastic beams, and (Özhan & Pakdemirli, 2009) for uniformly moving viscoelastic beams. Approximately analytical investigations on free vibration of axially moving elastic (Chen & Yang, 2007), forced vibration of axially moving viscoelastic beams (Yang & Chen, 2006), and parametric vibration of axially accelerating viscoelastic beams (Chen & Yang, 2005) and (Chen & Ding, 2008) demonstrated that the nonlinear partial-differential equation and the nonlinear integro-partial-differential equation yield the qualitatively same results but there are quantitative differences.

The nonlinear integro-partial-differential equation can also be obtained through uncoupling the governing equation for coupled longitudinal and transverse vibration under the quasi-static stretch assumption in small but finite stretching problems, and a special case of free vibration of axially moving elastic beam was treated in (Wickert, 1992). Under quasi-static stretch assumption, the dynamic tension to be a function of time alone. In traditional derivation in (Wickert, 1992), the nonlinear integro-partial-differential equation seems more exact than the nonlinear partial-differential equation because it is the transverse equation of motion in which the longitudinal displacement field is taken into account. However, the derivation here indicates that the nonlinear partial-differential equation can be reduced to the nonlinear integro-partial-differential equation based on the quasi-static stretch assumption. Numerical investigations on free vibration of axially moving elastic beams (Ding & Chen, 2008) and forced vibration of axially moving viscoelastic beams (Chen & Ding, 2009) indicated that the nonlinear integro-partial-differential equation is superior to the partial-differential equation, in the sense that approximates the coupled governing equation of planar motion better (some details in Subsection 4.2). However, since there has no decisive evidence to favor any models of transverse nonlinear vibration of axially moving beams, it is still an open problem.

3. Approximate analytical methods

3.1 Direct-perturbation approaches

As exact solutions are usually unavailable, approximate analytical methods are widely applied to investigate nonlinear vibration of axially moving beams. The approximate analytical methods can be applied to the nonlinear (integro-)partial-differential equations without discretization. Such a treatment is regarded as a direct-perturbation. The practice can be dated back to (Thurman & Mote, 1969) in which a modified Lindstedt method was used to calculate the fundamental frequency.

The method of multiple scales can be employed to analyze nonlinear vibration of axially moving beams. Actually, a general framework of the multi-scale analysis has been proposed for a linear gyroscopic continuous system under small nonlinear time-dependent disturbances (Chen & Zu, 2008). Consider a gyroscopic continuous system with a weak disturbance

$$Mv_{,tt} + Gv_{,t} + Kv = \varepsilon N\left(x,t\right), \tag{15}$$

where $v(x,t)$ is the generalized displacement of the system at spatial coordinate x and time t, linear, time-independent, spatial differential operators M, G and K represent mass, gyroscopic and stiffness operators respectively, ε stands for a small dimensionless parameter, and $N(x,t)$ expresses a nonlinear function of x and t that may explicitly contain v

and its spatial and temporal partial derivatives as well as its integral over a spatial region or a temporal interval. $N(x,t)$ is periodic in time with the period $2\pi/\omega$. Define an inner product

$$\langle f,g\rangle=\int_E f(x)\bar{g}(x)\mathrm{d}x, \tag{16}$$

for complex functions f and g defined in the gyroscopic continuum E, where the overbar denotes the complex conjugate. M, K are symmetric and G is skew symmetric in the sense

$$\langle Mf,g\rangle=\langle f,Mg\rangle,\langle Kf,g\rangle=\langle f,Kg\rangle,\langle Gf,g\rangle=-\langle f,Gg\rangle \tag{17}$$

for all functions f and g satisfying appropriate boundary conditions. A uniform approximation is sought in the form

$$v(x,t)=v_0(x,T_0,T_1)+\varepsilon v_1(x,T_0,T_1)+O\left(\varepsilon^2\right), \tag{18}$$

where $T_0=t$, $T_1=\varepsilon t$, and $O(\varepsilon^2)$ denotes the term with the same order as ε^2 or higher. Substitution of equation (18) into equation (15) yields

$$Mv_{0,T_0T_0}+Gv_{0,T_0}+Kv_0=0, \tag{19}$$

$$Mv_{1,T_0T_0}+Gv_{1,T_0}+Kv_1=N_1(x,T_0,T_1), \tag{20}$$

where $N_1(x,T_0,T_1)$ stands for a nonlinear function of x, T_0 and T_1, which usually depends explicitly on v_0 and its derivatives and integrals. In addition, $N_1(x,T_0,T_1)$ is periodic in T_0 with the period $2\pi/\omega$. Separation of variables leads to the solution of equation (19) as

$$v_0(x,T_0,T_1)=\sum_{j=1}^{\infty}A_j(T_1)\phi_j(x)\mathrm{e}^{\mathrm{i}\omega_jT_0}+cc, \tag{21}$$

where A_j denotes a complex function to be determined later, φ_j and ω_j represents respectively the complex modal function and the natural frequency given by

$$-\omega_j{}^2M\phi_j+\mathrm{i}\,\omega_jG\varphi_j+K\varphi_j=0 \tag{22}$$

and the boundary conditions, and cc stands for the complex conjugate of all preceding terms on the right side of the equation. If ω approaches a linear combination of natural frequencies of equation (19), the summation parametric resonance may occur. A detuning parameter σ is introduced to quantify the deviation of ω from the combination, and ω is described by

$$\omega=\sum_{j=1}^{\infty}c_j\omega_j+\varepsilon\sigma, \tag{23}$$

where c_j are real constants that are not all zero and only a finite of them are not zero. To investigate the summation parametric resonance, substitution of equations (21) and (23) into equation (20) leads to

$$Mv_{1,T_0T_0} + Gv_{1,T_0} + Kv_1 = \sum_{j=1}^{\infty} F_j(x,T_1)\mathrm{e}^{\mathrm{i}\omega_j T_0} + NST + cc, \tag{24}$$

where $F_j(x, T_1)$ (j=1,2,…) are complex functions dependent explicitly on $A_j(T_1)$ and their temporal derivatives as well as $\varphi_j(x)$ and their spatial derivatives and integrals. (Chen & Zu, 2008) proved that the solvability condition is the orthogonality of the coefficient of the resonant term in the first order equation and the corresponding modal function of the zero order equation, e.g.

$$\left\langle F_j(x,T_1),\varphi_j \right\rangle = 0. \tag{25}$$

It should be noticed that the solvability condition (25) holds providing the boundary conditions are appropriate. That is, M and K are symmetric and G is skew symmetric under the boundary conditions. In a specific problem, these requirements can be checked for a given the operators, boundary conditions and the modal functions. However, the examination depends only on the unperturbed linear part of the problem. For example, equation (25) holds for an axially moving beam under condition (9) (Chen & Zu, 2008). Usually, it is assumed that only the modes involved in the resonance (23) need to be considered in the linear solution (21), and the assumption is physically sound. Some case studies demonstrated mathematically that the mode uninvolved in the resonance has no effect on the steady-state response (Ding & Chen, 2008), (Chen & Wang, 2009), and (Chen et al., 2009). (Özhan & Pakdemirli, 2009) proposed multi-scale analysis on forced vibrations of general continuous systems with cubic nonlinearities in the primary resonance case.

The method of multiple scales has been applied in various transverse nonlinear vibration problems of axially moving beams. These problems include free (Öz et al, 2001) and (Chen & Yang, 2007), forced (Özhan & Pakdemirli, 2009), and parametric(Öz et al, 2001) and (Özhan & Pakdemirli, 2009) vibration of axially moving elastic beams, as well as forced (Yang & Chen, 2006) and parametric (Chakraborty & Mallik, 1999), (Chen & Yang, 2005) and (Chen & Ding, 2008) vibrations of axially moving viscoelastic beams. In addition to these works on the base of the Euler-Bernoulli beam theory, the method of multiple scales has also be applied to study free vibration of an axially moving beam with rotary inertia and temperature variation effects (Ghayesh & Khadem, 2008), parametric vibration of axially moving viscoelastic Rayleigh beams (Ghayesh & Balar, 2008), and forced (Tang et al. 2009) and parametric (Tang et al., 2010) vibrations of axially moving elastic Timoshenko beams, while the multi-scale analysis on axially moving viscoelastic Timoshenko beams has been only limited to linear parametric vibration (Chen et al., 2010).

Addition to the method of multiple scales, the method of asymptotic analysis is also an effective approach to treat parametric or nonlinear vibration. Based on the idea of Krylov, Bogoliubov, and Mitropolsky, (Wickert, 1992) developed a asymptotic method for general gyroscopic continuous systems with weak nonlinearities, and the method was specialized to free nonlinear vibration of an axially moving elastic beam with supercritical transport speed. (Maccari, 1999) proposed another asymptotic approach for analyzing transverse vibration of axially stationary beams, which are disturbed conservative continuous systems, and determined external force-response and frequency-response curves in the cases of primary resonance and subharmonic resonance for a weakly periodically forced beam with quadratic and cubic nonlinearities. The approach was extended to the gyroscopic continuous system with a weak nonlinear and time-dependent disturbance in order to analyze transverse vibration of an axially accelerating viscoelastic string constituted by the Kelvin model (Chen et al., 2008) and the standard linear solid model (Chen & Chen, 2009). The method of asymptotic analysis has been also presented for nonlinear parametric vibration of axially accelerating viscoelastic beams constituted by the Kelvin model (Chen et al., 2009) as well as linear parametric vibration of axially accelerating viscoelastic beams constituted by the Kelvin model (Chen & Wang, 2009) and the standard linear solid model (Wang & Chen, 2010).

Nonlinear normal modes whose shapes depend on the amplitude provide a possible direct treatment on nonlinear vibration of axially moving beams. (Chakraborty et al., 1999) used a temporal harmonic balance and a spatial perturbation technique to determine the nonlinear complex normal modes for free and forced vibrations of axially moving elastic beams. The approach was adopted to study the response of a parametrically excited axially moving beam both without and with an external harmonic force (Chakraborty & Mallik, 1998). The results were justified by the wave propagation analysis (Chakraborty & Mallik, 2000a,b).

3.2 Discretization-perturbation approaches

Discretization of governing equations is a commonly used approach to obtain approximate solutions of vibration problems of continuous systems. For the governing equations (11) of coupled vibration of axially moving beams, one assumes an approximate solution in the form

$$u(x,t)=\sum_{i=1}^{m}p_i(t)\phi_i(x), \tag{26}$$

$$v(x,t)=\sum_{i=1}^{n}q_i(t)\varphi_i(x), \tag{27}$$

where $p_i(t)$ and $q_i(t)$ are generalized coordinates, and $\phi_i(x)$ are $\varphi_i(x)$ base functions that are usually chosen to be the linear vibration mode shapes of axially stationary beams or moving beams (Wickert & Mote, 1990), (Chen & Yang, 2006), and (Tang et al. 2008). A weighted-residual procedure such as the Galerkin procedure can be applied to truncate equation (11) into m+n nonlinearly coupled second-order ordinary-differential equations. A general description of the Galerkin procedure is as follows. Denote the differences between the left and right sides of two equations in equation (11) as $F_u(x,u,v,t)$ and $F_v(x,u,v,t)$, which are nonlinear functions of x and t that may explicitly contain v and its spatial and temporal partial derivatives as well as its integral over a spatial region or a temporal interval. Then approximate solution (31) and (32) satisfies

$$\begin{aligned}&\left\langle F_u\left(x,\sum_{i=1}^{m}p_i(t)\phi_i(x),v\sum_{i=1}^{n}q_i(t)\varphi_i(x),t\right),\zeta_j(x)\right\rangle\\&=0,\left\langle F_v\left(x,\sum_{i=1}^{m}p_i(t)\phi_i(x),v\sum_{i=1}^{n}q_i(t)\varphi_i(x),t\right),\psi_k(x)\right\rangle=0,(j=1,\cdots,m;k=1,\cdots,n)\end{aligned} \tag{28}$$

where $\zeta_j(x)$ and $\psi_k(x)$ are the weight functions.

After the discretization, various perturbation techniques such as the method of multiple scales can be employed to analyze the resulting nonlinear ordinary-differential equations approximately. Such a treatment is regarded as a discretization-perturbation

In practical problems, *m* and *n* in the discretization expressions (26) and (27) are rather small, and they are usually 1 or 2. (Riedel & Tan, 2002) applied the method of multiple scales to the discretized equations (*m*=*n*=2) to determine the forced response of an axially moving elastic beam with internal resonance. The method of multiple scales was also applied to the discretized problem of coupled vibration of axially moving beams (Feng & Hu, 2002, 2003). (Sze et al., 2005) presented a general description of discretization of the governing equation of an axially moving elastic beam, and used incremental harmonic balance method to a concrete case of *m*=*n*=2 for forced response with internal resonance. In both studies, the mode shapes of axially stationary beams were chosen as the base functions and the weight functions.

Discretization-perturbation approaches have also been used in analyzing transverse nonlinear vibration of axially moving beams. In this case, equation (27) will be substituted into equation (13) or (14) and then the Galerkin procedure can be used to discretize equation (13) or (14) into *n* nonlinearly coupled second-order ordinary-differential equations that can be solved approximately via various perturbation techniques. The Lindstedt-Poincaré method was applied to discretized governing equations to evaluate transverse response of axially moving beams (Pellicano & Zirilli, 1997) and to analyze parametric instability of axially moving elastic beams subjected to multifrequency excitations (Parker & Lin, 2001). The method of normal forms was used to evaluate free vibration of axially moving elastic beams with internal resonances (Pellicano & Zirilli, 1997) as well as forced and parametric vibration of axially moving elastic beams (Pellicano et al., 2001). In (Pellicano & Zirilli, 1997), (Parker & Lin, 2001), and (Pellicano et al., 2001), the mode shapes of axially moving beams were chosen as the base functions and the weight functions, and their orthogonality were employed. The stationary mode shapes can also serve as the base functions and the weight functions to discretize governing equations. Based on the discretization, the Lindstedt-Poincaré method was applied to determine the forced response of axially moving elastic beams (Chen et al., 2007), and the method of multiple scales was applied to evaluate the response of an axially moving viscoelastic beams subjected to multifrequency external excitations (Yang et al., 2009). In their studies, *n*=2 (Chen et al., 2007) and *n*=1 (Yang et al., 2009), respectively.

4. Numerical approaches

4.1 Galerkin procedure

Numerical calculation is an effective approach to studying nonlinear vibration of axially moving beams. Based on the numerical solutions of the governing equations, some changing tendencies of vibration characteristics, such as frequencies or amplitudes, with related parameters can be predicted, the approximate analytical results can be verified, and the nonlinear dynamical behaviours can be revealed.

Among numerical approaches, the Galerkin procedure can be applied to discretize the governing equation of nonlinear vibration of axially moving beams. The Galerkin discretization is not only the priority of discretization-perturbations reviewed in Subsection 3.2, but also feasible approach to numerical solutions. Using the 3 order Galerkin discretization of governing equation (in the type of equation (18)) for transverse motion of axially moving viscoelastic beams excited by the changing tension, (Marynowski, 2002) and (Marynowski & Kapitaniak, 2002) numerically investigated the effects of different viscoelastic models, such as the Kelvin model, the Maxwell model, and the standard linear solid model, on the dynamic response and found that different viscoelastic models yield very close numerical results for small damping. The Galerkin procedure has been mainly use to calculate long time nonlinear dynamical behaviors, which will be addressed in Subsection 5.1.

In the application of the Galerkin discretization, the main problem in the actual computations is the complexity of the resulting discretized equations. If the number of terms retained in the Galerkin discretization is rather large, the explicit expression of nonlinear terms is very difficult to obtain. (Chen & Yang, 2006b) proposed a technique to simplify the nonlinear terms in the equations derived from the Galerkin discretization. All nonlinear terms are regrouped to combine the repeated terms and cancel the zero terms. Therefore, the resulting equations can be easily coded for computers and then be effectively calculated. For example, the Galerkin discretization of the governing equation (18) for transverse motion (in the dimensionless form) is

$$\begin{aligned}&\sum_{i=1}^{n}\ddot{q}_i(t)\langle\varphi_i,\psi_k\rangle+2\gamma\sum_{i=1}^{n}\dot{q}_i(t)\langle\varphi_i',\psi_k\rangle+(\gamma^2-1)\sum_{i=1}^{n}q_i(t)\langle\varphi_i'',\psi_k\rangle\\&+k_{\mathrm{f}}{}^2\sum_{i=1}^{n}q_i(t)\langle\varphi_i'''',\psi_k\rangle+\alpha\sum_{i=1}^{n}\dot{q}_i(t)\langle\varphi_i'''',\psi_k\rangle-\frac{3}{2}k_1^2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{n}q_i(t)q_j(t)q_s(t)\left\langle\varphi_i'\varphi_j'\varphi_s'',\psi_k\right\rangle\\&-\alpha k_2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{n}q_i(t)q_j(t)\dot{q}_s(t)\left(\left\langle2\varphi_i'\varphi_s'\varphi_j'',\psi_k\right\rangle+\left\langle\varphi_i'\varphi_j'\varphi_k'',\psi_k\right\rangle\right)=0\quad(k=1,2,\ldots,n)\end{aligned}\tag{29}$$

If both the base and weight functions are chosen as sine functions, the stationary mode shapes for the simply supported beams, equation (29) can be cast into a from convenient to compute. Evaluating the corresponding inner products, regrouping the nonlinear terms to combine the same terms and canceling all null terms in the resulting equation, one obtains

$$\begin{aligned}&\ddot{q}_k(t)+4\sum_{j+k\text{ is odd}}^{n}\frac{kj}{k^2-j^2}\left[2\gamma\dot{q}_j(t)+\dot{\gamma}q_j(t)\right]+\left(\gamma^2-1\right)k^2\pi^2q_k(t)+k_{\mathrm{f}}^2k^4\pi^4q_k+\alpha k^4\pi^4\dot{q}_k\\&=\frac{7k\pi^4}{16}\left\{\sum_{s=k+1}^{n+k}\left\{(s-k)q_{l-n}\sum_{i=\max\{1,s-n\}}^{\min\{s-1,n\}}\left[i(s-i)q_j\left(k_1^2q_{s-j}+\alpha k_2\dot{q}_{s-j}\right)\right]\right\}+\sum_{l=2}^{n-k}\left\{(n+l)q_{k+s}\sum_{j=1}^{s-1}\left[j(s-j)q_j\left(k_1^2q_{s-j}\right.\right.\right.\right.\\&\left.\left.\left.\left.+\alpha k_2\dot{q}_{s-j}\right)\right]\right\}\right\}-\frac{5k\pi^4}{16}\sum_{s=2}^{k-1}\left\{(k-s)q_{k-s}\sum_{i=1}^{s-1}\left[j(s-j)q_j\left(k_1^2q_{s-j}+\alpha k_2\dot{q}_{s-j}\right)\right]\right\}-\frac{k\alpha k_2\pi^4}{8}\left\{\sum_{s=k+1}^{n+k}\left[(s\right.\right.\\&\left.\left.-k)\dot{q}_{s-k}\sum_{i=\max\{1,s-n\}}^{\min\{s-1,n\}}j(s-j)q_jq_{s-j}\right]+\sum_{s=2}^{n-k}\left[(k+s)\dot{q}_{k+s}\sum_{i=1}^{s-1}j(s-j)q_jq_{s-j}\right]+\sum_{l=2}^{k-1}\left[(k-s)\dot{q}_{n-s}\sum_{i=1}^{l-1}j(s-j)q_jq_{s-j}\right]\right\}\end{aligned}\tag{30}$$

where the sum is defined to be zero if its lower limit is larger than its upper limit. Although equation (30) seems rather complicated, it is very efficient when used for computer implementing, because almost all repeated nonlinear terms are put together, and terms with zero coefficients are eliminated. In fact, equation (29) contains $2n^3$ nonlinear terms, while equation (30) contains less than $2n^2$ nonlinear terms. For large n, the difference is significant.

It should be remarked that, based on stationary mode shapes, the even order Galerkin discretization can take the linear gyroscopic terms into full account, while the odd order discretization will miss some effects of the gyroscopic terms.

4.2 Finite difference

The finite difference method is a numerical procedure to solve partial differential equations. The method can be used to discretize both spatial coordinates and time or to discretize spatial coordinates only. In the former case, the procedure consists of four steps: 1 Discretize the continuous spatial domain and temporal interval, on which a partial differential equation is defined, into a discrete finite difference grid; 2 Approximate the individual exact partial derivatives in the partial differential equation by algebraic finite difference approximations; 3 Substitute the finite difference approximations into the partial differential equation in order to derive a set of algebraic finite difference equations; 4 Solve the resulting algebraic finite difference equations.

The finite difference method can be applied to calculated nonlinear vibration of axially moving beams. For example, the method will be employed to solve numerically equation (11) (Chen & Ding, 2010). Introduce the $L\times T$ equispaced mesh grid with time step τ and space step h: $x_j=jh$ (j=0, 1,2,…,L, $h=l/L$); $t_n=n\tau$ (n=0,1,2,…,N, $\tau=T/N$), where T is the calculation termination time. Denote the function values $u(x,t)$ and $v\ (x,t)$ at (x_j,t_n) as $u^n{}_j$ and $v^n{}_j$. Application of centered difference approximations to the spatial, temporal and mixed partial derivatives leads to

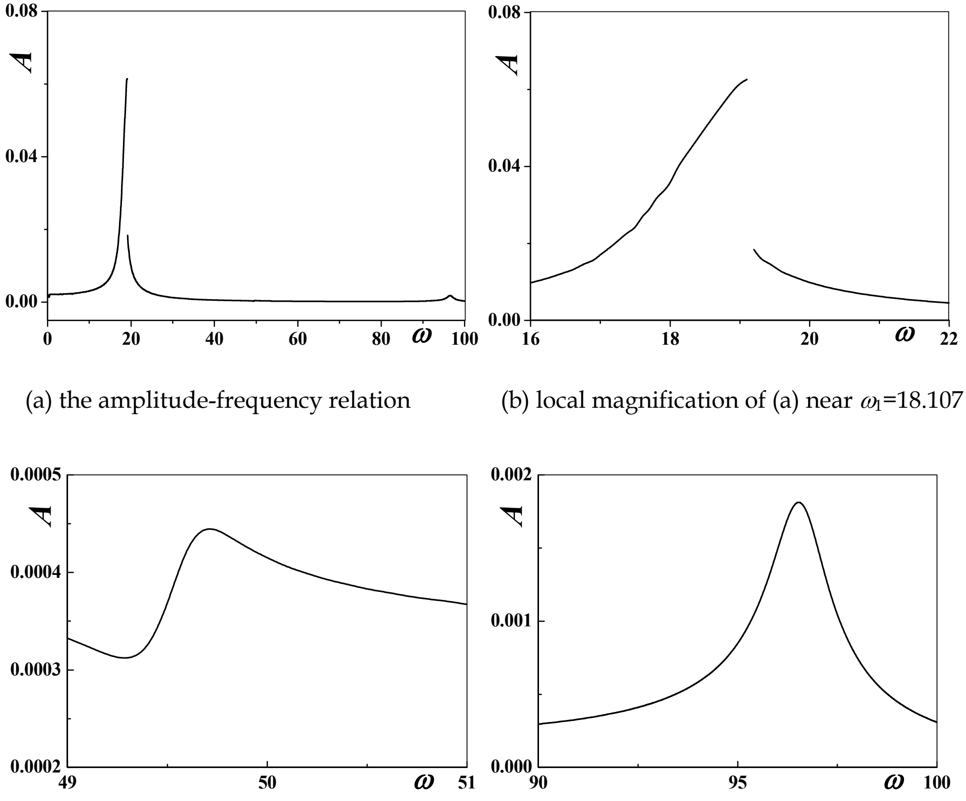
$$
\begin{aligned}
&u,_x=\frac{u_{j+1}^{n}-u_{j-1}^{n}}{2h},\ u,_{tt}=\frac{u_{j}^{n+1}-2u_{j}^{n}+u_{j}^{n-1}}{\tau^2},\ u,_{tt}=\frac{u_{j+1}^{n}-2u_{j}^{n}+u_{j-1}^{n}}{h^2},\\
&u,_{xt}=\frac{u_{j+1}^{n+1}-u_{j+1}^{n-1}-u_{j-1}^{n+1}+u_{j-1}^{n-1}}{4h\tau},
\end{aligned}
\tag{31}
$$

and

$$
\begin{aligned}
&v,_x=\frac{v_{j+1}^{n}-v_{j-1}^{n}}{2h},\ v,_{xx}=\frac{v_{j+1}^{n}-2v_{j}^{n}+v_{j-1}^{n}}{h^2},\ v,_{xxx}=\frac{-v_{j-2}^{n}+2v_{j-1}^{n}-2v_{j+1}^{n}+v_{j+2}^{n}}{2h^3},\\
&v,_{xxxx}=\frac{v_{j-2}^{n}-4v_{j-1}^{n}+6v_{j}^{n}-4v_{j+1}^{n}+v_{j+2}^{n}}{h^4},v,_{xxxxx}=\frac{-v_{j-3}^{n}+4v_{j-2}^{n}-5v_{j-1}^{n}+5v_{j+1}^{n}-4v_{j+2}^{n}+v_{j+3}^{n}}{2h^5},\\
&v,_{tt}=\frac{v_{j}^{n+1}-2v_{j}^{n}+v_{j}^{n-1}}{\tau^2},\ v,_{xt}=\frac{v_{j+1}^{n+1}-v_{j+1}^{n-1}-v_{j-1}^{n+1}+v_{j-1}^{n-1}}{4h\tau},\\
&v,_{xxxxt}=\frac{v_{j+2}^{n+1}-4v_{j+1}^{n+1}+6v_{j}^{n+1}-4v_{j-1}^{n+1}+v_{j-2}^{n+1}-v_{j+2}^{n-1}+4v_{j+1}^{n-1}-6v_{j}^{n-1}+4v_{j-1}^{n-1}-v_{j-2}^{n-1}}{2\tau h^4},
\end{aligned}
\tag{32}
$$

Substitution of equations (31) and (32) into equation (11) leads to a set of algebraic equations with respect to $u^n{}_j$ and $v^n{}_j$ that can be solved as under the boundary conditions (8) and (9) for prescribed parameters and initial conditions. Then the resulting grid values $u^n{}_j$ and $v^n{}_j$ are used in the finite difference schemes as an approximation to the continuous solutions $u(x,t)$ and $v(x,t)$ to equation (11). When the external transverse load is a spatially uniformly distributed periodic force, the amplitude of the beam center displacement changes with the force frequency, which is shown in Fig. 2.

The finite difference method was applied to examine the validity of the two nonlinear transverse models (equations (13) and (14)) and to determine the superiority in the sense of approximating the coupled governing equation (11) of planar vibration. For forced vibration of axially moving viscoelastic beams (Chen & Ding, 2010), the steady-state transverse responses of the beam center calculated from the two transverse models are contrasted with the results based on the coupled equations of planar vibration. Qualitatively, the three models predict the same tendencies with the changing parameters. Quantitatively, there are certain differences. In the view of both the center amplitude and the beam shape, the nonlinear integro-partial-differential equation yields the results closer to those from the governing equation of coupled vibration. The similar result was obtained by the finite difference method for response in free vibrations of axially moving elastic beams (Ding & Chen, 2009a).



(a) the amplitude-frequency relation

(b) local magnification of (a) near ω_1=18.107

(c) local magnification of (a) near ω_2=49.706

(d) local magnification of (a) near ω_3=97.140

Fig. 2. The response amplitude changing with the external excitation frequency

The finite difference method was used to confirm the analytical results of nonlinear transverse vibration of axially moving beams. For free vibration of axially moving elastic beams, (Pellicano & Zirilli, 1997) compared the beam center displacement changing with time via the Lindstedt- Poincaré method and the normal form method with the numerical solutions via the finite difference method, and found that they are in good agreement. For parametric vibration of axially accelerating viscoelastic beams, (Chen & Ding, 2008) compared the stable steady-state response of the beam center via the method of multiple scales with the numerical solutions via the finite difference method, and demonstrated that they have the same qualitative tendencies changing with the related parameters and are quantitatively with rather high precision.

4.3 Differential quadrature

The differential quadrature method, initiated from the idea of integral quadrature, is an efficient discretization technique to seek accurate numerical solutions using a considerably small number of grid points. The method can be used to discretize both spatial coordinates and time or to discretize spatial coordinates only. In later case, the differential quadrature discretization of a partial-differential equation yields a set of differential-algebraic equations via the following four steps. 1 Discretize the continuous spatial domain, on which a partial differential equation is defined, by grid points; 2 Approximate the individual exact partial derivatives in the partial differential equation by a linear weighted sum of all the functional values at all grid points; 3 Substitute the differential quadrature approximations into the partial differential equation to obtain a set of ordinary-differential-algebraic equations; 4 Solve the resulting ordinary-differential-algebraic equations. The two extensively decisive issues in the applications of the differential quadrature method are to choose grid points and to determine the weighting coefficients for the discretization of a derivative of necessary order.

The differential quadrature method can be applied to calculated nonlinear vibration of axially moving beams. Equation (11) is treated as an example to show the application of the differential quadrature method. Introduce *N* unequally spaced grid points as (Bert & Malik, 1996) and (Shu, 2001)

$$x_i = \frac{1}{2}\left[1 - \cos\frac{(i-1)\pi}{N-1}\right] \quad (i = 1, 2, \ldots, N). \tag{33}$$

The quadrature rules for the derivatives of a function at the grid points yield

$$\begin{gathered}
u_{,x}(x_i,t) = \sum_{j=1}^{N} A_{ij}^{(1)} u(x_j,t),\; u_{,xx}(x_i,t) = \sum_{j=1}^{N} A_{ij}^{(2)} u(x_j,t), \\
v_{,x}(x_i,t) = \sum_{j=1}^{N} A_{ij}^{(1)} v(x_j,t),\; v_{,xx}(x_i,t) = \sum_{j=1}^{N} A_{ij}^{(2)} v(x_j,t), \\
v_{,xxxx}(x_i,t) = \sum_{j=1}^{N} A_{ij}^{(4)} v(x_j,t),\; v_{,xxxxx}(x_i,t) = \sum_{j=1}^{N} A_{ij}^{(5)} v(x_j,t),
\end{gathered} \tag{34}$$

where the weighting coefficients are the expression

$$A_{ij}^{(1)} = \frac{\prod_{k=1,k\neq i}^{N}\left(x_i - x_k\right)}{\left(x_i - x_j\right)\prod_{k=1,k\neq j}^{N}\left(x_j - x_k\right)} \quad \left(i,j = 1,2,\cdots,N;\ j\neq i\right) \tag{35}$$

and the recurrence relationship

$$\begin{aligned} A_{ij}^{(r)} &= r\left[A_{ii}^{(r-1)}A_{ij}^{(1)} - \frac{A_{ij}^{(r-1)}}{x_i - x_j}\right] \left(r = 2,3,4,5; i,j = 1,2,\cdots,N;\ j\neq i\right), \\ A_{ii}^{(r)} &= -\sum_{k=1,k\neq i}^{N} A_{ik}^{(r)} \left(r = 1,2,3,4,5; i = 1,2,\cdots,N\right). \end{aligned} \tag{36}$$

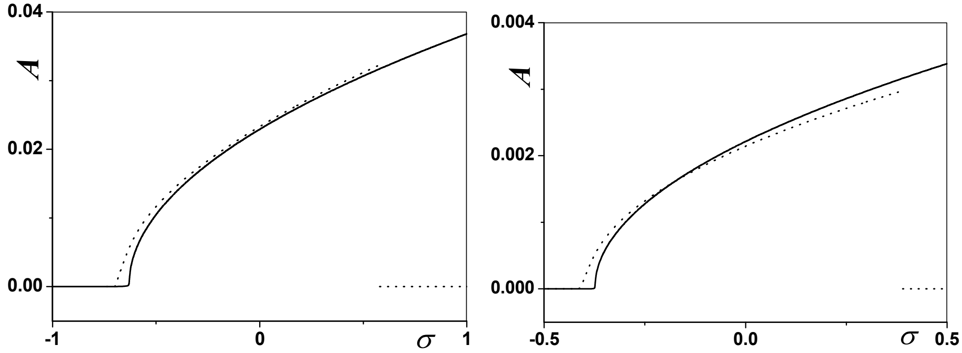
Consider the beam with simply supports at both ends (*K*=0 in equation (9)). Substitution of equation (34) into equation (11) and modification of the weighting coefficient matrices to implement the boundary conditions (Wang & Bert 1993) lead to the ordinary-differential-algebraic equations

$$\begin{aligned} &\ddot{u}_j + 2\gamma\sum_{j=1}^{N} A_{ij}^{(1)}\dot{u}_j + \sum_{j=1}^{N}\left(\dot{\gamma}A_{ij}^{(1)} + \gamma^2 A_{ij}^{(2)}\right)u_j + \dot{\gamma} = \\ &\sum_{j=1}^{N} A_{ij}^{(1)}\left\{\frac{\left(1 + \sum_{j=1}^{N} A_{ij}^{(1)}u_j\right)}{\rho A\left(\varepsilon_j + 1\right)}\left[P_0 + A\left(E\varepsilon_j + \eta\dot{\varepsilon}_j + \eta\gamma\sum_{j=1}^{N} A_{ij}^{(1)}\varepsilon_j\right)\right]\right\} + \frac{f_u\left(x_j,t\right)}{\rho A}, \\ &\ddot{v}_j + \sum_{j=1}^{N}\left(2\gamma A_{ij}^{(1)} + \frac{\eta I}{\rho A}A_{ij}^{(5)}\right)\dot{v}_j + \sum_{j=1}^{N}\left(\dot{\gamma}A_{ij}^{(1)} + \gamma^2 A_{ij}^{(2)} + \frac{EI}{\rho A}A_{ij}^{(4)} + \frac{\gamma}{\rho A}A_{ij}^{(5)}\right)v_j \\ &= \frac{1}{\rho A}\sum_{j=1}^{N} A_{ij}^{(1)}\left\{\frac{\sum_{j=1}^{N} A_{ij}^{(1)}v_j}{\varepsilon_j + 1}\left[P_0 + A\left(E\varepsilon_j + \eta\dot{\varepsilon}_j + \eta\gamma\sum_{j=1}^{N} A_{ij}^{(1)}\varepsilon_j\right)\right]\right\} + \frac{f_v\left(x_j,t\right)}{\rho A}, \quad (j = 2,3,\ldots,N-1)\ , \\ &v_1 = v_N = u_1 = u_N = 0, \end{aligned} \tag{37}$$

which can be numerically solved via the convenient integration routines.

The differential quadrature method was applied to check the validity and the superiority of the two nonlinear transverse models. For free vibration of axially moving elastic beams (Ding & Chen, 2009a), the transverse responses of the beam center calculated from the two transverse models are contrasted with the results based on the coupled equations of planar vibration. The computational investigation leads to the following conclusions: 1 The differences between the two models are both relatively small for not very large vibration; 2 The model differences increase with the vibration amplitude and the axial speed; 3 The integro-partial-differential equation yields better results.

The differential quadrature method was used to validate the analytical results of nonlinear transverse vibration of axially moving beams. (Chen et al., 2009) developed a differential quadrature scheme to verify the approximate analytical results of stable steady-state response in parametric vibration of axially accelerating viscoelastic beams. Figure 3 shows the comparison, in which the solid and dot lines represent the results of the asymptotic analysis method and the differential quadrature method respectively. The amplitudes from both methods are almost coincided, especially near the exact-resonance (σ=0) and in the first resonance. The differential quadrature method was also applied to confirm the analytical results of the stability regions in linear parametric vibration of axially accelerating beams constituted by the Kelvin model (Chen & Wang, 2009) and the standard linear solid model (Wang & Chen, 2009)



(a) the first principal parametric resonance (b) the second principal parametric resonance

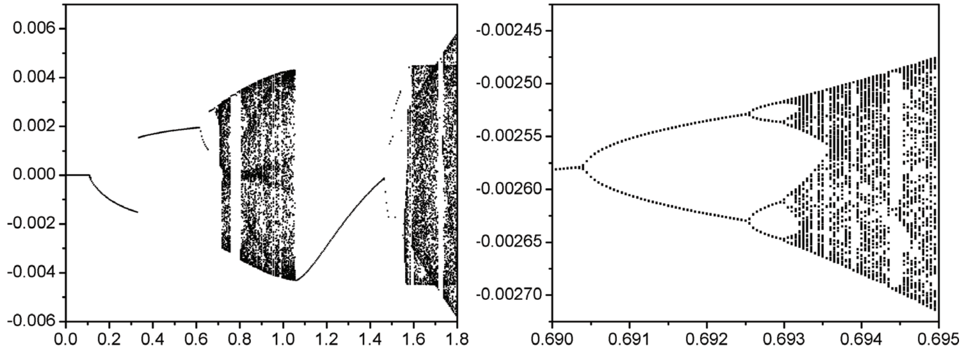
Fig. 3. Comparison of analytical and numerical results

5. Nonlinear dynamical behaviours

5.1 Galerkin discretization

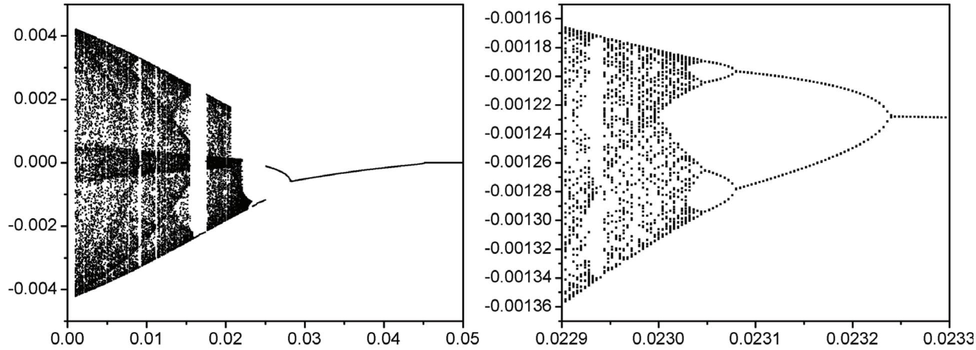
Axially moving beams undergo periodic vibrations in the aforementioned researches. Nonlinear system may exhibit chaos, steady-state response sensitive to initial conditions thus unpredictable after a certain time and recurrent but either periodic or quasiperiodic hence like a random single with the continuous frequency spectrum. Besides, the dynamical behaviors of nonlinear system may change qualitatively at the critical value of the parametric variation, and the qualitative change is termed as bifurcation.

Many investigations on bifurcation and chaos are based on the Galerkin discretization of various transverse models of axially moving beams. For transverse free vibration of accelerating elastic beams in the supercritical regime, based on 1 order Galerkin discretization, (Ravindra & Zhu, 1998) applied Melnikov's criterion to find out the parameter condition of occurring chaos and performed numerical simulations to show both period-doubling and intermittent routes to chaos. For transverse harmonically forced vibration of axially moving elastic beams in the supercritical regime, based on 8 order Galerkin discretization, (Pellicano & Vestroni, 2002) observed intricate scenario of chaos, including cascade of bifurcations, blue-sky catastrophes and coexisting chaotic and periodic orbits. Actually, they also considered 12 order Galerkin discretization and found that a few number of degree-of-freedom is sufficient to furnish a good spatial representation and to follow the actual dynamical behaviors. For transverse parametric vibration of axially moving viscoelastic beams excited by the time-dependent tension, based on 4 order Galerkin discretization, (Marynowski, 2004) and (Marynowski & Kapitaniak, 2007) observed the inverse period doubling and inverse Hope bifurcation and occurrences of regular and chaotic motions for beams constituted by the Kelvin model and the standard linear solid model respectively. For transverse parametric vibration of axially accelerating viscoelastic beams, based on 4 order Galerkin discretization, (Chen & Yang, 2006b) constructed numerically the bifurcation diagrams in the case that the axial speed perturbation amplitude, the mean axial speed, or the viscosity coefficient is respectively varied while other parameters are fixed. They also calculated the largest Lyapunov exponent from the discretized governing equation. Numerical results show that, with the increasing speed perturbation amplitude, the increasing mean speed, and the decreasing viscosity coefficient, the equilibrium loses its stability and bifurcates into a periodic motion, and the periodic motion becomes chaotic motion via period doubling bifurcation. In addition, the chaotic motion and the periodic motion exchange alternately for the sufficiently large speed



(a) From equilibrium to chaos (b) Local magnification

Fig. 4. Bifurcation versus the dimensionless speed fluctuation amplitude



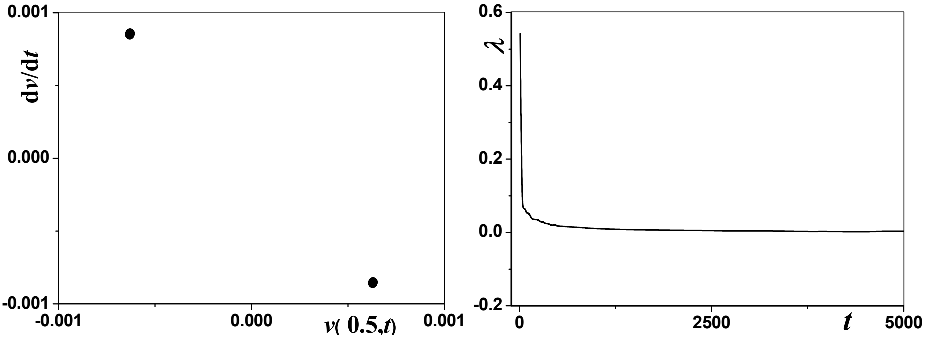
(a) From chaos to equilibrium (b) Local magnification

Fig. 5. Bifurcation versus the dimensionless viscosity coefficient

perturbation amplitude and mean speed, and for the sufficiently small viscosity coefficient. Figures 4 and 5 show respectively the bifurcation diagrams versus the dimensionless speed fluctuation amplitude and the dimensionless viscosity coefficient.

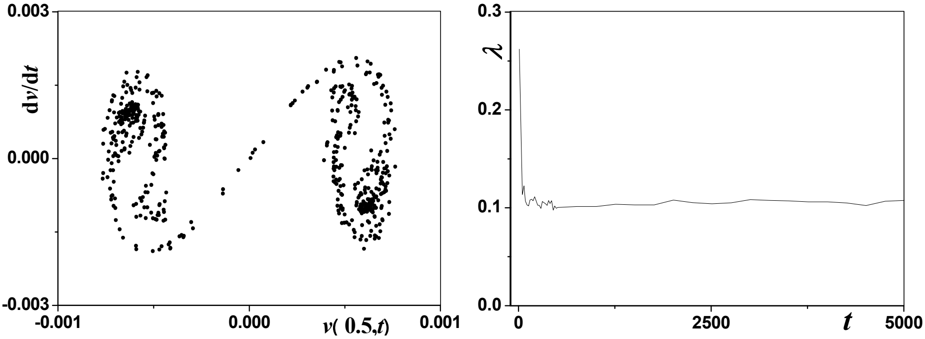
5.2 Differential quadrature and time series

The differential quadrature method is an effective numerical technique for initial and boundary problems, and it has much higher precision than the few term Galerkin discretization. However, it has not been applied to calculate nonlinear behaviors of axially moving materials until (Ding & Chen, 2009b). They used the differential quadrature method to investigate bifurcation and chaos of an axially accelerating viscoelastic beam constituted by the Kelvin model. Based on the numerical solutions, analysis of the time series yielded the Lyapunov exponent to identify periodic and chaotic motions. Numerical results show that, with the increasing mean axial speed, the equilibrium loses its stability and bifurcates into a periodic motion, and the periodic motion becomes chaotic motion. The chaotic motion and the periodic motion exchange alternately for the sufficiently large mean axial speed and speed perturbation amplitude. Figures 6 and 7 show the Poincaré map and the largest Lyapunov exponent of periodic and chaotic motions respectively.



(a) the Poincaré map (b) the largest Lyapunov exponent

Fig. 6. Periodic motion of the beam centre



(a) the Poincaré map (b) the largest Lyapunov exponent

Fig. 7. Chaotic motion of the beam centre

6. Energetics and conserved quantity

6.1 Energetics

It is well known that the total mechanical energy in free vibration of an undamped axially stationary elastic beam with pinned or fixed ends is constant. However, many investigations found that the total mechanical energy associated with free vibration of an axially moving elastic beam is not constant even if the beam travels between two motionless supports. (Barakat, 1968) considered the energetics of an axially moving beam and found that energy flux through the supports can invalidate the linear theories of axially moving beams at sufficiently high transporting speed. (Tabarrok, 1974) showed that the total energy of a traveling beam without tension is periodic in time. (Wickert & Mote, 1989) presented the temporal variation of the total energy related to the local rate of change and calculated the temporal variation of energy associated with modes of moving beams. Considering the case that there were nonconservative forces acting on two boundaries, (Lee & Mote 1997) presented a generalized treatment of energetics of translating beams. (Renshaw et al., 1998) examined the energy of axially moving beams from both Lagrangian and Eulerian views and found that Lagrangian and Eulerian energy functionals are not conserved for axially moving beams. (Zhu & Ni, 2000) investigated energetics of axially moving strings and beams with arbitrarily varying lengths. (Chen & Zu, 2004) proposed energetics of axially moving beams with geometric nonlinearity due to small but finite stretching of the beams. Hence the variation of the total mechanical energy is a fundamental feature of free transverse vibration of axially moving beams. However, all aforementioned investigations on energetics and conserved quantities of axially moving beams have only been limited to transverse vibration, in which longitudinal motion is assumed to be uncoupled and thus neglectable. Actually, the energetics of coupled vibration of axially moving elastic strings (Chen, 2006) can be extended to beams.

Assume that the axially moving beam described at the beginning of Subsection 2.1 is elastic (η=0), is without external excitations (f_u=0, f_v=0), and moves in a constant axially speed (γ=c). Consider the total mechanical energy in a specified spatial domain, the span (0, L). The total mechanical energy consists of the kinetic energy of all material particles and the potential energy resulted from the initial tension, the disturbed tension, and the bending moment caused by the beam deflection due to its motion

$$\mathscr{E}=\int_0^L\left\{\frac{\rho A}{2}\left[\left(c+u_{,t}+cu_{,x}\right)^2+\left(v_{,t}+cv_{,x}\right)^2\right]+\left(P_0+\frac{1}{2}EA\varepsilon\right)\varepsilon+\frac{1}{2}EIv_{,xx}^2\right\}\mathrm{d}x \tag{38}$$

Then the time-rate of energy change is

$$\frac{\mathrm{d}\mathscr{E}}{\mathrm{d}t}=\int_0^L\frac{\partial}{\partial t}\left\{\frac{\rho A}{2}\left[\left(c+u_{,t}+cu_{,x}\right)^2+\left(v_{,t}+cv_{,x}\right)^2\right]+\left(P_0+\frac{1}{2}EA\varepsilon\right)\varepsilon+\frac{1}{2}EIv_{,xx}^2\right\}\mathrm{d}x \tag{39}$$

Interchanging the order of differentiation and integration and inserting $u_{,tt}$ and $v_{,tt}$ solved from equation (11), after some mathematical manipulations, one can express the time-rate of energy change in the boundary values

$$\frac{\mathrm{d}\boldsymbol{\mathscr{e}}}{\mathrm{d}t}=\left[\begin{array}{l}(c+u_{,t}+cu_{,x})\dfrac{(P_0+EA\varepsilon)(1+u_{,x})}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}+(v_{,t}+cv_{,x})\dfrac{(P_0+EA\varepsilon)v_{,x}}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}\\ +EIv_{,xx}(v_{,xt}+cv_{,xx})-EIv_{,xxx}(v_{,t}+cv_{,x})\end{array}\right]\Bigg|_0^l \tag{40}$$
$$-c\left\{\frac{1}{2}\rho A\left[(c+u_{,t}+cu_{,x})^2+(v_{,t}+cv_{,x})^2\right]+\left(P_0\varepsilon+\frac{1}{2}EA\varepsilon^2\right)+\frac{1}{2}EI{v_{,xx}}^2\right\}\Bigg|_0^l .$$

Notice that

$$P_u=\frac{(P_0+EA\varepsilon)(1+u_{,x})}{\sqrt{(1+u_{,x})^2+v_{,x}^2}},\ P_v=\frac{(P_0+EA\varepsilon)v_{,x}}{\sqrt{(1+u_{,x})^2+v_{,x}^2}} \tag{41}$$

are respectively the longitudinal and transverse components of the tension in the beam, $EIv_{,xx}$ is the bending moment and $EIv_{,xxx}$ are the shear, while $c+u_{,t}+cu_{,x}$ and $v_{,t}+cv_{,x}$ are respectively the absolute velocity in the longitudinal and transverse directions and $v_{,tx}+cv_{,xx}$ is the absolute angle velocity. Hence the first term in equation (40) stands for the difference of power of the beam tension, the beam bending moment, and the beam shear acting at two ends. Meanwhile,

$$\hat{\boldsymbol{\mathscr{e}}}=\frac{1}{2}\rho A\left[(c+u_{,t}+cu_{,x})^2+(v_{,t}+cv_{,x})^2\right]+\left(P_0\varepsilon+\frac{1}{2}EA\varepsilon^2\right)+\frac{1}{2}EI{v_{,xx}}^2 \tag{42}$$

is the total mechanical energy per unit length. Hence the second term in equation (40) stands for the energy change due to the axial motion of the beam. Physically, equation (40) means the change rate of the energy consisting of two parts: the power of the beam tension, moment and shear applying at two ends and the energy variation resulted from the axial motion.

For a beam with the simple support (K=0 in equation (9)) or the fixed ends ($K\to\infty$ in equation (9)), equation (40) leads to, respectively,

$$\frac{\mathrm{d}\boldsymbol{\mathscr{e}}}{\mathrm{d}t}=c\left[\left(EA-\rho Ac^2\right)\varepsilon\left(1+\frac{\varepsilon}{2}\right)+EI\left(\frac{1}{2}v_{,xx}-v_{,xxx}v_{,x}\right)\right]\Bigg|_0^L, \tag{43}$$

$$\frac{\mathrm{d}\boldsymbol{\mathscr{e}}}{\mathrm{d}t}=c\left[\left(EA-\rho Ac^2\right)u_{,x}\left(1+\frac{u_{,x}}{2}\right)+EIv_{,xx}\right]\Bigg|_0^L. \tag{44}$$

For an axially stationary beam, c=0. Equation (40) becomes

$$\frac{\mathrm{d}\boldsymbol{\mathscr{e}}}{\mathrm{d}t}=\left[u_{,t}\frac{(P_0+EA\varepsilon)(1+u_{,x})}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}+v_{,t}\frac{(P_0+EA\varepsilon)v_{,x}}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}+EIv_{,xx}v_{,xt}-EIv_{,xxx}v_{,t}\right]\Bigg|_0^L. \tag{45}$$

If the axially stationary beam is with pinned or fixed ends, equation (45) leads to the conservation of the mechanical energy, which is a well known fact.

6.2 Conserved quantity

Although the total mechanical energy of axially moving beams is generally not constant, there does exist an alternative conserved quantity. (Renshaw et al. 1998) presented both Eulerian and Lagrangian conserved functionals for axially moving beams. (Chen & Zu, 2004) generalized their results to nonlinear free vibration of axially moving beams. They adopted the partial-differential equation (a special case of equation (13)) for axially moving beams undergoing nonlinear transverse vibration. (Chen & Zhao, 2005) also present a conserved functional for a beam modeled by an integro-partial-differential equation derived from the quasi-static assumption (a special case of equation (14)). They applied the conserved functional to verify that the straight equilibrium configuration is stable for beams at low axial speed.

Define the functional

$$\boldsymbol{\mathcal{J}}=\int_0^L\left\{\frac{\rho A}{2}\left[\left(u_{,t}{}^2-c^2u_{,x}{}^2\right)+\left(v_{,t}{}^2-c^2v_{,x}{}^2\right)\right]+\left(P_0+\frac{1}{2}EA\varepsilon\right)\varepsilon+\frac{1}{2}EIv_{,xx}{}^2\right\}\mathrm{d}x \tag{46}$$

Evaluation of the temporal differentiation by parts yield

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{\mathcal{J}}}{\mathrm{d}t}=\int_0^L&\left\{u_{,t}\left[\rho Au_{,tt}+2\rho Acu_{,xt}+\left(\rho Ac^2-EA\right)u_{,xx}-\left(EA-P_0\right)\frac{(1+u_{,x})v_{,x}v_{,xx}-v_{,x}{}^2u_{,xx}}{\left[(1+u_{,x})^2+v_{,x}^2\right]^{3/2}}\right]\right.\\&\left.+v_{,t}\left[\rho Av_{,tt}+2\rho Acv_{,xt}+\left(\rho Ac^2-EA\right)u_{,xx}+EIv_{,xxxx}-\left(EA-P_0\right)\frac{(1+u_{,x})^2v_{,xx}-(1+u_{,x})v_{,x}u_{,xx}}{\left[(1+u_{,x})^2+v_{,x}^2\right]^{3/2}}\right]\right\}\mathrm{d}x\\&-\rho Ac\left[(u_{,t}+cu_{,x})u_{,t}+(v_{,t}+cv_{,x})v_{,t}\right]\Big|_0^L+\left[\frac{(P_0+EA\varepsilon)(1+u_{,x})}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}u_{,t}+\frac{(P_0+EA\varepsilon)v_{,x}}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}v_{,t}+EIv_{,xx}v_{,xt}\right]\Bigg|_0^L.\end{aligned} \tag{47}$$

Substitution of equation (11) with η=0, f_u=0, f_v=0, and γ=c into equation (47) leads to

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{\mathcal{J}}}{\mathrm{d}t}=&-\rho Ac\left[(u_{,t}+cu_{,x})u_{,t}+(v_{,t}+cv_{,x})v_{,t}\right]\Big|_0^L\\&+\left[\frac{(P_0+EA\varepsilon)(1+u_{,x})}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}u_{,t}+\frac{(P_0+EA\varepsilon)v_{,x}}{\sqrt{(1+u_{,x})^2+v_{,x}^2}}v_{,t}+EIv_{,xx}v_{,xt}\right]\Bigg|_0^L.\end{aligned} \tag{48}$$

At a pinned or fixed end, $u_{,t}$=0, $v_{,t}$=0, $v_{,xx}$=0 or $v_{,x}$=0 (hence $v_{,xt}$=0). Therefore, equation (48) results in $\mathrm{d}\boldsymbol{\mathcal{J}}/\mathrm{d}t$=0. There exists functional (46) that is conserved under pinned or fixed boundary conditions for beams moving with a constant axial speed c.

The conserved quantity in a mechanical system is not only mathematically the first integral leading to a reduction in the order of the system, but also reflects the physical essence of the system closely related to the symmetries of the system. Therefore, it is theoretically significant to investigate the conserved quantities. The conserved quantity in a mechanical system can be used to check and develop numerical simulation algorithms. It is also useful for stability analysis and controller design.

7. Concluding remarks

Because an axially moving beam is an effective mechanical model that can be used in diverse engineering fields, many research activities in the area have been witnessed. The chapter summarizes some resent works on modeling, analysis and simulations of nonlinear vibrations of axially moving beams. It will remain to be an active research field. There are many promising topics for future researches, including but surely not limited to the follows: (1) modeling slender structures via sophisticated beam theories such as three-dimensional beams or composite beams, (2) incorporating functionally graded, theromviscoelastic or other advanced materials, (3) accounting for aerodynamic forces and heating and other actions coupled with the vibration, (4) considering complex constraints and coupling such as belts in drive systems, (5) developing analytical approaches especially for coupled vibrations and strongly nonlinear vibrations, (6) investigating convergence, consistency, and stability of numerical procedures, (7) exploring energetics of nonlinear and time-dependent beams under general constraint conditions, (8) understanding complicated dynamical behaviors such as global bifurcations, chaos, patterns, and spatio-temporal chaos.

8. Acknowledgments

This work was supported by the National Outstanding Young Scientists Foundation of China (No. 10725209), the National Natural Science Foundation of China (No. 90816001), the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20093108110005), Shanghai Subject Chief Scientist Project (No. 09XD1401700), Shanghai Leading Talent Program, Shanghai Leading Academic Discipline Project (No. S30106), and the Program for Changjiang Scholars and Innovative Research Team in University (No. IRT0844).

9. References

Abrate, A. S. (1992). Vibration of belts and belt drives. *Mechanism and Machine Theory*, 27, 6, 645-659, ISSN 0094-114X

Barakat, R. (1968). Transverse vibrations of a moving thin rod. *The Journal of the Acoustical Society of America*, 43, 533-539, ISSN 0001-4966

Behdinan, K.; Stylianou M.C. & Tabarrok, B. (1997). Dynamics of flexible sliding beams—non-linear analysis part I: formulation. *Journal of Sound and Vibration*, 208, 4, 517-539, ISSN 0022-460X

Bert, C. W. & Malik, M. (1996). The differential quadrature method in computational mechanics: a review. *Applied Mechanics Reviews*, 49, 1-28, ISSN 0003-6900

Chakraborty, G. & Mallik, A.K. (1998). Parametrically excited nonlinear traveling beams with and without external forcing. *Nonlinear Dynamics*, 17, 4, 301-324, ISSN 1573-269X

Chakraborty, G. & Mallik, A.K. (1999). Stability of an accelerating beam. *Journal of Sound and Vibration*, 227, 2, 309-320, ISSN 0022-460X

Chakraborty, G. & Mallik, A.K. (2000a). Wave propagation in and vibration of a travelling beam with and without non-linear effects, part I: free vibration. *Journal of Sound and Vibration*, 236, 2, 277-290, ISSN 0022-460X

Chakraborty, G. & Mallik, A.K. (2000b). Wave propagation in and vibration of a travelling beam with and without non-linear effects, part II: forced vibration. *Journal of Sound and Vibration*, 236, 2, 291-305, ISSN 0022-460X

Chakraborty, G.; Mallik, A.K. & Hatwal, H. (1999).Non-linear vibration of a travelling beam. *International Journal of Non-Linear Mechanics*, 34, 655-670, ISSN 0020-7462

Chen, L.H.; Zhang, W. & Liu, Y.Q. (2007). Modeling of nonlinear oscillations for viscoelastic moving belt using generalized Hamilton's Principle. *ASME Journal of Vibration and Acoustics*, 129, 128-132, ISSN 1528-8927

Chen, L.Q. (2005). Analysis and control of transverse vibrations of axially moving strings. *Applied Mechanics Reviews*, 58, 91-116, ISSN 0003-6900

Chen, L.Q. (2005). Principal parametric resonance of axially accelerating viscoelastic strings constituted by the Boltzmann superposition principle, *Proceedings of the Royal Society of London A*: *Mathematical, Physical and Engineering Sciences*, 461, 2061), 2701-2720, ISSN 1471-2946

Chen, L.Q. (2006). The energetics and the stability of axially moving strings undergoing planer motion. *International Journal of Engineering Science*, 44, 1346-1352, ISSN 0020-7225

Chen, L.Q. & Chen, H. (2009). Asymptotic analysis on nonlinear vibration of axially accelerating viscoelastic strings with the standard linear solid model. *Journal of Engineering Mathematics*, accepted

Chen, L.Q.; Chen, H. & Lim, C.W. (2008). Asymptotic analysis of axially accelerating viscoelastic strings. *International Journal of Engineering Science*, 46(10): 976-985, ISSN 0020-7225

Chen, L.Q. & Ding, H. (2008). Steady-state responses of axially accelerating viscoelastic beams: approximate analysis and numerical confirmation. *Science in China Series G: Physics, Mechanics & Astronomy*, 51(11): 1701-1721, ISSN 1862-2844

Chen, L.Q. & Ding H. (2009). Steady-state transverse response in planar vibration of axially moving viscoelastic beams. *ASME Journal of Vibration and Acoustics*, in press, ISSN 1528-8927

Chen, L.Q.; Tang, Y.Q. & Lim, C.W. (2010). Dynamic stability in parametric resonance of axially accelerating viscoelastic Timoshenko beams. *Journal of Sound and Vibration*, 329, 547-565, ISSN 0022-460X

Chen, L.Q. & Wang, B. (2009). Stability of axially accelerating viscoelastic beams: asymptotic perturbation analysis and differential quadrature validation, *European Journal of Mechanics A/Solid*, 28(4): 786-791, ISSN 0997-7538

Chen, L.Q.; Wang, B. & Ding, H. (2009). Nonlinear parametric vibration of axially moving beams: asymptotic analysis and differential quadrature verification. *Jorurnal of Physics: Conference Series*, 181, 012008, ISSN 1742-6596

Chen, L.Q. & Yang, X.D. (2005). Steady-state response of axially moving viscoelastic beams with pulsating speed: comparison of two nonlinear models. *International Journal of Solids and Structures*, 42, 37-50, ISSN 0020-7683

Chen, L.Q. & Yang, X.D. (2006a). Vibration and stability of an axially moving viscoelastic beam with hybrid supports. *European Journal of Mechanics A/Solids*, 25, 996-1008, ISSN 0997-7538

Chen, L.Q. & Yang, X.D. (2006b). Transverse nonlinear dynamics of axially accelerating viscoelastic beams based on 4-term Galerkin truncation. *Chaos, Solitons and Fractals*, 2006, 27, 3, 748-757, ISSN 0960-0779

Chen, L.Q. & Yang, X.D. (2007). Nonlinear free vibration of an axially moving beam: comparison of two models. *Journal of Sound and Vibration*, 299, 348-354, ISSN 0022-460X

Chen, L.Q. & Zhao, W.J. (2005). A conserved quantity and the stability of axially moving nonlinear beams. *Journal of Sound and Vibration*, 286, 663-668, ISSN 0022-460X

Chen, L.Q.; Zhang, W. & Zu, J.W. (2009). Nonlinear dynamics in transverse motion of axially moving strings. *Chaos, Solitons & Fractals*, 40, 1, 78-90, ISSN 0960-0779

Chen, L.Q. & Zu, J.W. (2004). Energetics and Conserved Functional of Moving Materials Undergoing Transverse Nonlinear Vibration. *ASME Journal of Vibration and Acoustics*, 126, 452-455, ISSN 1528-8927

Chen, L.Q. & Zu, J.W. (2008). Solvability condition in multi-scale analysis of gyroscopic continua. *Journal of Sound and Vibration*, 309, 338-342, ISSN 0022-460X

Chen, S.H.; Huang, J.L. & Sze, K.Y. (2007). Multidimensional Lindstedt-Poincaré method for nonlinear vibration of axially moving beams. *Journal of Sound and Vibration*, 306, 1-11, ISSN, 0022-460X

D'Angelo C. III; Alvarado, N.T.; Wang, K.W., and Mote, C.D.Jr. (1985). Current research on circular saw and band saw vibration and stability. *The Shock and Vibration Digest.* 17, 5, 11-23, ISSN 1741-3184

Ding, H. & Chen, L.Q. (2008). Stability of axially accelerating viscoelastic beams: multi-scale analysis with numerical confirmations. *European Journal of Mechanics A/Solid*, 27(6): 1108-1120, ISSN 0997-7538

Ding, H. & Chen, L.Q. (2009a). On two transverse nonlinear models of axially moving beams. *Science in China Series E: Technological Sciences*, 52(3): 743-751, ISSN 1862-281X

Ding, H. & Chen, L.Q. (2009b). Nonlinear dynamics of axially accelerating viscoelastic beams based on differential quadrature. *Acta Mechanica Sinica Solida*, 22, 3, 267-275 ISSN 0894-9166

Feng, Z.H. & Hu, H.Y. (2002). Nonlinear dynamics modeling and periodic vibration of a cantilever beam subjected to axial movement of basement. *Acta Mechanica Solida Sinica* , 15, 2, 133-139, ISSN 0894-9166

Feng, Z.H. & Hu, H.Y. (2003). Principal parametric and three-to-one internal resonances of flexible beams undergoing a large linear motion. *Acta Mechanica Sinica*, 19, 4, 355-364, ISSN 0567-7718

Ghayesh, M.H. (2008). Nonlinear transversal vibration and stability of an axially moving viscoelastic string supported by a partial viscoelastic guide. *Journal of Sound and Vibration*, 314, 757-774, ISSN 0022-460X

Ghayesh, M.H. & Balar, S. (2008). Non-linear parametric vibration and stability of axially moving visco-elastic Rayleigh beams. *International Journal of Solids and Structures*, 45, 6451-6467, ISSN 0020-7683

Ghayesh, M.H. & Khadem, S.E. (2008). Rotary inertia and temperature effects on non-nonlinear vibration, steady-state response and stability of an axially moving beam with time-dependent velocity. *International Journal of Mechanical Sciences*, 50, 389-404, ISSN 0020-7403

Hochlenert, D.; Spelsberg-Korspeter, G. & Hagedorn, P. (2007). Friction induced vibrations in moving continua and their application to brake squeal. *ASME Journal of Applied Mechanics*, 74, 542-549, ISSN 1528-9036

Humer, A. & Irschik, H. 2009. Onset of transient vibrations of axially moving beams with large displacements, finite deformations and an initially unknown length of the reference configuration. *Zeitschrift fur Angewandte Mathematik und Mechanik*, 89, 4, 267-278, ISSN 0044-2267

Hwang, S.-J. & Perkins, N. C. (1992a). Supercritical stability of an axially moving beam part 1: model and equilibrium analysis. *Journal of Sound and Vibration*, 154, 381-396, ISSN 0022-460X

Hwang, S.-J. & Perkins, N. C. (1992b). Supercritical stability of an axially moving beam part 2: vibration and stability analysis. *Journal of Sound and Vibration*, 154, 397-409, ISSN 0022-460X

Hwang, S.-J. & Perkins, N. C. (1994). High speed stability of coupled band/wheel systems: theory and experiment. *Journal of Sound and Vibration*, 169, 459-483, ISSN 0022-460X

Koivurova, H. & Salonen, E.M. (1999). Comments on nonlinear formulations for travelling string and beam problems. *Journal of Sound and Vibration*, 225, 5, 845-856, ISSN 0022-460X

Kong, L. & Parker, R. G. (2004). Coupled belt-pulley vibration in serpentine drives with belt bending stiffness. *ASME Journal of Applied Mechanics*, 71, 109-119, ISSN 1528-9036

Lee, S. . & Mote, C.D.Jr. (1997). A generalized treatment of the energetics of translating continua, part 2: beams and fluid conveying pipes, *Journal of Sound and Vibration* 204, 735-753, ISSN 0022-460X

Marynowski, K. (2002). Non-linear dynamic analysis of an axially moving viscoelastic beam. *Journal of Theoretical And Applied Mechanics*, 40, ISSN 465-482, 0973-6085

Marynowski, K. (2004). Non-linear vibrations of an axially moving viscoelastic web with time-dependent tension. *Chaos, Solitons & Fractals*, 21, 481-490, ISSN 0960-0779

Marynowski, K. (2006). Two-dimensional rheological element in modelling of axially moving viscoelastic web. *European Journal of Mechanics A/Solids*, 25, 729-744, ISSN 0997-7538

Marynowski, K. & Kapitaniak, T. (2002). Kelvin-Voigt versus Bügers internal damping in modeling of axially moving viscoelastic web. *International Journal of Non-Linear Mechanics*, 37, 1147-1161, ISSN 0020-7462

Marynowski, K. & Kapitaniak, T. (2007). Zener internal damping in modelling of axially moving viscoelastic beam with time-dependent tension. *International Journal of Non-Linear Mechanics*, 42, 118-131, ISSN 0020-7462

Mockensturm, E. M. & Guo, J. (2005). Nonlinear vibration of parametrically excited, viscoelastic, axially moving strings. *ASME Journal of Applied Mechanics*, 72, 374-380, ISSN 1528-9036

Mote C.D.Jr. (1972). Dynamic stability of axially moving materials. *The Shock and Vibration Digest*. 4, 4, 3-13, ISSN 1741-3184

Mote C.D.Jr.; Schajer, G.S. & Wu, W.Z. (1982). Band saw and circular saw vibration and stability. *The Shock and Vibration Digest*. 14, 2, 19-25, ISSN 1741-3184

Maccari, A. (1999). The asymptotic perturbation method for nonlinear continuous systems. *Nonlinear Dynamics*, 19, 1-18, ISSN 1573-269X

Öz, H.R.; Pakdemirli, M. & Boyaci, H. (2001). Non-linear vibrations and stability of an axially moving beam with time-dependent velocity. *International Journal of Non-Linear Mechanics*, 36, 107-115, ISSN 0020-7462

Özhan, B. B. and Pakdemirli, M. (2009) A general solution procedure for the forced vibrations of a continuous system with cubic nonlinearities: Primary resonance case. *Journal of Sound and Vibration*, 325, 894-906, ISSN 0022-460X

Parker, R. G. & Lin, Y. (2001). Parametric instability of axially moving media subjected to multifrequency tension and speed fluctuations. *ASME Journal of Applied Mechanics*, 68, 49-57, ISSN 1528-9036

Pellicano, F.; Fregolent, A.; Bertuzzi, A. & Vestroni F. (2001). Primary and parametric non-linear resonances of a power transmission belt. *Journal of Sound and Vibration*, 244, 669-684, ISSN 0022-460X

Pellicano, F. & Vestroni, F. (2000). Nonlinear dynamics and bifurcations of an axially moving beam. *ASME Journal of Vibration and Acoustics*, 122, 21-30, ISSN 1528-8927

Pellicano, F. & Vestroni, F. (2002). Complex dynamic of high-speed axially moving systems. *Journal of Sound and Vibration*, 258, 31-44, ISSN 0022-460X

Pellicano, F. & Zirilli, F. (1997). Boundary layers and non-linear vibrations in an axially moving beam. *International Journal of Non-Linear Mechanics*, 33, 691-711, ISSN 0020-7462

Pratiher, B. & Dwivedy, S.K. (2008). Non-linear vibration of a single link viscoelastic Cartesian manipulator. *International Journal of Non-Linear Mechanics*, 43, 683-696, ISSN 0020-7462

Ravindra, B. & Zhu, W.D. (1998). Low dimensional chaotic response of axially accelerating continuum in the supercritical regime. *Archive of Applied Mechanics*, 68, 195-205, ISSN 1432-0681

Renshaw, A.A.; Rahn, C.D.; Wickert, J. & Mote, C.D.Jr. (1998). Energy and conserved functionals for axially moving materials. *ASME Journal of Vibration and Acoustics*, 120, 634-636, ISSN 1528-8927

Riedel, C. H. & Tan, C. A. (2002). Coupled, forced response of an axially moving strip with internal resonance. *International Journal of Non-Linear Mechanics*, 37, 101-116, ISSN 0020-7462

Spelsbrg-Korspeter, G; kirillov, & O.N. Hagedorn, P. (2008). Modeling and stability analysis of an axially moving beam with frictional contact. *ASME Journal of Applied Mechanics*, 75, 031001, ISSN 1528-9036

Shu, C. (2001). *Differential Quadrature and Its Application in Engineering*. Springer, ISBN 978-1-85233-209-9, Berlin

Sze, K.Y.; Chen, S.H. & Huang, J.L. (2005). The incremental harmonic balance method for nonlinear vibration of axially moving beams. *Journal of Sound and Vibration*, 281, 611-626, ISSN 0022-460X

Tang Y.Q.; Chen, L.Q., & Yang X.D. (2008). Natural frequencies, modes and critical speeds of axially moving Timoshenko beams with different boundary conditions. *International Journal of Mechanical Sciences*, 50 (10-11): 1448-1458, ISSN 0020-7403

Tang, Y.Q.; Chen, L.Q. & Yang, X.D. (2009). Non-linear vibrations of axially moving Timoshenko beams under weak and strong external excitations. *Journal of Sound and Vibration*, 320, 4/5, 1078-1099, ISSN 0022-460X

Tang, Y.Q.; Chen, L.Q. & Yang X.D. (2010). Parametric resonance of axially moving Timoshenko beams with time-dependent speed. *Nonlinear Dynamics*, 58, 715-724, ISSN 1573-269X

Thurman, A. L. & Mote Jr. C. D. (1969). Free, periodic, nonlinear oscillation of an axially moving strip. *ASME Journal of Applied Mechanics* 36, 83-91, ISSN 1528-9036

Tabarrok, B., Leech, C. M. & Kim, Y. I. (1974). On the dynamics of an axially moving beam. *Journal of The Franklin Institute*, 297, ISSN 201-220, 0016-0032

Ulsoy, A.G. & Mote, C.D.Jr. (1978). Band saw vibration and stability. *The Shock and Vibration Digest.* 10, 1, 3-15, ISSN 1741-3184

Vu-Quoc, L. & Li, S. (1995). Dynamics of sliding geometrically-exact beams: large angle maneuver and parametric resonance. *Computer Methods in Applied Mechanics and Engineering*, 120, 1995, ISSN 0045-7825

Wang, B. & Chen, L.Q. (2009). Asymptotic stability analysis with numerical confirmation of an axially accelerating beam constituted by the standard linear solid viscoelastic model. *Journal of Sound and Vibration*, 328, 456-466, ISSN 0022-460X

Wang, K. W. (1991). Dynamic stability analysis of high speed axially moving bands with end curvatures. *ASME Journal of Vibration and Acoustics*, 113, 62-68, ISSN 1528-8927

Wang, K. W. & Mote C. D. Jr. (1986). Vibration coupling analysis of Band/wheel mechanical systems. *Journal of Sound and Vibration*. 109, 237-258, ISSN 0022-460X

Wang, K. W. & Mote C. D. Jr. (1987). Band/wheel system vibration under impulsive boundary excitation. *Journal of Sound and Vibration*. 115, 203-216, ISSN 0022-460X

Wang, X. & Bert, C. W. (1993). A new approach in applying differential quadrature to static and free vibration analyses of beams and plates. *Journal of Sound and Vibration*, 162, 566-572, ISSN 0022-460X

Wickert, J.A. (1992). Non-linear vibration of a traveling tensioned beam. *International Journal of Non-Linear Mechanics*, 27, 503-517, ISSN 0020-7462

Wickert, J.A. & Mote, C.D.Jr. (1989). On the energetics of axially moving continua. *The Journal of the Acoustical Society of America*, 85, 1365-1368, ISSN 0001-4966

Wickert, J.A. & Mote, C.D.Jr. (1990). Classical vibration analysis of axially moving continua. *ASME Journal of Applied Mechanics*, 57, 738-744, ISSN 1528-9036

Wickert, J.A. & Mote, Jr. C. D. (1988). Current research on the vibration and stability of axially-moving materials. *The Shock and Vibration Digest.* 20, 5, 3-13, ISSN 1741-3184

Yang, T.Z.; Fang, B.; Chen, Y. & Zhen, Y.X. (2009). Approximate solutions of axially moving viscoelastic beams subject to multi-frequency excitations. *International Journal of Non-Linear Mechanics*, 44, 240-248, ISSN 0020-7462

Yang, X. D. & Chen, L.Q. (2005). Bifurcation and chaos of an axially accelerating viscoelastic beam. *Chaos, Solitons & Fractals*, 23(1): 249-258, ISSN 0960-0779

Yang, X. D. & Chen, L.Q. (2006). Non-linear forced vibration of axially moving viscoelastic beams. *Acta Mechanica Solida Sinica*, 19, 4, 365-373, ISSN 0894-9166

Zhang, L. & Zu, J.W. (1999). Nonlinear vibration of parametrically excited moving belts. *ASME Journal of Applied Mechanics*, 66, 396-402, ISSN 1528-9036

Zhang, W. & Song, C.Z. (2007). Higher-dimensional periodic and chaotic oscillations for viscoelastic moving belt with multiple internal resonances. *International Journal of Bifurcation and Chaos*, 17, 1637-1660, ISSN 1793-6551

Zhu, W.D. (2000). Vibration and stability of time-dependent translating media. *The Shock and Vibration Digest*. 32, 5, 369-379, ISSN 1741-3184

Zhu, W.D. & Ni, J.(2000). Energetics and stability of translating media with an arbitrarily varying length. *ASME Journal of Vibration and Acoustics*, 122, 295-304, ISSN 1528-8927

8

The 3D Nonlinear Dynamics of Catenary Slender Structures for Marine Applications

Ioannis K. Chatjigeorgiou and Spyros A. Mavrakos
National Technical University of Athens
Greece

1. Introduction

Riser systems are inextricable parts of integrated floating production and offloading systems as they are used to convey oil from the seafloor to the offshore unit. Risers are installed vertically or they are laid obtaining a catenary configuration. From the theoretical point of view they can be formulated as slender structures obeying to the principles of the Euler-Bernoulli beams. Riser-type catenary slender structures and especially Steel Catenary Risers (SCRs) attract the attention of industry for many years as they are very promising for deep water applications. According to the Committee V.5 of the International Ship and Offshore Structures Congress (ISSC, 2003), “flexible risers have been qualified to 1500m and are expected to be installed in depths up to 3000m in the next few years”. In such huge depths where the suspended length of the catenary will unavoidably count several kilometers, the equivalent elastic stiffness of the structure will be quite low enabling large displacements. The later remark implies that even small excitations could cause significant excursions in both in-plane and out-of-plane directions. Therefore a 2D formulation, although adequate in predicting the associated dynamics in the reference plane of the static equilibrium, it would be certainly a short approximation.

Furthermore, in deep water installations, for practical reasons mainly, the riser should be configured nearly as a vertical structure in order to avoid suspending more material. The nearly vertical configuration which ends in a sharp increase of the curvature close to the bottom, results in extreme bending moments at the touch down region. The static bending moment which is applied in the plane of reference of the catenary is further amplified due to the imposed excitation set by the motions of the floating structure. It has been generally acknowledged that the heave motion is the worst loading condition as it causes several effects, which depending on the properties of the excitation, can be applied individually or in combination between each other. Indicative examples are the seafloor interaction, buckling-like effects, “compression loading” and heave induced out-of-plane motions.

For the formulation of the seafloor interaction, various approaches have been proposed and it appears that the associated effects continue to attract the attention of the research community (Leira et al., 2004; Aubeny et al., 2006; Pesce et al., 2006; Clukey et al., 2008). “Compression loading” has been studied mainly in 2D (Passano & Larsen, 2006 & 2007; Chatjigeorgiou et al., 2007; Chatjigeorgiou, 2008), while buckling-like effects and possible destabilizations are mainly considered for completely vertical structures (Kuiper & Mertikine, 2005; Gadagi & Benaroya, 2006; Chandrasekaran et al., 2006; Kuiper et al., 2008).

The content of the present work falls in the last category of the effects that were mentioned previously. The main concern of the study is to identify the details of the out-of-plane response which is induced due to motions imposed in the catenary's plane of reference and in particular due to heave excitation. Relevant effects called as "Heave Induced Motions" have been investigated experimentally in the past by Joint Industry Projects (JIP). According to HILM (Heave Induced Lateral Motions of Steel Catenary Risers) JIP led by Institut français du pétrole (Ifp), the phenomenon was first recorded during the HCR (Highly Compliant Riser Large Scale Model Tests) JIP led by PMB Engineering, in which a steel catenary riser was excited by heave motion in a stillwater lake. The pipe was subjected to out-of-plane cyclic motions. The same behaviour was observed during the HILM JIP measurements (LeCunff et al., 2005).

Apparently, the associated phenomena can be captured numerically only by treating the governing 3D dynamical system. To this end, the associated system is properly elaborated and solved numerically using an efficient finite differences numerical scheme.

2. Definitions

A fully immersed catenary slender structure is considered. The catenary is modeled as an Euler-Bernoulli slender beam, having the following geometrical and physical properties: suspended length L, outer diameter d_o, inner diameter d_i, submerged weight w_o, mass m, hydrodynamic mass m_a, cross sectional area A and moment of inertia I. The quantities d_o, d_i, A and I, correspond to the unstretched condition, while w_o, m and m_a are defined per unit unstretched length. The Young modulus of elasticity is denoted by E and accordingly EA and EI define the elastic and bending stiffness respectively. Finally, it is assumed that the catenary conforms to a linear stress-strain relation.

Next the generalized motion and loading vectors (Fig. 1) are defined. These are

$$\vec{V}(s;t) = \begin{bmatrix} u & v & w & \phi & \theta \end{bmatrix}^T \tag{1}$$

$$\vec{F}(s;t) = \begin{bmatrix} T & S_n & S_b & M_b & M_n \end{bmatrix}^T \tag{2}$$

where u, v, w are the tangential (axial), normal and bi-normal velocities, respectively, ϕ is the Eulerian angle which is formed between the tangent of the line and the horizontal in the reference plane of the catenary, θ is the Eulerian angle in the out-of-plane direction, T is the tension, S_n and S_b are the in-plane and the out-of-plane shear forces and finally M_b and M_n are the bending moments around the corresponding Lagrangian axes $\vec{b}$ and $\vec{n}$, namely the generalized loading that causes bending in the in-plane and the out-of-plane direction, respectively. The moments M_b and M_n are associated with the corresponding curvatures Ω_b and Ω_n according to $M_j=EI\Omega_j$, for $j=n,b$.

In the general case where steady current is presented, the relative velocities should be considered. These are written as $v_{tr} = u - U_t$, $v_{nr} = v - U_n$ and $v_{br} = w - U_b$, where U_t, U_n and U_b are the components of the steady current parallel to $\vec{t}$, $\vec{n}$ and $\vec{b}$, respectively. The elements of the vectors defined through Eqs. (1) and (2) are all functions of time t and the unstretched Lagrangian coordinate s.

3. Dynamic system

The 3D dynamic equilibrium of the submerged catenary is governed by ten partial differential equations. These equations are provided in the following without further details on the derivation procedure. For more details the reader is referenced to the works of Howell (1992), Burgess (1993), Triantafyllou (1994) and Tjavaras et al. (1998).

$$m\left(\frac{\partial u}{\partial t}+w\frac{\partial \theta}{\partial t}-v\frac{\partial \phi}{\partial t}\cos\theta\right)=\frac{\partial T}{\partial s}+S_b\Omega_n-S_n\Omega_b-w_0\sin\phi\cos\theta+R_{dt} \tag{3}$$

$$m\left(\frac{\partial v}{\partial t}+\frac{\partial \phi}{\partial t}(u\cos\theta+w\sin\theta)\right)+m_a\frac{\partial v_{nr}}{\partial t}=\frac{\partial S_n}{\partial s}+\Omega_b(T+S_b\tan\theta)-w_0\cos\phi+R_{dn} \tag{4}$$

$$m\left(\frac{\partial w}{\partial t}-v\frac{\partial \phi}{\partial t}\sin\theta-u\frac{\partial \theta}{\partial t}\right)+m_a\frac{\partial v_{br}}{\partial t}=\frac{\partial S_b}{\partial s}-S_n\Omega_b\tan\theta-T\Omega_n-w_0\sin\phi\sin\theta+R_{db} \tag{5}$$

$$\frac{1}{EA}\frac{\partial T}{\partial t}=\frac{\partial u}{\partial s}+\Omega_n w-\Omega_b v \tag{6}$$

$$\left(1+\frac{T}{EA}\right)\frac{\partial \phi}{\partial t}\cos\theta=\frac{\partial v}{\partial s}+\Omega_b(u+w\tan\theta) \tag{7}$$

$$-\left(1+\frac{T}{EA}\right)\frac{\partial \theta}{\partial t}=\frac{\partial w}{\partial s}-\Omega_b v\tan\theta-\Omega_n u \tag{8}$$

$$EI\frac{\partial \Omega_n}{\partial s}=EI\Omega_b^2\tan\theta+S_b\left(1+\frac{T}{EA}\right)^3 \tag{9}$$

$$EI\frac{\partial \Omega_b}{\partial s}=EI\Omega_n\Omega_b\tan\theta-S_b\left(1+\frac{T}{EA}\right)^3 \tag{10}$$

$$\frac{\partial \theta}{\partial s}=\Omega_n \tag{11}$$

$$\frac{\partial \phi}{\partial s}\cos\theta=\Omega_b \tag{12}$$

In Eqs. (3)-(5) R_{dt}, R_{dn} and R_{db} denote the nonlinear drag forces which are expressed using the Morison's formula. Thus,

$$R_{dt}=-\frac{1}{2}\pi\rho d_o C_{dt}v_{tr}\left|v_{tr}\right|(1+e)^{1/2} \tag{13}$$

$$R_{dn}=-\frac{1}{2}\rho d_o C_{dn}v_{nr}\left|v_{nr}^2+v_{br}^2\right|^{1/2}(1+e)^{1/2} \tag{14}$$

$$R_{db} = -\frac{1}{2}\rho d_o C_{db} v_{br} \left| v_{nr}^2 + v_{br}^2 \right|^{1/2} (1+e)^{1/2} \tag{15}$$

where C_{dt}, C_{dn} and C_{db} are the drag coefficients in tangential, normal and bi-normal directions respectively. Normally, for a cylindrical structure, the in-plane and the out-of-plane drag coefficients are equal while the tangential coefficient is very small and the associated term can be ignored without loss of accuracy. Finally, e denotes the axial strain deformation, which for a linear stress-strain relation is written as $e=T/EA$.

4. Numerical solution of the governing system using finite differences

The numerical method employed herein, is the finite differences box approximation (Hoffman, 1993). Unlike the very popular finite element methods, the existing works which are related to the application of numerical approximations that rely on finite differences, concern mainly the dynamics of cables and mooring lines which have a negligible bending stiffness (Burgess, 1993; Tjavaras et al., 1998; Ablow & Schechter, 1983; Howell, 1991; Chatjigeorgiou & Mavrakos, 1999 & 2000; Gobat & Grosenbaugh, 2001 & 2006; Gobat et al., 2002). The employment of the bending stiffness in mathematical formulations of cable dynamics is done for special applications such as low tension cables, towing cables, highly extensible cables and mooring lines in which the cycling loading leads to slacking conditions, i.e. cancellation of the total tension.

With regard to the studies on pipes, for which the omission of the bending stiffness will unavoidably lead to loss of important information, the finite differences approximation has been used mainly for the solution of the static equilibrium problem (Zare & Datta, 1988; Jain 1994) or as a numerical scheme for the integration in the time domain, alternative to Houbolt, Wilson-θ and Newmark-β methods (Patel & Seyed, 1995). As far as the dynamic equilibrium problem is concerned, box approximation has been employed recently by Chatjigeorgiou (2008) for the development of a solution tool that treats the two dimensional nonlinear dynamics of marine catenary risers.

For the governing system at hand (Eqs. (3)-(12)), the recommended procedure for employing a finite differences approximation requires that the set of equations should be first cast in a matrix-vector form. Thus, the concerned equations are written as

$$\mathbf{M}\frac{\partial \mathbf{Y}}{\partial t} + \mathbf{K}\frac{\partial \mathbf{Y}}{\partial s} + \mathbf{F}(\mathbf{Y}_{,s,t}) = 0 \tag{16}$$

where $\mathbf{Y} = [u \quad v \quad w \quad T \quad \phi \quad \theta \quad S_b \quad S_n \quad \Omega_n \quad \Omega_b]^{\mathbf{T}}$. The mass and stiffness matrices, **M** and **K**, and the forcing vector **F** are defined in Appendix A.

Next, Eq. (16) is discretized in both time and space using the finite differences box approximation. This is the approach taken by several authors mentioned in the references section of the present work. With this scheme, the discrete equations are written using what look like traditional backward differences, but because the discetization is applied on the half-grid points the method is second-order accurate. The result is a four point average, centered around the half-grid point. Thus, Eq. (16) becomes

$$\left(\mathrm{M}_k^{i+1} + \mathrm{M}_k^i\right)\left(\frac{\mathrm{Y}_k^{i+1} - \mathrm{Y}_k^i}{\Delta t}\right) + \left(\mathrm{M}_{k-1}^{i+1} + \mathrm{M}_{k-1}^i\right)\left(\frac{\mathrm{Y}_{k-1}^{i+1} - \mathrm{Y}_{k-1}^i}{\Delta t}\right)$$

$$
\begin{aligned}
&+\left(\mathrm{K}_{k-1}^{i+1}+\mathrm{K}_{k}^{i+1}\right)\left(\frac{\mathrm{Y}_{k}^{i+1}-\mathrm{Y}_{k-1}^{i+1}}{\Delta s}\right)+\left(\mathrm{K}_{k-1}^{i}+\mathrm{K}_{k}^{i}\right)\left(\frac{\mathrm{Y}_{k}^{i}-\mathrm{Y}_{k-1}^{i}}{\Delta s}\right)\\
&+\left(\mathrm{F}_{k}^{i+1}+\mathrm{F}_{k-1}^{i+1}+\mathrm{F}_{k}^{i}+\mathrm{F}_{k-1}^{i}\right)=0
\end{aligned}
\qquad (17)
$$

According to the matrix-vector Eq. (17) the governing partial differential equations are defined in the center of [*i,i*+1] and [*k*-1,*k*], namely at [*i*+1/2, *k*-1/2]. The subscripts *k* define the spatial grid points (the nodes) and the superscripts *i* define the temporal grid points (the time steps). For *n* nodal points (*k*=1 corresponds to the touch down point at *s*=0 and *k*=*n* corresponds to the top terminal point where the excitation is applied) Eq. (17) defines a system of $10\cdot(n\text{-}1)$ equations to be solved for the $10\cdot n$ dependent variables at time step *i*+1. The ten equations needed to complete the problem are provided by boundary conditions.

The algebraic equivalents of the governing Eqs. (3)-(12) are derived using the grid transformation proposed by Eq. (17). The associated algebraic equations are given in Appendix B of the present paper. The boundary conditions which are needed to complete the final $10\cdot n$ algebraic system correspond to zero bending moments at both ends of the catenary, zero motions at the bottom fixed point and specified time depended excitations at the top in three directions. The final system is solved efficiently by the relaxation method.

5. Discussion on the contribution of the nonlinearities

The nonlinearities involved in the problem are either geometric or hydrodynamic nonlinearities. Here the current is ignored and accordingly, the hydrodynamic action is represented by the nonlinear drag forces induced due to the motions of the structure. It is noted that the presence of current could stimulate possible vortex-induced-vibration phenomena, the study of which exceeds the purposes of the present contribution. In addition the structure is slender and therefore the diffraction phenomena are negligible. This makes the drag forces the most determinative factor of hydrodynamic nature. Other hydrodynamic effects involved in the problem are the added inertia forces which are expressed through the added mass coefficients in the normal and the bi-normal directions.

Apart from the drag forces the dynamic equilibrium of the catenary involves also geometric nonlinearities. Apparently, the most important are the internal loading-curvature terms. The term "internal loading" refers to the tension and the shear forces. The question which easily arises is how nonlinear contributions influence the motions of the structure, namely the axial, the normal and the bi-normal displacements. It is evident that any excitation will induce displacements in the same direction but the question herein concerns the details of the motions which are induced in the other directions. The later remark is intimately connected with the so called "compression loading", i.e. the amplification of the bending moments at the touch down region due to the dynamic components. The importance of the subject regarding the in-plane bending moment has been extensively discussed by Passano and Larsen (2006) and Chatjigeorgiou et al. (2007). Here the discussion is extended to the out-of-plane bending moments as well.

In order to distinguish between the linear and the nonlinear effects it is indispensable to go through the equivalent linearized dynamic problem. It is assumed that the generalized loading terms and the Eulerian angles consist of a static and a dynamic component. These will be denoted in the sequel by the indexes 0 and 1 respectively. In addition small motions are considered. Thus the velocities are given by $u=\partial p/\partial t$, $v=\partial q/\partial t$ and $w=\partial r/\partial t$, where *p*, *q* and r are the motions in the axial, normal and bi-normal directions. Thus, the vector of the unknowns of the linear problem $\vec{Y}(s;t)=\begin{bmatrix}p & q & r & T & \phi & \theta & S_b & S_n & \Omega_n & \Omega_b\end{bmatrix}^{\mathbf{T}}$ becomes

$$\vec{Y}(s;t)=Y_0(s)+\vec{Y}_1(s;t) \tag{18}$$

where

$$Y_0(s)=\begin{bmatrix}0 & 0 & 0 & T_0 & \phi_0 & \theta_0 & S_{b0} & S_{n0} & \Omega_{n0} & \Omega_{b0}\end{bmatrix}^{\mathbf{T}} \tag{19}$$

and

$$\vec{Y}_1(s;t)=\begin{bmatrix}p & q & r & T_1 & \phi_1 & \theta_1 & S_{b1} & S_{n1} & \Omega_{n1} & \Omega_{b1}\end{bmatrix}^{\mathbf{T}} \tag{20}$$

The linearization procedure is outlined succinctly in the following. First, Eq. (18) is introduced into the nonlinear system of Eqs. (3)-(12). After short mathematical manipulations it can be seen that the resulting products will include the terms that define the static equilibrium problem as well as nonlinear components. Static equilibrium terms cancel each other while in the context of the linearized problem, the nonlinear terms are ignored. The compatibility relations given by Eqs. (6)-(8), are integrated with respect to time t. Finally, it is noted that the static terms Ω_{n0}, θ_0 and S_{b0} are zero. This is due to the two-dimensional static configuration of the catenary.

By employing the above procedure, the system of Eqs. (3)-(12) is reduced to the equivalent linearized system.

$$m\frac{\partial^2 p}{\partial t^2}=\frac{\partial T_1}{\partial s}-S_{n0}\Omega_{b1}-S_{n1}\Omega_{b0}-w_0\cos\phi_0\phi_1 \tag{21}$$

$$(m+m_a)\frac{\partial^2 q}{\partial t^2}=\frac{\partial S_{n1}}{\partial s}+T_1\Omega_{b0}+\Omega_{b1}T_0+w_0\sin\phi_0\phi_1-c_n\omega|q|\frac{\partial q}{\partial t} \tag{22}$$

$$(m+m_a)\frac{\partial^2 r}{\partial t^2}=\frac{\partial S_{b1}}{\partial s}-S_{n0}\Omega_{b0}\theta_1-T_0\Omega_{n1}-w_0\sin\phi_0\theta_1-c_b\omega|r|\frac{\partial r}{\partial t} \tag{23}$$

$$T_1=EA\left(\frac{\partial p}{\partial s}-\Omega_{b0}q\right) \tag{24}$$

$$\phi_1=\frac{\partial q}{\partial s}+\Omega_{b0}p \tag{25}$$

$$\theta_1=-\frac{\partial r}{\partial s} \tag{26}$$

$$EI\frac{\partial\Omega_{n1}}{\partial s}-EI\Omega_{b0}\theta_1=S_{b1} \tag{27}$$

$$EI\frac{\partial \Omega_{b1}}{\partial s} = -S_{n1} \tag{28}$$

$$\frac{\partial \theta_1}{\partial s} = \Omega_{n1} \tag{29}$$

$$\frac{\partial \phi_1}{\partial s} = \Omega_{b1} \tag{30}$$

In Eqs. (22) and (23) c_n=4/(3π)$\rho C_{dn} d_o$ and c_b=4/(3π)$\rho C_{db} d_o$ denote the linearized damping coefficients which are determined through the linearization process of the nonlinear drag forces R_{dn} and R_{db}. Also, the drag force in tangential direction was considered negligible, whereas the elastic strain e was set equal to zero.

Eqs. (21)-(30) consists of two major groups, namely one set that governs the coupled axial and normal motions (Eqs. (21), (22), (24), (25), (28) and (30)) and one set that governs the bi-normal or out-of-plane motions (Eqs. (23), (26), (27) and (29)). Provided that the solution of the static equilibrium problem is known, the two systems can be treated separately, which implies that, at least in the context of the linear problem, the in-plane motions do not influence the out-of-plane motions and vise versa. Thus, the axial and normal motions induced out-of plane vibrations is only due to the nonlinear terms and especially due to the geometric nonlinearities. This can be traced back to the fact that the out-of-plane static components Ω_{n0}, S_{b0} and θ_0, were assumed equal to zero. In fact, this is the actual case when the structure is perfect with no initial deformations, even marginal, and the excitations coincide absolutely with the unit vectors $\vec{t}$ and $\vec{n}$ for the in-plane motions and $\vec{b}$ for the out of plane motions.

For the linear problem, which by default assumes that the motions are relatively small, the in-plane and out-of-plane motions and their consequences, as regards the moments, the shear forces and the tension, can be considered uncoupled without loss of accuracy. Nevertheless, this is not a valid approach for the nonlinear problem. For a perfect structure however and assuming only in-plane excitations it will be easy to confirm, through the solution of the dynamic problem, that no out-of-plane motions are induced. This is a shortcoming of the theoretical methods which is associated with the disability to represent the marginal structural imperfections of the static configuration. However it is no difficult to invent numerical tricks to override this practical problem. In the present contribution for example, the numerical results which refer to the heave excitation induced out-of-plane motions, were obtained by exciting the structure at the top with a combined motion that consists of a vertical and a bi-normal component. The later is applied for a limited amount of time, which is enough to produce non-zero out-of-plane angles, bending moments and shear forces. Thus, at the cut-off time step the structure has obtained a 3D shape that explicitly diverges from the perfect in-plane configuration and is accordingly used as the initial condition for the subsequent time steps of the numerical simulation.

6. Numerical results and discussion

The numerical results which are presented in the following refer to the SCR that was used as a model by Passano and Larsen (2006). The same model was employed also by Chatjigeorgiou (2008). The physical and geometrical properties of the structure are: outer diameter 0.429m, wall thickness 0.0022m, Young modulus of elasticity 207GPa, mass per unit unstretched length 262.9kg/m, added mass per unit unstretched length 148.16kg/m, submerged weight per unit unstretched length 915.6N/m, suspended length 2024m, elastic stiffness $0.5823 \cdot 10^{10}$N and bending stiffness $0.1209 \cdot 10^{9}$Nm^{2}. The drag coefficients in normal and bi-normal directions were assumed equal to unity while the tangential drag coefficient was set equal to zero. Finally, with regard to the installation characteristics, the catenary was assumed suspended in water depth 1800m by applying a pretension at the top equal to 1860kN.

This work focuses mainly on the out-of-plane dynamics of the catenary, induced due to both in-plane and out-of-plane motions. More interesting from the academic point of view is the former type of excitation as in this case the out-of-plane motions are driven by nonlinearities.

6.1 Bi-normal (sway) excitation

Normally, nonlinear phenomena are stimulated at high frequencies and large amplitudes or by combining both properties, at high excitation velocities. Therefore in order to expose and study the associated impacts, the structure should be subjected to relatively severe loading. The details of the sway excitation are examined having the structure excited with a harmonic motion at the top with amplitude y_a=1.0m and circular frequency ω=2.0rad/s.

The solution in the time domain and especially the one that accounts for the nonlinear terms calculates the time histories of all time varying components at any point along the structure, providing huge data records, which admittedly, are hard to be handled. In addition, in a nonlinear formulation the records of the output signals will contain the contribution of sub- and super-harmonics which are difficult to be identified by inspecting only the time histories. Therefore, in order to present the results in a friendly and understandable format, all records were processed using Fast Fourier Transformation (FFT) and adopted to 3D spectrums. The spectrums reveal the prevailing frequencies at any point along the catenary. For the test case mentioned before, the 3D spectrums for the dynamic tension T_1, the normal velocity v, the in-plane dynamic bending moment M_{b1}, and the out-of-plane dynamic bending moment M_{n1} are depicted respectively in Figs. 2-5. It is noted that the out-of-plane dynamic bending moment also represents the total out-of-plane bending moment as the corresponding static counterpart is zero.

Fig. 5 shows that the out-of-plane bending moment responds at the excitation frequency. This occurs for all points along the catenary. The maximum value occurs just before the top terminal point where the excitation is applied. In addition, the variation of the out-of-plane bending moment as a function of s exhibits a dentate configuration with a notable increase at the touch down area. It is also important to note that no other harmonics are stimulated and the response is restricted to the frequency of excitation only.

Figs. 2-4 demonstrate that the in-plane response due to the sway excitation is much more complicated as various harmonics are detected. The most significant contribution comes from the double of the excitation frequency (4.0rad/s) while it is visually evident that there are peaks at $1/2\omega$, $3/2\omega$, 2ω, $5/2\omega$ and so on. The non-zero values of the spectral densities for $\omega \rightarrow 0$ or $T \rightarrow \infty$, which exhibit a different pattern for the various dynamic components, imply that the sway excitation causes a quasi-static application of the corresponding component. In addition, the non-zero values for $T \rightarrow \infty$, manifest that the response is in general non periodic and it is composed by a fundamental frequency that tends to infinity and practically a boundless number of harmonics.

6.2 In-plane heave excitation induced out-of-plane response

Here a single excitation case is examined that refers to excitation amplitude in heave z_a=1.0m with circular frequency ω=1.5rad/s. Again, a relatively high excitation velocity was assumed, in order to investigate the effect of nonlinearities. In the specific static configuration the heave motion acts nearly as an axial loading which, depending on the conditions, may result in "compression loading".

The details of the in-plane and the out-of-plane response due to the applied heave excitation are examined with the aid of Figs. 6-19. Figs. 6-8 are given as a part of the discussion, started in section 5, on the dependence of the out-of-plane motions, shear forces and bending moments by the initial static configuration. Figs. 6-7 demonstrate a dependence of the concerned variables on the amplitude of the sway excitation that is applied for practical reasons and for a short time, just to provide an initial out-of-plane deformation to the structure. Apparently, the records of the response, which in the specific case correspond to the location where the maximum static bending moment M_{b0} occurs, are different for different amplitudes. Nevertheless, the output signals converge for large amplitudes. The attainment of convergence is better shown in shear force S_{b1} (Fig. 8), as in this case the associated time history contains abnormal signals which however, do not dilute periodicity. Nevertheless, it should be noted that the impotence to formulate accurately the marginal static deformations in the out-of-plane direction, which it turn leads to the necessity to apply artificially non-zero values of $M_{n0}(s)$ and $\theta_0(s)$, constitutes in this connection, a numerical uncertainty.

Next, we focus for a while in Figs. 6-8. Fig. 8 is a little bit confusing whereas a careful inspection in Fig. 7 indicates the existence of a base harmonic and an additional harmonic. The two harmonics are more evident in the time history of the out-of-plane velocity w (Fig. 6) and it can be shown that they correspond to 0.75rad/s and 2.25rad/s. In other words none of the harmonics coincides with the excitation frequency. In particular, the concerned harmonics correspond to $1/2\omega$ and $3/2\omega$ where ω is the frequency of the excitation. Apparently the occurrence of these harmonics makes the motion of the structure quite complicated. The latter remark is graphically shown in Figs. 9-11 which demonstrate the path that is followed (in particular by node no 3 in a discretization grid of 100 nodes at s=41m from touch down point) as seen from behind (v=f(w)), from above (u=f(w)) and from the side (v=f(u)), respectively. It is noted that in Figs. 9-11 v and u respond following the excitation frequency ω while w responds having contributions from both $1/2\omega$ and $3/2\omega$. Fig. 9 shows that the general impression that the orbit of the structure follows a reclined "eight" configuration is not absolutely true. In fact, the motion is more complicated, mainly due to the contribution of $3/2\omega$. The reclined "eight" path or using a more symbolic term the "butterfly" motion, is more appropriate to be used in order to describe the motion of the structure from above, i.e. the function u=f(w). Finally, the fundamental frequency of the response for v and u which are both in-plane components is equal to the excitation frequency. This is shown with a more descriptive fashion in Fig. 11 where the function v=f(u) is represented by two coinciding closed loops.

Figs. 9-11 have been plotted using the numerical predictions of two periods of the steady state response. Another way to verify that the in-plane motions conform to the frequency of excitation is to observe that the two loops of Fig. 11 practically coincide. However, this is not the case when the out-of-plane motion is considered, which it is driven by a subharmonic and a superharmonic of the excitation frequency. In this case, each of the loops in Figs. 9 and 10 (right or left) is covered during one period of the excitation. Nevertheless, the fundamental frequency for the response of w, and in general for all out-of-plane components, is the half of the excitation frequency and accordingly the steady state motion at any point along the structure is completed after two excitation periods.

The contribution of the various harmonics, which are stimulated due to the heave excitation, to both the in-plane and the out-of-plane dynamic components, is better shown in the 3D spectral densities depicted in Figs. 12-17. Figs. 12-14 show in-plane components, namely the dynamic tension T_1 (Fig. 12), the normal velocity v (Fig. 13) and the in-plane dynamic bending moment M_{b1} (Fig. 14). In the respective plots it is immediately apparent that the in-plane components are primarily governed by the excitation frequency (ω=1.5rad/s in the present case study), while it is evident that the in-plane response is affected by additional harmonics that coincide with integer multipliers of the excitation frequency ω, i.e., 2ω, 3ω etc. The 2ω superharmonic is easily detectable in all three figures, whereas 3ω is seen (admittedly with relative difficulty), only in the dynamic tension spectral density (Fig. 12). It should be stated however that it exists, together with the higher integer multipliers, in all in-plane dynamic components.

Figs. 15-17 provide the 3D spectral densities of out-of-plane dynamic components, namely the bi-normal velocity w (Fig. 15), the out-of-plane dynamic bending moment M_{n1} (Fig. 16) and the out-of-plane dynamic shear force S_{b1} (Fig. 17). For enriching the discussion that preceded with regard the dominant harmonics of the out-of-plane response due to the heave excitation, it is again underlined that the motion herein is governed by frequencies that correspond to $1/2\omega$, $3/2\omega$, $5/2\omega$ etc. The occurrence of all three of them can be detected only in Fig. 15 (again, the latter is seen with relative difficulty), while for M_{n1} and S_{b1} the response appears to be governed by $1/2\omega$. Moreover, we could positively claim that there is a slight contribution from $3/2\omega$.

The question which easily arises is what exactly these findings mean. To provide an answer we could generalize the visual observations on the 3D spectral densities of the out-of-plane components and speculate that the contributing harmonics correspond to $(n/2)\cdot\omega$ for n=1,2,…. In addition, in order to be consistent with the above discussion we could claim that the even terms of the sequence are negligible. As far as the in-plane response is concerned, the logical sequence is to assume that the constituent harmonics could be approximated by the same simple formula, but in this case, the components which could be omitted are the odd terms of the sequence.

Correlating the above findings with the Mathieu equation, should not be considered as a significant discovery as many authors did the same in the past. Nevertheless most of the works in this subject discuss vertical slender structures (risers or tethers) (Gadagi & Benaroya, 2006; Chandrasekaran et al., 2006; Kuiper et al., 2008; Park & Jung, 2002) for marine applications where the heaving motions produce buckling and the associated dynamic behaviour is directly connected to Mathieu equation. To extend the discussion in the context of catenary structures, effort has been made to associate the numerical predictions depicted graphically in 3D spectral densities to the solution(s) of Mathieu equation. The issue for which we are mainly interested is that the global response consists of harmonics $(n/2)\cdot\omega$ for n=1,2,…, or equivalently $n\cdot(2\omega)$ for n=1,2,…, provided that the excitation frequency is 2ω. The Mathieu equation which is satisfied by periodic solutions is given for reference in the following:

$$\frac{d^2y(\tau)}{d\tau^2}+\left(a-2q\cos 2\tau\right)y(\tau)=0 \tag{31}$$

where $\tau=\omega t$ and q is referred as the Mathieu parameter. The solutions of Mathieu Eq. (31) associated with the characteristic values a, are given by (Abramowitz & Stegun, 1970; McLachlan, 1947; Meixner & Schäfke, 1954)

$$\mathrm{ce}_{2m}(\omega t;q)=\sum_{r=0}^{\infty}A_{2r}^{2m}(q)\cos[2r\omega t] \tag{32}$$

$$\mathrm{ce}_{2m+1}(\omega t;q)=\sum_{r=0}^{\infty}A_{2r+1}^{2m+1}(q)\cos[(2r+1)\omega t] \tag{33}$$

$$\mathrm{se}_{2m+1}(\omega t;q)=\sum_{r=0}^{\infty}B_{2r+1}^{2m+1}(q)\sin[(2r+1)\omega t] \tag{34}$$

$$\mathrm{se}_{2m+2}(\omega t;q)=\sum_{r=0}^{\infty}B_{2r+2}^{2m+2}(q)\sin[(2r+2)\omega t] \tag{35}$$

where ce_m and se_m are the even and odd periodic Mathieu functions and A and B are the associated constants depending on the Mathieu parameter q. It is immediately apparent that a stable solution of Mathieu Eq. (31) will include contributions originating from an infinite number of harmonics. In any case the first harmonic will be equal to $\omega/2$ provided that the excitation frequency is ω. It is reminded that according to the numerical results that describe the in-plane and the out-of-plane dynamic behaviour of the catenary structure due to heave excitation, the response was assumed to include the same type and number of harmonics regardless whether they are significant or not. The answer to the question why the in-plane motions are governed by the harmonics ω, 2ω, 3ω,…, and the out-of-plane motions by the harmonics $\omega/2$, $3\omega/2$, $5\omega/2$,…is apparently a difficult task that requires deep and comprehensive investigation and it could be the subject for a future work.

7. Conclusion

The 3D dynamic behaviour of catenary slender structures for marine applications was considered. The investigation was based on the results obtained by solving the complete nonlinear governing system that consists of ten partial differential equations. The solution method employed was the finite differences box approximation. Particular attention was given to the out-of-plane variables which are induced due to heave excitation.

The main finding in this context was the contribution of several harmonics that influence the global response of the structure. In fact it was shown that under in-plane heave excitation at the top terminal point the in-plane variables, motions and generalized loading components, are governed by the harmonics ω, 2ω, 3ω,…, whereas the out-of-plane variables by the harmonics $\omega/2$, $3\omega/2$, $5\omega/2$,…

For the heave induced out-of-plane motions, the fundamental frequency is exactly the half of the excitation frequency. This leads to cyclic motions which are completed during a time interval that is equal to the double of the excitation period. It was shown graphically that the orbit of the structure resembles a "butterfly" configuration. This interesting behaviour was correlated to the even and odd periodic solutions of the canonical form of Mathieu equation. Finally, the contribution of the nonlinearities was studied by deriving the equivalent linearized system and it was commented that the out-of-plane motions induced due to in-plane excitation are driven by the geometric nonlinear terms.

8. References

Ablow, C.M & Schechter, S. (1983). Numerical simulation of undersea cable dynamics. *Ocean Engineering*; 10, 443-457

Abramowitz, M. & Stegun I.A. (1970). *Handbook of mathematical functions*, Dover Publications Inc, New York

Aubeny, C.P., Biscotin, G. & Zhang, J (2006). *Seafloor interaction with steel catenary risers*, Final Project Report, MMS Project No 510, Texas A&M University

Burgess, J.J. (1993). Bending stiffness in a simulation of undersea cable deployment., *International Journal of Offshore and Polar Engineering*, 3, 197-204

Chandrasekaran, S., Chandak, N.R. & Anupam, G. (2006). Stability analysis of TLP tethers, *Ocean Engineering,* 33, 471-482.

Chatjigeorgiou, I.K. & Mavrakos, S.A. (1999). Comparison of numerical methods for predicting the dynamic behavior of mooring lines. *Proceedings of the 9th International Conference on Offshore and Polar Engineering (ISOPE 1999)*, Brest, France, Vol. II, 332-339

Chatjigeorgiou, I.K. & Mavrakos, S.A. (2000). Comparative evaluation of numerical schemes for 2D mooring dynamics, *International Journal of Offshore and Polar Engineering,* 10, 301-309

Chatjigeorgiou, I.K., Passano, E. & Larsen, C.M. (2007). Extreme bending moments on long catenary risers due to heave excitation, *Proceedings of the 26th International Conference on Offshore Mechanics and Arctic Engineering (OMAE 2007)*, San Diego, California, USA, Paper No 29384.

Chatjigeorgiou, I.K. (2008). A finite differences formulation for the linear and nonlinear dynamics of 2D catenary risers, *Ocean Engineering,* 35, 616-636.

Clukey, E., Jacob, P. & Sharma, P. (2008). Investigation of riser seafloor interaction using explicit finite element methods. *Offshore Technology Conference*, Houston, Texas, OTC 19432

Gadagi, M.M. & Benaroya, H. (2006). Dynamic response of an axially loaded tendon of a tension leg platform, *Journal of Sound and Vibration,* 293, 38-58.

Gobat, J.I. & Grosenbaugh, M.A. (2001). Application of the generalized-*a* method to the time integration of the cable dynamics equations, *Computer Methods in Applied Mechanics and Engineering,* 190, 4817-4829.

Gobat, J.I., Grosenbaugh, M.A. & Triantafyllou, M.S. (2002). Generalized-*a* time integration solutions for hanging chain dynamics. *Journal of Engineering Mechanics – ASCE,* 128, 677-687

Gobat, J.I. & Grosenbaugh, M.A. (2006). Time-domain numerical simulation of ocean cable structures, *Ocean Engineering,* 33, 1373-1400.

Hoffman, J.D. (1993). *Numerical methods for engineers and scientists,* McGraw-Hill, New York

Howell, C.T. (1991). Numerical analysis of 2-D nonlinear cable equations with applications to low tension problems, *Proceedings of the 1st International Offshore and Polar Engineering Conference (ISOPE 1991)*, Edinburgh, United Kingdom, Vol. II, 203-209.

Howell, C.T. *Investigation of the dynamics of low tension cables*, PhD Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts

ISSC (2003). *Report of the committee V.5: Floating Production Systems*. Elsevier Science, Oxford, Eds A.E. Mansour, R.C. Ertekin.

Jain, A.K. Review of flexible risers and articulated storage systems, *Ocean Engineering,* 21, 733-750.

Kuiper, G.L. & Metrikine, AV. (2005). Dynamic stability of a submerged, free-hanging riser conveying fluid, *Journal of Sound and Vibration,* 280, 1051–1065.

Kuiper, G.L., Brugmans, J. & Metrikine, AV. (2008). Destabilization of deep-water risers by a heaving platform, *Journal of Sound and Vibration,* 310, 541-557.

LeCunff, C., Biolley, F. & Damy, G. (2005). Experimental and numerical study of heave induced lateral motion (HILM), *Proceedings of the 24th International Conference on Offshore Mechanics and Arctic Engineering (OMAE 2005)*, Halkidiki, Greece, Paper No 67019

Leira, B.J., Karunakaran, D., Giertsen, E., Passano, E. & Farnes, K-A. Analysis guidelines and application of a riser-soil interaction model including trench effects, *Proceedings of the 23rd International Conference on Offshore Mechanics and Arctic Engineering (OMAE 2004)*, Vancouver, Canada, Paper No 51527

McLachlan N.W. (1947). *Theory and applications of Mathieu functions*, Dover Publications, New York.

Meixner J. & Schäfke F.W. (1954). *Mathieusche funktionen und sphäroidfunktionen,* Springer, Berlin

Milinazzo, F., Wilkie, M. & Latchman, S.A. (1987). An efficient algorithm for simulating the dynamics of towed cable systems, *Ocean Engineering,* 14, 513-526.

Park, H-I., & Jung, D-H. (2002). A finite element method for dynamic analysis of long slender marine structures under combined parametric and forcing excitations, *Ocean Engineering,* 29, 1313-1325.

Passano, E. & Larsen, C.M. (2006). Efficient analysis of a catenary riser, *Proceedings of the 25th International Conference on Offshore Mechanics and Arctic Engineering (OMAE 2006)*, Hamburg, Germany, Paper No 92308

Passano, E. & Larsen, C.M. (2007). Estimating distributions for extreme response of a catenary riser. *Proceedings of the 26th International Conference on Offshore Mechanics and Arctic Engineering (OMAE 2007)*, San Diego, California, USA, Paper No 29547.

Patel, H.M. & Seyed, F.B. (1995). Review of flexible risers modelling and analysis techniques, *Engineering Structures,* 17, 293-304.

Pesce, C.P., Martins, C.A. & Silveira, L.M.Y. (2006). Riser-soil interaction: Local dynamics at TDP and a discussion on the eigenvalue and the VIV problems, *Journal of Offshore Mechanics and Arctic Engineering,* 128, 39-55.

Tjavaras, A.A., Zhu, Q., Liu, Y., Triantafyllou, M.S. & Yue, D.K.P. (1998). The mechanics of highly extensible cables, *Journal of Sound and Vibration,* 213, 709-737

Triantafyllou, M.S. (1994). Cable mechanics for moored floating structures. *Proceedings of the 7th International Conference on the Behaviour of Offshore Structures (BOSS 1994)*, Boston, Massachusetts, Vol. 2, 57-77.

Zare, K. & Datta, T.K. (1988). Vibration of Lazy-“S” risers due to vortex shedding under lock-in, *Proceedings of the 20th Offshore Technology Conference*, OTC 5795.

Appendix A. Mass matrix M, stiffness matrix K and forcing vector F of Eq. (16)

$$\mathrm{M}=\begin{bmatrix} -m & 0 & 0 & 0 & mv\cos\theta & -mw & 0 & 0 & 0 & 0 \\ 0 & -m-m_a & 0 & 0 & -m(u\cos\theta+w\sin\theta) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -m-m_a & 0 & mv\sin\theta & mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\dfrac{1}{EA} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\left(1+\dfrac{T}{EA}\right)\cos\theta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \left(1+\dfrac{T}{EA}\right) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (A.1)$$

$$\mathrm{K}=\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & EI & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & EI \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos\theta & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (A.2)$$

$$\mathrm{F}=\begin{bmatrix} S_b\Omega_n - S_n\Omega_b - w_0\sin\phi\cos\theta + R_{dt} \\ \Omega_b\left(T+S_b\tan\theta\right) - w_0\cos\phi + R_{dn} + m_a\dfrac{\partial U_n}{\partial t} \\ -S_n\Omega_b\tan\theta - T\Omega_n - w_0\sin\phi\sin\theta + R_{db} + m_a\dfrac{\partial U_b}{\partial t} \\ \Omega_n w - \Omega_b v \\ \Omega_b\left(u + w\tan\theta\right) \\ -\Omega_b v\tan\theta - \Omega_n u \\ EI\Omega_b^2\tan\theta + S_b\left(1+\dfrac{T}{EA}\right)^3 \\ EI\Omega_n\Omega_b\tan\theta - S_b\left(1+\dfrac{T}{EA}\right)^3 \\ -\Omega_n \\ -\Omega_b \end{bmatrix} \quad (A.3)$$

Appendix B. Algebraic expansions of the nonlinear system of dynamic equilibrium Eqs. (3)-(12) using the finite differences box scheme

$$\begin{aligned}
E_1 &= \frac{T_k^{i+1} + T_k^i - T_{k-1}^{i+1} - T_{k-1}^i}{2\Delta s} + \frac{1}{4}\left(S_{bk}^{i+1}\Omega_{nk}^{i+1} + S_{bk}^{i}\Omega_{nk}^{i} + S_{bk-1}^{i+1}\Omega_{nk-1}^{i+1} + S_{bk-1}^{i}\Omega_{nk-1}^{i}\right) \\
&- \frac{1}{4}\left(S_{nk}^{i+1}\Omega_{bk}^{i+1} + S_{nk}^{i}\Omega_{bk}^{i} + S_{nk-1}^{i+1}\Omega_{bk-1}^{i+1} + S_{nk-1}^{i}\Omega_{bk-1}^{i}\right) \\
&- \frac{w_0}{4}\left(\sin\phi_k^{i+1}\cos\theta_k^{i+1} + \sin\phi_k^{i}\cos\theta_k^{i} + \sin\phi_{k-1}^{i+1}\cos\theta_{k-1}^{i+1} + \sin\phi_{k-1}^{i}\cos\theta_{k-1}^{i}\right) \\
&- \frac{1}{2}\pi\rho d_o C_{dt}\frac{1}{4}\left[v_{trk}^{i+1}\left|v_{trk}^{i+1}\right|\left(1 + e_k^{i+1}\right)^{1/2} + v_{trk}^{i}\left|v_{trk}^{i}\right|\left(1 + e_k^{i}\right)^{1/2} \right. \\
&\left. v_{trk-1}^{i+1}\left|v_{trk-1}^{i+1}\right|\left(1 + e_{k-1}^{i+1}\right)^{1/2} + v_{trk-1}^{i}\left|v_{trk-1}^{i}\right|\left(1 + e_{k-1}^{i}\right)^{1/2} \right] \\
&- m\left[\frac{u_k^{i+1} + u_{k-1}^{i+1} - u_k^i - u_{k-1}^i}{2\Delta t} + \frac{1}{4}\left(w_k^{i+1} + w_{k-1}^{i+1} + w_k^i + w_{k-1}^i\right)\frac{\theta_k^{i+1} + \theta_{k-1}^{i+1} - \theta_k^i - \theta_{k-1}^i}{2\Delta t} \right. \\
&\left. - \frac{1}{4}\left(v_k^{i+1}\cos\theta_k^{i+1} + v_k^{i}\cos\theta_k^{i} + v_{k-1}^{i+1}\cos\theta_{k-1}^{i+1} + v_{k-1}^{i}\cos\theta_{k-1}^{i}\right)\frac{\phi_k^{i+1} + \phi_{k-1}^{i+1} - \phi_k^i - \phi_{k-1}^i}{2\Delta t} \right] = 0
\end{aligned} \tag{B.1}$$

$$\begin{aligned}
E_2 &= \frac{S_{nk}^{i+1} + S_{nk}^{i} - S_{nk-1}^{i+1} - S_{nk-1}^{i}}{2\Delta s} + \frac{1}{4}\left(T_k^{i+1}\Omega_{bk}^{i+1} + T_k^{i}\Omega_{bk}^{i} + T_{k-1}^{i+1}\Omega_{bk-1}^{i+1} + T_{k-1}^{i}\Omega_{bk-1}^{i}\right) \\
&+ \frac{1}{4}\left(\Omega_{bk}^{i+1}S_{bk}^{i+1}\tan\theta_k^{i+1} + \Omega_{bk}^{i}S_{bk}^{i}\tan\theta_k^{i} + \Omega_{bk-1}^{i+1}S_{bk-1}^{i+1}\tan\theta_{k-1}^{i+1} + \Omega_{bk-1}^{i}S_{bk-1}^{i}\tan\theta_{k-1}^{i}\right) \\
&- \frac{w_0}{4}\left(\cos\phi_k^{i+1} + \cos\phi_k^{i} + \cos\phi_{k-1}^{i+1} + \cos\phi_{k-1}^{i}\right) \\
&- \frac{1}{2}\rho d_o C_{dn}\frac{1}{4}\left[v_{nrk}^{i+1}\left|\left(v_{nrk}^{i+1}\right)^2 + \left(v_{brk}^{i+1}\right)^2\right|^{1/2}\left(1 + e_k^{i+1}\right)^{1/2} + v_{nrk}^{i}\left|\left(v_{nrk}^{i}\right)^2 + \left(v_{brk}^{i}\right)^2\right|^{1/2}\left(1 + e_k^{i}\right)^{1/2} \right. \\
&\left. v_{nrk-1}^{i+1}\left|\left(v_{nrk-1}^{i+1}\right)^2 + \left(v_{brk-1}^{i+1}\right)^2\right|^{1/2}\left(1 + e_{k-1}^{i+1}\right)^{1/2} + v_{nrk-1}^{i}\left|\left(v_{nrk-1}^{i}\right)^2 + \left(v_{brk-1}^{i}\right)^2\right|^{1/2}\left(1 + e_{k-1}^{i}\right)^{1/2} \right] \\
&- m\frac{v_k^{i+1} + v_{k-1}^{i+1} - v_k^i - v_{k-1}^i}{2\Delta t} - m_a\frac{v_{nrk}^{i+1} + v_{nrk-1}^{i+1} - v_{nrk}^{i} - v_{nrk-1}^{i}}{2\Delta t} \\
&- \frac{m}{4}\left(u_k^{i+1}\cos\theta_k^{i+1} + u_k^{i}\cos\theta_k^{i} + u_{k-1}^{i+1}\cos\theta_{k-1}^{i+1} + u_{k-1}^{i}\cos\theta_{k-1}^{i}\right)\frac{\phi_k^{i+1} + \phi_{k-1}^{i+1} - \phi_k^i - \phi_{k-1}^i}{2\Delta t} \\
&- \frac{m}{4}\left(w_k^{i+1}\sin\theta_k^{i+1} + w_k^{i}\sin\theta_k^{i} + w_{k-1}^{i+1}\sin\theta_{k-1}^{i+1} + w_{k-1}^{i}\sin\theta_{k-1}^{i}\right)\frac{\phi_k^{i+1} + \phi_{k-1}^{i+1} - \phi_k^i - \phi_{k-1}^i}{2\Delta t} = 0
\end{aligned} \tag{B.2}$$

$$\begin{aligned}
E_3 &= \frac{S_{b\,k}^{i+1} + S_{b\,k}^{i} - S_{b\,k-1}^{i+1} - S_{b\,k-1}^{i}}{2\Delta s} - \frac{1}{4}\left(T_k^{i+1}\Omega_{n\,k}^{i+1} + T_k^{i}\Omega_{n\,k}^{i} + T_{k-1}^{i+1}\Omega_{n\,k-1}^{i+1} + T_{k-1}^{i}\Omega_{n\,k-1}^{i}\right) \\
&- \frac{1}{4}\left(\Omega_{b\,k}^{i+1}S_{n\,k}^{i+1}\tan\theta_k^{i+1} + \Omega_{b\,k}^{i}S_{n\,k}^{i}\tan\theta_k^{i} + \Omega_{b\,k-1}^{i+1}S_{n\,k-1}^{i+1}\tan\theta_{k-1}^{i+1} + \Omega_{b\,k-1}^{i}S_{n\,k-1}^{i}\tan\theta_{k-1}^{i}\right) \\
&- \frac{w_0}{4}\left(\sin\phi_k^{i+1}\sin\theta_k^{i+1} + \sin\phi_k^{i}\sin\theta_k^{i} + \sin\phi_{k-1}^{i+1}\sin\theta_{k-1}^{i+1} + \sin\phi_{k-1}^{i}\sin\theta_{k-1}^{i}\right) \\
&- \frac{1}{2}\rho d_o C_{db}\frac{1}{4}\Bigg[v_{br\,k}^{i+1}\left|\left(v_{nr\,k}^{i+1}\right)^2 + \left(v_{br\,k}^{i+1}\right)^2\right|^{1/2}\left(1+e_k^{i+1}\right)^{1/2} + v_{br\,k}^{i}\left|\left(v_{nr\,k}^{i}\right)^2 + \left(v_{br\,k}^{i}\right)^2\right|^{1/2}\left(1+e_k^{i}\right)^{1/2} \\
&v_{br\,k-1}^{i+1}\left|\left(v_{nr\,k-1}^{i+1}\right)^2 + \left(v_{br\,k-1}^{i+1}\right)^2\right|^{1/2}\left(1+e_{k-1}^{i+1}\right)^{1/2} + v_{br\,k-1}^{i}\left|\left(v_{nr\,k-1}^{i}\right)^2 + \left(v_{br\,k-1}^{i}\right)^2\right|^{1/2}\left(1+e_{k-1}^{i}\right)^{1/2}\Bigg] \\
&- m\frac{w_k^{i+1} + w_{k-1}^{i+1} - w_k^{i} - w_{k-1}^{i}}{2\Delta t} - m_a\frac{v_{br\,k}^{i+1} + v_{br\,k-1}^{i+1} - v_{br\,k}^{i} - v_{br\,k-1}^{i}}{2\Delta t} \\
&+ \frac{m}{4}\left(v_k^{i+1}\sin\theta_k^{i+1} + v_k^{i}\sin\theta_k^{i} + v_{k-1}^{i+1}\sin\theta_{k-1}^{i+1} + v_{k-1}^{i}\sin\theta_{k-1}^{i}\right)\frac{\phi_k^{i+1} + \phi_{k-1}^{i+1} - \phi_k^{i} - \phi_{k-1}^{i}}{2\Delta t} \\
&+ \frac{m}{4}\left(u_k^{i+1} + u_{k-1}^{i+1} + u_k^{i} + u_{k-1}^{i}\right)\frac{\theta_k^{i+1} + \theta_{k-1}^{i+1} - \theta_k^{i} - \theta_{k-1}^{i}}{2\Delta t} = 0
\end{aligned} \tag{B.3}$$

$$\begin{aligned}
E_4 &= \frac{u_k^{i+1} + u_k^{i} - u_{k-1}^{i+1} - u_{k-1}^{i}}{2\Delta s} + \frac{1}{4}\left(w_k^{i+1}\Omega_{n\,k}^{i+1} + w_k^{i}\Omega_{n\,k}^{i} + w_{k-1}^{i+1}\Omega_{n\,k-1}^{i+1} + w_{k-1}^{i}\Omega_{n\,k-1}^{i}\right) \\
&- \frac{1}{4}\left(v_k^{i+1}\Omega_{b\,k}^{i+1} + v_k^{i}\Omega_{b\,k}^{i} + v_{k-1}^{i+1}\Omega_{b\,k-1}^{i+1} + v_{k-1}^{i}\Omega_{b\,k-1}^{i}\right) - \frac{1}{EA}\frac{T_k^{i+1} + T_{k-1}^{i+1} - T_k^{i} - T_{k-1}^{i}}{2\Delta t} = 0
\end{aligned} \tag{B.4}$$

$$\begin{aligned}
E_5 &= \frac{v_k^{i+1} + v_k^{i} - v_{k-1}^{i+1} - v_{k-1}^{i}}{2\Delta s} + \frac{1}{4}\left(u_k^{i+1}\Omega_{b\,k}^{i+1} + u_k^{i}\Omega_{b\,k}^{i} + u_{k-1}^{i+1}\Omega_{b\,k-1}^{i+1} + u_{k-1}^{i}\Omega_{b\,k-1}^{i}\right) \\
&+ \frac{1}{4}\left(w_k^{i+1}\Omega_{b\,k}^{i+1}\tan\theta_k^{i+1} + w_k^{i}\Omega_{b\,k}^{i}\tan\theta_k^{i} + w_{k-1}^{i+1}\Omega_{b\,k-1}^{i+1}\tan\theta_{k-1}^{i+1} + w_{k-1}^{i}\Omega_{b\,k-1}^{i}\tan\theta_{k-1}^{i}\right) \\
&- \frac{1}{4}\Big[\left(1+e_k^{i+1}\right)\cos\theta_k^{i+1} + \left(1+e_k^{i}\right)\cos\theta_k^{i} + \left(1+e_{k-1}^{i+1}\right)\cos\theta_{k-1}^{i+1} \\
&+ \left(1+e_{k-1}^{i}\right)\cos\theta_{k-1}^{i}\Big]\frac{\phi_k^{i+1} + \phi_{k-1}^{i+1} - \phi_k^{i} - \phi_{k-1}^{i}}{2\Delta t} = 0
\end{aligned} \tag{B.5}$$

$$\begin{aligned}
E_6 &= \frac{w_k^{i+1} + w_k^{i} - w_{k-1}^{i+1} - w_{k-1}^{i}}{2\Delta s} - \frac{1}{4}\left(u_k^{i+1}\Omega_{n\,k}^{i+1} + u_k^{i}\Omega_{n\,k}^{i} + u_{k-1}^{i+1}\Omega_{n\,k-1}^{i+1} + u_{k-1}^{i}\Omega_{n\,k-1}^{i}\right) \\
&- \frac{1}{4}\left(v_k^{i+1}\Omega_{b\,k}^{i+1}\tan\theta_k^{i+1} + v_k^{i}\Omega_{b\,k}^{i}\tan\theta_k^{i} + v_{k-1}^{i+1}\Omega_{b\,k-1}^{i+1}\tan\theta_{k-1}^{i+1} + v_{k-1}^{i}\Omega_{b\,k-1}^{i}\tan\theta_{k-1}^{i}\right) \\
&+ \frac{1}{4}\Big[\left(1+e_k^{i+1}\right) + \left(1+e_k^{i}\right) + \left(1+e_{k-1}^{i+1}\right) + \left(1+e_{k-1}^{i}\right)\Big]\frac{\theta_k^{i+1} + \theta_{k-1}^{i+1} - \theta_k^{i} - \theta_{k-1}^{i}}{2\Delta t} = 0
\end{aligned} \tag{B.6}$$

$$\begin{aligned}E_7 &= EI\frac{\Omega_{nk}^{i+1}+\Omega_{nk}^{i}-\Omega_{nk-1}^{i+1}-\Omega_{nk-1}^{i}}{2\Delta s}-\frac{EI}{4}\Big(\Omega_{bk}^{i+1}\Omega_{bk}^{i+1}\tan\theta_k^{i+1}+\Omega_{bk}^{i}\Omega_{bk}^{i}\tan\theta_k^{i}\\ &+\Omega_{bk-1}^{i+1}\Omega_{bk-1}^{i+1}\tan\theta_{k-1}^{i+1}+\Omega_{bk-1}^{i}\Omega_{bk-1}^{i}\tan\theta_{k-1}^{i}\Big)\\ &-\frac{1}{4}\left[S_{bk}^{i+1}\left(1+e_k^{i+1}\right)^3+S_{bk}^{i}\left(1+e_k^{i}\right)^3+S_{bk-1}^{i+1}\left(1+e_{k-1}^{i+1}\right)^3+S_{bk-1}^{i}\left(1+e_{k-1}^{i}\right)^3\right]=0\end{aligned}\tag{B.7}$$

$$\begin{aligned}E_8 &= EI\frac{\Omega_{bk}^{i+1}+\Omega_{bk}^{i}-\Omega_{bk-1}^{i+1}-\Omega_{bk-1}^{i}}{2\Delta s}-\frac{EI}{4}\Big(\Omega_{nk}^{i+1}\Omega_{bk}^{i+1}\tan\theta_k^{i+1}+\Omega_{nk}^{i}\Omega_{bk}^{i}\tan\theta_k^{i}\\ &+\Omega_{nk-1}^{i+1}\Omega_{bk-1}^{i+1}\tan\theta_{k-1}^{i+1}+\Omega_{nk-1}^{i}\Omega_{bk-1}^{i}\tan\theta_{k-1}^{i}\Big)\\ &+\frac{1}{4}\left[S_{nk}^{i+1}\left(1+e_k^{i+1}\right)^3+S_{nk}^{i}\left(1+e_k^{i}\right)^3+S_{nk-1}^{i+1}\left(1+e_{k-1}^{i+1}\right)^3+S_{nk-1}^{i}\left(1+e_{k-1}^{i}\right)^3\right]=0\end{aligned}\tag{B.8}$$

$$E_9=\frac{\theta_k^{i+1}+\theta_k^{i}-\theta_{k-1}^{i+1}-\theta_{k-1}^{i}}{2\Delta s}-\frac{1}{4}\left(\Omega_{nk}^{i+1}+\Omega_{nk}^{i}+\Omega_{nk-1}^{i+1}+\Omega_{nk-1}^{i}\right)=0\tag{B.9}$$

$$\begin{aligned}E_{10}&=\frac{1}{4}\left(\cos\theta_k^{i+1}+\cos\theta_k^{i}+\cos\theta_{k-1}^{i+1}+\cos\theta_{k-1}^{i}\right)\frac{\phi_k^{i+1}+\phi_k^{i}-\phi_{k-1}^{i+1}-\phi_{k-1}^{i}}{2\Delta s}\\ &-\frac{1}{4}\left(\Omega_{bk}^{i+1}+\Omega_{bk}^{i}+\Omega_{bk-1}^{i+1}+\Omega_{bk-1}^{i}\right)=0\end{aligned}\tag{B.10}$$

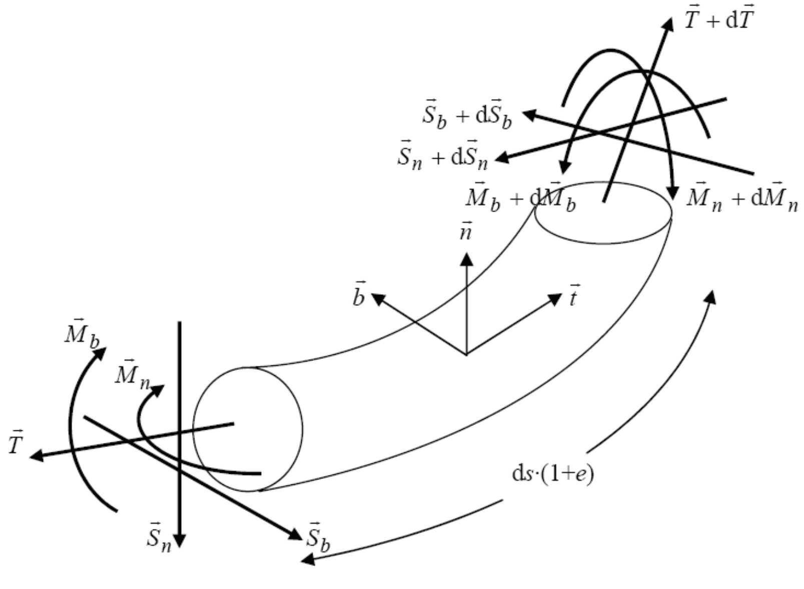

$\vec{T}+\mathrm{d}\vec{T}$
$\vec{S}_b+\mathrm{d}\vec{S}_b$
$\vec{S}_n+\mathrm{d}\vec{S}_n$
$\vec{M}_b+\mathrm{d}\vec{M}_b$
$\vec{M}_n+\mathrm{d}\vec{M}_n$
$\vec{n}$
$\vec{b}$
$\vec{t}$
$\vec{M}_b$
$\vec{M}_n$
$\vec{T}$
$\mathrm{d}s\cdot(1+e)$
$\vec{S}_n$
$\vec{S}_b$

Fig. 1. Stretched catenary segment and balance of internal loading.

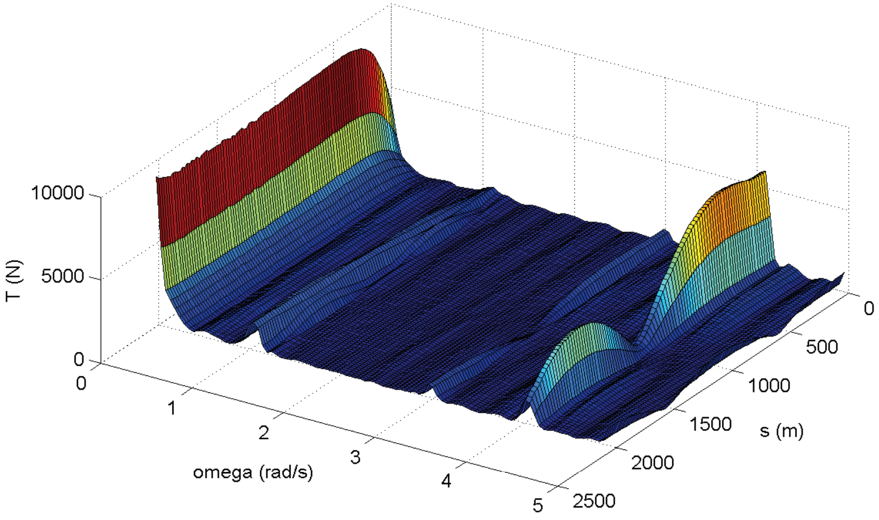


Fig. 2. Spectral densities of the dynamic tension T_1 along the catenary under sway excitation at the top, with amplitude y_a=1.0m and circular frequency ω=2.0rad/s.

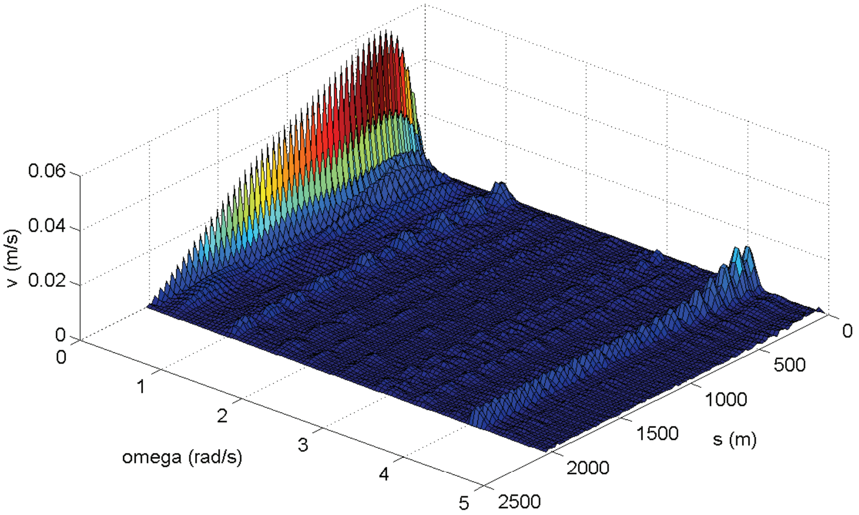


Fig. 3. Spectral densities of the normal velocity v along the catenary under sway excitation at the top, with amplitude y_a=1.0m and circular frequency ω=2.0rad/s.

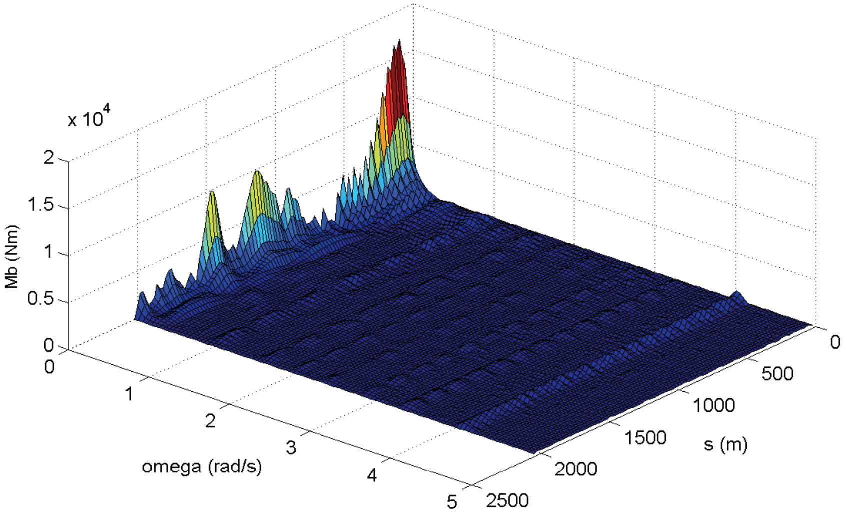


Fig. 4. Spectral densities of the in-plane dynamic bending moment M_{b1} along the catenary under sway excitation at the top, with amplitude y_a=1.0m and circular frequency ω=2.0rad/s.

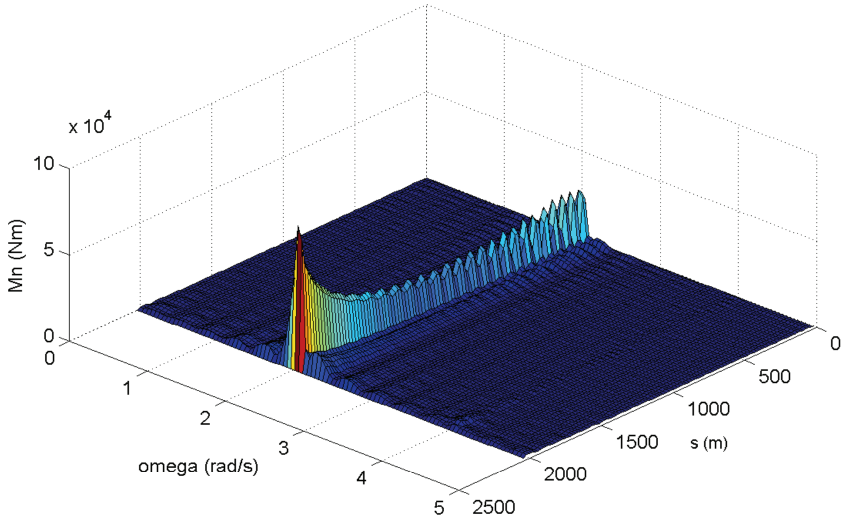


Fig. 5. Spectral densities of the out-of-plane dynamic bending moment M_{n1} along the catenary under sway excitation at the top, with amplitude y_a=1.0m and circular frequency ω=2.0rad/s.

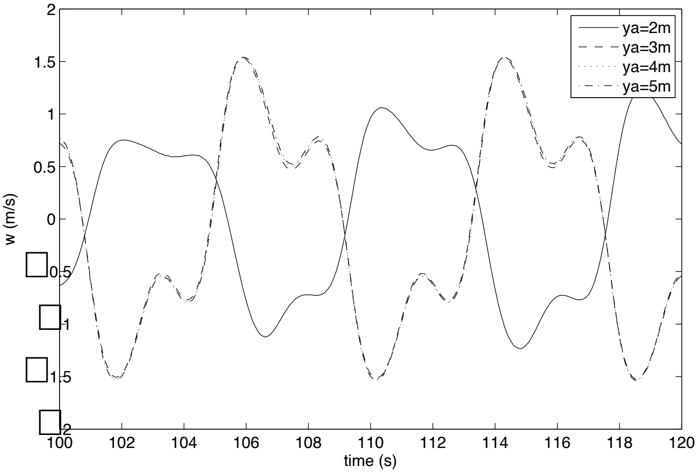


Fig. 6. Effect of the initial, short-time, sway displacement on the out-of-plane velocity *w* due to heave excitation with amplitude z_a=1.0m and circular frequency ω=1.5rad/s. The time history depicts the variation of *w* at the location of the max static in-plane bending moment M_{b0}, namely at *s*≈41m from touch down (at node *k*=3 in a discretization grid of 100 nodes)

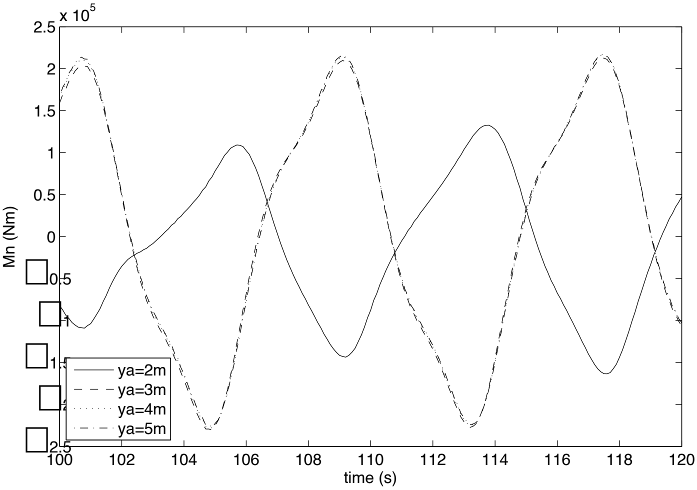


Fig. 7. Effect of the initial, short-time, sway displacement on the out-of-plane dynamic bending moment M_{n1} due to heave excitation with amplitude z_a=1.0m and circular frequency ω=1.5rad/s. The time history depicts the variation of M_{n1} at the location of the max static in-plane bending moment M_{b0}, namely at *s*≈41m from touch down (at node *k*=3 in a discretization grid of 100 nodes)

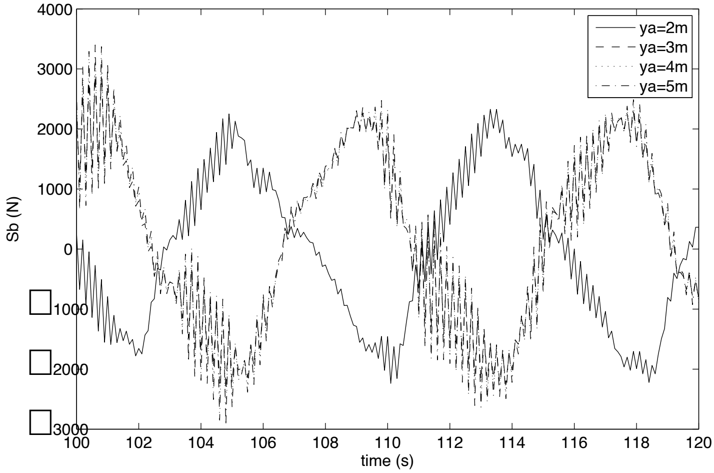


Fig. 8. Effect of the initial, short-time, sway displacement on the out-of-plane dynamic shear force S_{b1} due to heave excitation with amplitude z_a=1.0m and circular frequency ω=1.5rad/s. The time history depicts the variation of S_{b1} at the location of the max static in-plane bending moment M_{b0}, namely at s≈41m from touch down (at node k=3 in a discretization grid of 100 nodes)

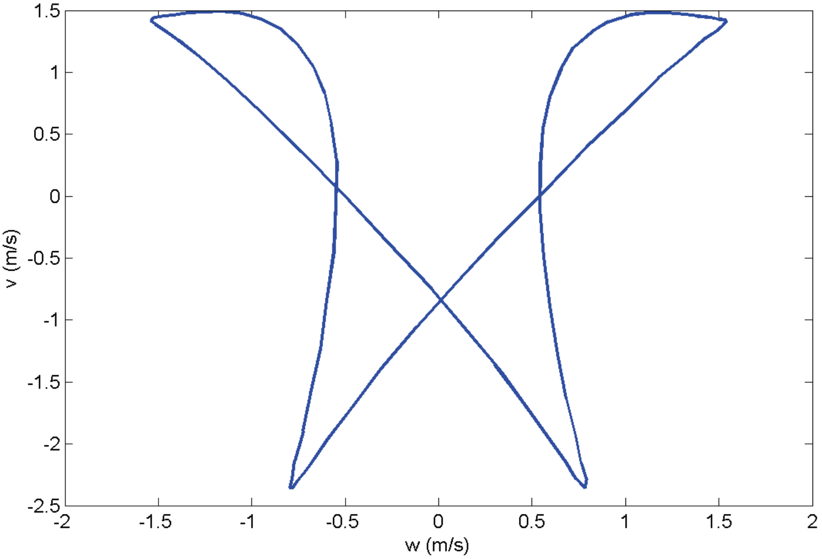


Fig. 9. Orbit of node no 3 (in a discretization grid of 100 nodes at s=41m from touch down) as seen from behind (v=f(w)), under heave excitation at the top with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

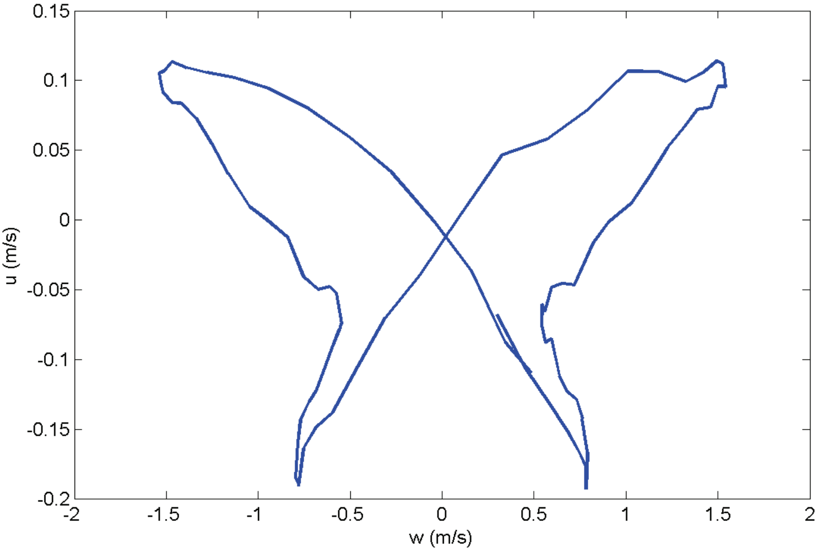


Fig. 10. Orbit of node no 3 (in a discretization grid of 100 nodes at s=41m from touch down) as seen from above (u=f(w)), under heave excitation at the top with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

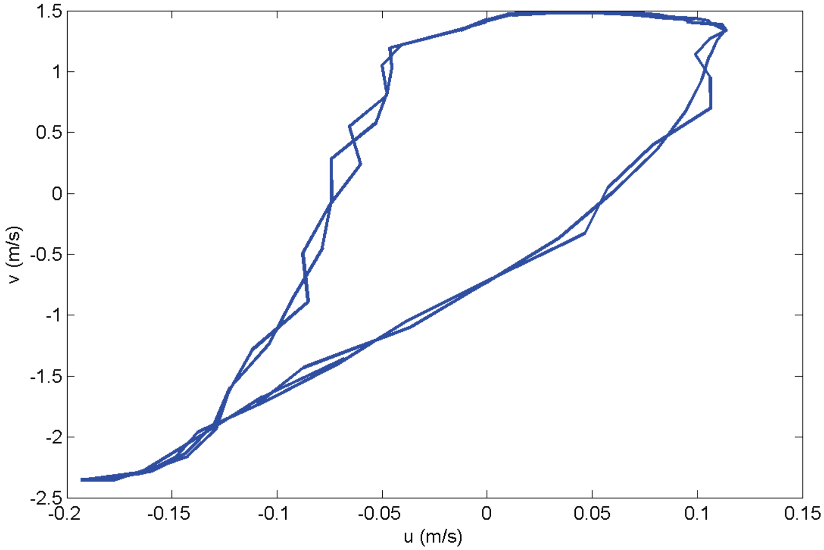


Fig. 11. Orbit of node no 3 (in a discretization grid of 100 nodes at s=41m from touch down) as seen from the side (v=f(u)), under heave excitation at the top with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

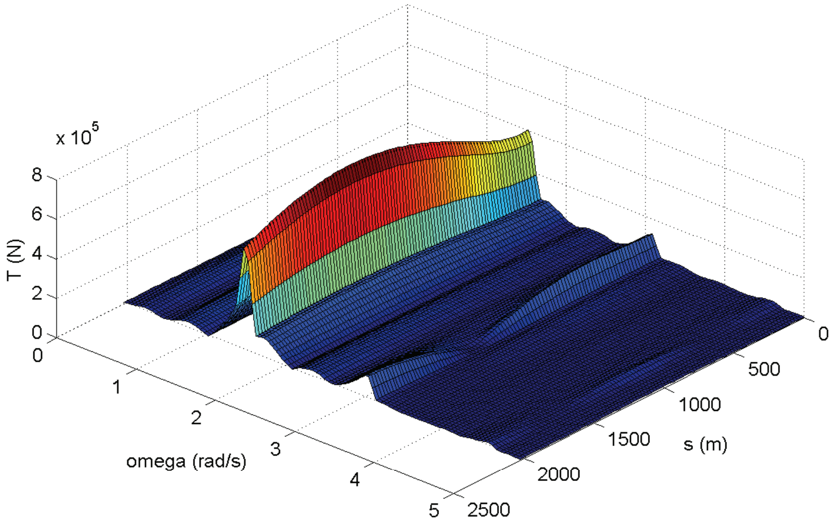


Fig. 12. Spectral densities of the dynamic tension T_1 along the catenary under heave excitation at the top, with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

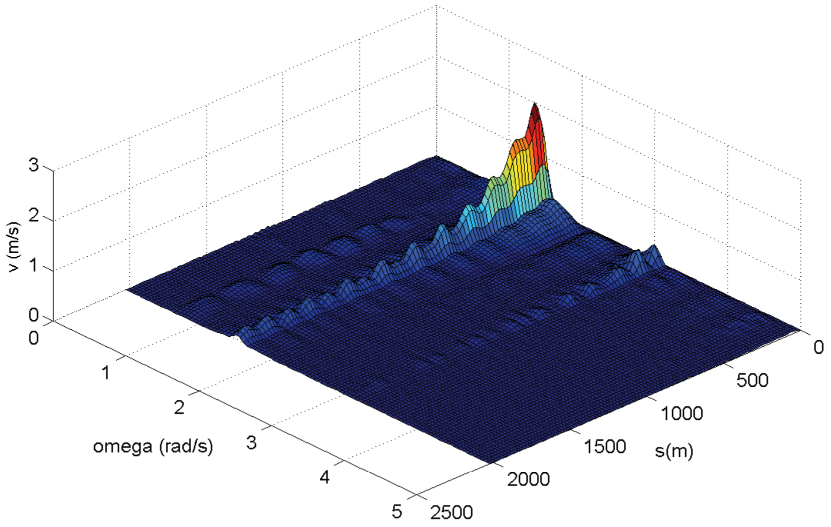


Fig. 13. Spectral densities of the normal velocity v along the catenary under heave excitation at the top, with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

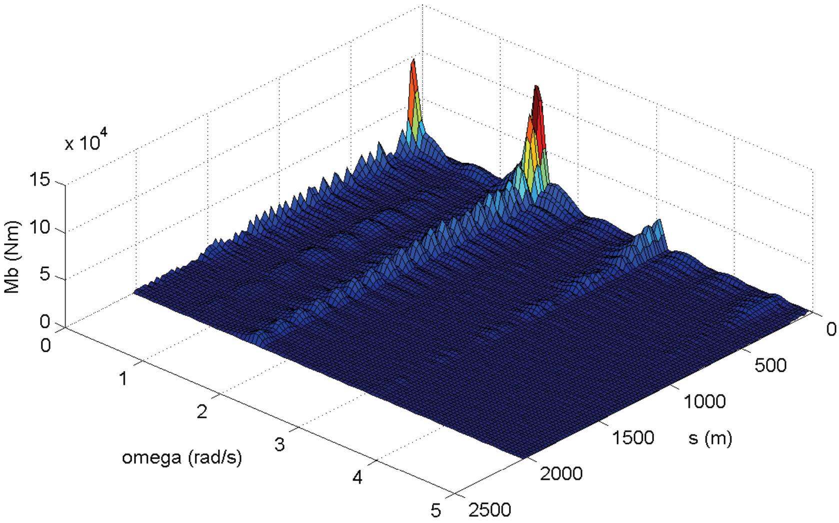


Fig. 14. Spectral densities of the in-plane dynamic bending moment M_{b1} along the catenary under heave excitation at the top, with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

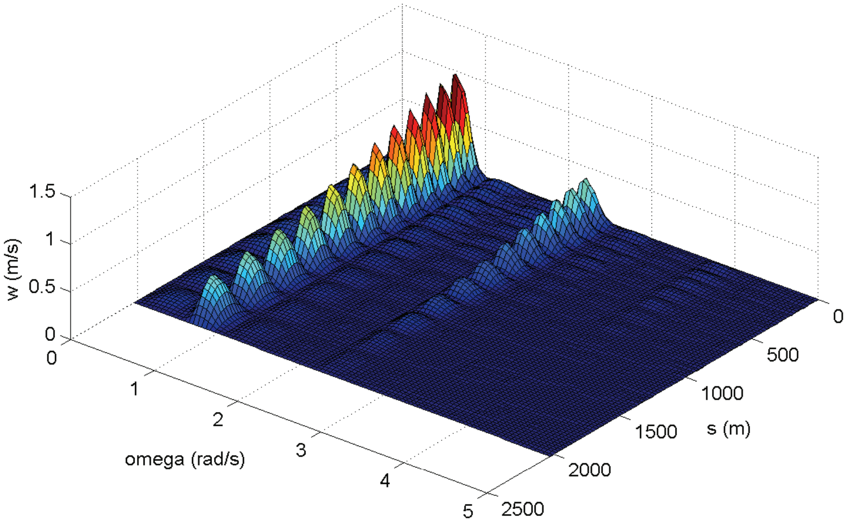


Fig. 15. Spectral densities of the bi-normal velocity w along the catenary under heave excitation at the top, with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

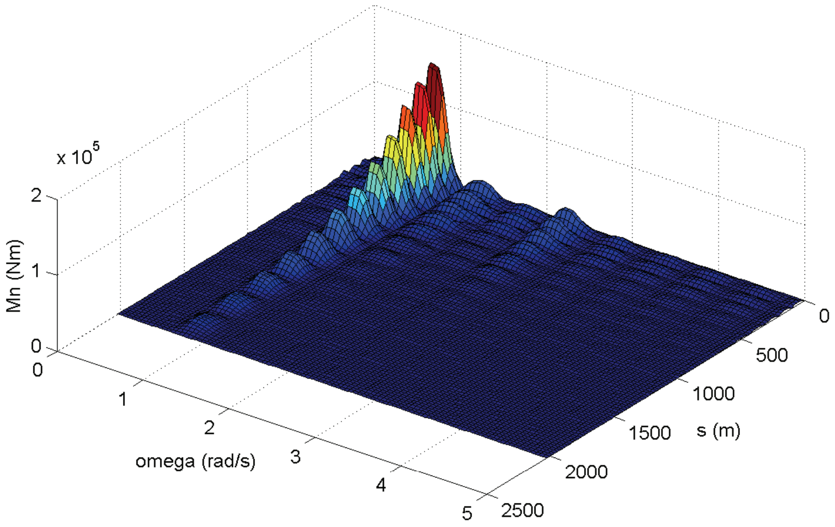


Fig. 16. Spectral densities of the out-of-plane dynamic bending moment M_{n1} along the catenary under heave excitation at the top, with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

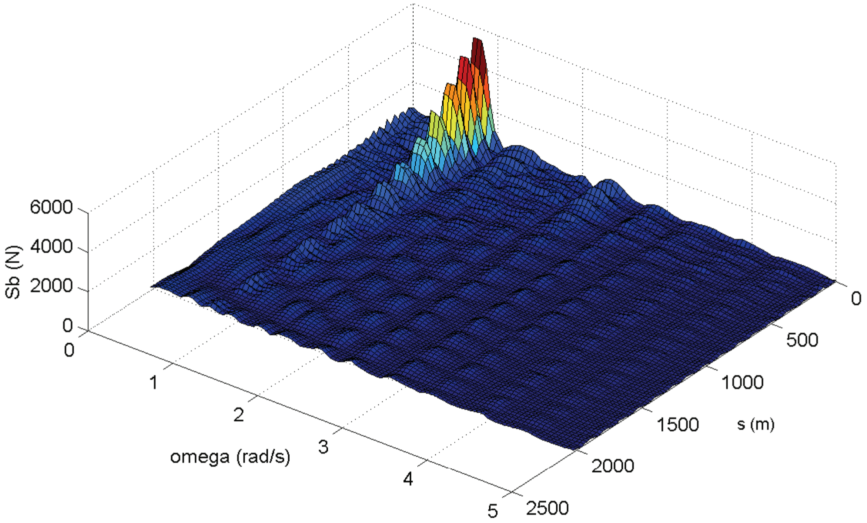


Fig. 17. Spectral densities of the out-of-plane dynamic shear force S_{b1} along the catenary under heave excitation at the top, with amplitude z_a=1.0m and circular frequency ω=1.5rad/s.

9

Nonlinear Dynamics Traction Battery Modeling

Antoni Szumanowski
Warsaw University of Technology,
Poland

1. Introduction

This chapter presents a method of determining electromotive force (EMF) and battery internal resistance as time functions, which are depicted as functions of state of charge (SOC). The model is based on battery discharge and charge characteristics under different constant currents that are tested by a laboratory experiment.

Further the method of determining the battery SOC according to the battery modeling result is considered. The influence of temperature on battery performance is analyzed according to laboratory-tested data and the theoretical background for calculating the SOC is obtained. The algorithm of battery SOC indication is depicted in detail. The algorithm of battery SOC "online" indication considering the influence of temperature can be easily used in practice by microprocessor. NiMH and Li-ion battery are taken under analyze. In fact, the method also can be used for different types of contemporary batteries, if the required test data are available.

Hybrid electric (HEVs) and electric (EVs) vehicles are remarkable solutions for the world wide environmental and energy problem caused by automobiles. The research and development of various technologies in HEVs is being actively conducted [1]-[8]. The role of battery as power source in HEVs is significant. Dynamic nonlinear modeling and simulations are the only tools for the optimal adjustment of battery parameters according to analyzed driving cycles. The battery's capacity, voltage and mass should be minimized, considering its over-load currents. This is the way to obtain the minimum cost of battery according to the demands of its performance, robustness, and operating time.

The process of battery adjustment and its management is crucial during hybrid and electric drives design. The generic model of electrochemical accumulator, which can be used in every type of battery, is carried out. This model is based on physical and mathematical modeling of the fundamental electrical impacts during energy conservation by a battery. The model is oriented to the calculation of the parameters EMF and internal resistance. It is easy to find direct relations between SOC and these two parameters. If the EMF is defined and the function versus the SOC ($k \in <0,1>$) is known, it is simple to depict the discharge/charge state of a battery.

The model is really nonlinear because the correlative parameters of equations are functions of time [or functions of SOC because $SOC = f(t)$] during battery operation. The modeling method presented in this chapter must use the laboratory data (for instance voltage for different constant currents or internal resistance versus the battery SOC) that are expressed in a static form. These data have to be obtained discharging and charging tests. The considered generic model is easily adapted to different types of battery data and is expressed in a dynamic way using approximation and iteration methods.

An HEV operation puts unique demands on battery when it operates as the auxiliary power source. To optimize its operating life, the battery must spend minimal time in overcharge and or overdischarge. The battery must be capable of furnishing or absorbing large currents almost instantaneously while operating from a partial-state-of-charge baseline of roughly 50% [9]. For this reason, knowledge about battery internal loss (efficiency) is significant, which influences the battery SOC.

There are many studies dedicated to determine the battery SOC [10]-[22]; however, these solutions have some limitations for practical application [23]. Some solutions for practical application are based on a loaded terminal voltage [17]-[20] or a simple calculation the flow of charge to/from a battery [21]-[22], which is the integral that is based on current and time. Both solutions are not considered the strong nonlinear behavior of a battery. It is possible to determine transitory value of the SOC "online" in real drive conditions with proper accuracy, considering the nonlinear characteristic of a battery by resolving the mathematical model that is presented in this paper.

This is the background for optimal battery parameters as well as the proper battery management system (BMS) design - particularly in the case of SOC indication [25]. The high power (HP) NiMH and LiIon batteries so common used in HEV were considered.

Finally, for instance, the plots of battery voltage, current and SOC as alterations in time for real experimental hybrid drive equipped with BMS especially design according to presented original battery modeling method, are attached.

2. Battery dynamic modeling

2.1 Battery physical model

The basis enabling the formulation of the energy model of an electrochemical battery is battery physical model shown in Fig.1.

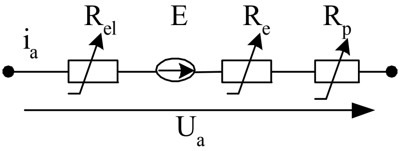


Fig. 1. Substitute circuit for nonlinear battery modeling

2.2 Mathematical modeling

The internal resistance can be expressed in an analytical way [7], where:

$$R_w(i_a,\tau,Q) = R_{el}(\tau,Q) + R_e(Q) + bE(i_a,\tau,Q)I_a^{-1} \tag{1}$$

$bE(i_a,\tau,Q)I_a^{-1}$ is the resistance of polarization.

b is the coefficient that expresses the relative change of the polarization's EMF on the cell's terminals during the flow of the I_a current in relation to the EMF E for nominal capacity. Electrolyte resistance R_{el} and electrode resistance R_e are inversely proportional to temporary capacity of the battery. During real operation, the capacity of the battery is changeable with respect to current and temperature [7], i.e.,

$$Q_u\left(i_a,t,\tau\right)=Q_\tau\left(\tau\right)-K_w\left(i_a\left(t\right),t\right) \tag{2}$$

or

$$Q_u\left(i_a,t,\tau\right)=Q_\tau\left(\tau,i_a\right)-\int_0^t i_a\left(t\right)\mathrm{d}t \tag{3}$$

Where:

$K_w\left(i_a\left(t\right),t\right)$ is the nonlinear function that is used to calculate the battery discharged capacity

$\int_0^t i_a\left(t\right)\mathrm{d}t$ is the function that is used to calculate the used charge, which has been drawn from the battery since the instant time t=0 till the time t

$Q_\tau\left(\tau,i_a\right)$ is the battery capacity as a function of temperature and load current, and

$$K_w=i_a^{n(\tau)}t \tag{4}$$

where K_w is the discharge capacity of the battery, n is the Peukert's constant, which varies for different types of batteries.

Assuming temperature influence:

$$Q_u\left(i,t,\tau\right)=c_\tau\left(\tau\right)Q_{\tau n}\left(\frac{i_a(t)}{I_n}\right)^{-\beta}-\int_0^t i_a\left(t\right)\mathrm{d}t \tag{5}$$

where the $c_\tau\left(\tau\right)$ coefficient can be defined as the temperature index of nominal capacity [7], i.e.,

$$c_\tau\left(\tau\right)=\frac{Q_\tau}{Q_{\tau n}}=\frac{1}{1+\alpha\left|\left(\tau_n-\tau\right)\right|} \tag{6}$$

where α is the temperature capacity index (we can assume α ≈ 0.01 deg^{-1}).

According to the Peukert equation, we can get the following:

$$\frac{Q(i_a)\overline{U}}{Q_{\tau n}\overline{U}}=\left(\frac{i_a(t)}{I_n}\right)^{-\beta(\tau)} \tag{7}$$

The left –hand side of the equation (7) is the quotient of the electric power that is drawn from the battery during the flow of $i_a \neq I_n$ current and the electric power that is drawn from the battery during loading with the rated current. The quotient mentioned above defines the usability index of the accumulated power, i.e.,

$$\eta_A\left(i_a,\tau\right)=\left(\frac{i_a(t)}{I_n}\right)^{-\beta(\tau)} \tag{8}$$

When $i_a < I_n$, the value of the index can exceed 1.
During further solution of (5), it can be transformed by means of (8), i.e.,

$$Q_u(i,t,\tau) = c_\tau(\tau)\eta_A(i_a,\tau)Q_{\tau n} - \int_0^t i_a(t)\mathrm{d}t \tag{9}$$

Therefore, the real battery SOC can be expressed in the following way [7]:

$$k = \frac{Q_u}{Q_{\tau n}} = \frac{c_\tau(\tau)\eta_A(i_a,\tau)Q_{\tau n} - \int_0^t i_a(t)\mathrm{d}t}{Q_{\tau n}} \tag{10}$$

where $k = 1$ for a nominally charged battery, $0 \le k \le 1$, and thus

$$k = c_\tau(\tau)\eta_A(i_a,\tau) - \frac{1}{Q_{\tau n}}\int_0^t i_a(t)\mathrm{d}t \tag{11}$$

For practical application, it's necessary to transform aforementioned equations for determining the internal resistance R_w and EMF as functions of k (SOC) [7], i.e.,

$$R_w(i_a,\tau,Q) = \frac{l_1}{Q_u(i_a,t,\tau)} + \frac{l_2}{Q_u(i_a,t,\tau)} + \frac{bE(i_a,\tau,Q)}{i_a(t)} \tag{12}$$

$$R_w(i_a,t,\tau) = lk^{-1} + b\frac{E(k)}{i_a(t)} \tag{13}$$

where $l = (l_1 + l_2)Q_{\tau n}^{-1}$, $l \approx const$ is a piecewise constant, assuming that the temporary change of the battery capacity is significantly smaller than its nominal capacity; the coefficient l is experimentally determined under static conditions. $E(k)$ is the temporary value of polarization's EMF, which is dependent on the SOC.
The EMF as a function of k is deduced from the well-know battery voltage equation, including the momentary value of voltage and internal resistance, because the values R_w and EMF are unknown. The solution can be obtained by a linearization and iterative method, which is explained by following Fig.2 and following:

$$b(k) = \frac{E(k) - E^*_{\min}}{E^*_{\max}} \tag{14}$$

Take under consideration (12)-(14), it's then possible to obtain the following:

$$\begin{cases} R_w(k_n) = \dfrac{E(k_n) - E^*{}_{\min}}{E^*{}_{\max}}\dfrac{E(k_n)}{I_n} + \dfrac{l(k_n)}{k_n} \\ R_w(k_{n-1}) = \dfrac{E(k_{n-1}) - E^*{}_{\min}}{E^*{}_{\max}}\dfrac{E(k_{n-1})}{I_n} + \dfrac{l(k_{n-1})}{k_{n-1}} \end{cases} \tag{15}$$

Obviously, $E(k)$ is the function that we need. To obtain it, it's necessary to use the known functions $u_a(k)$, which are obtained by laboratory tests.

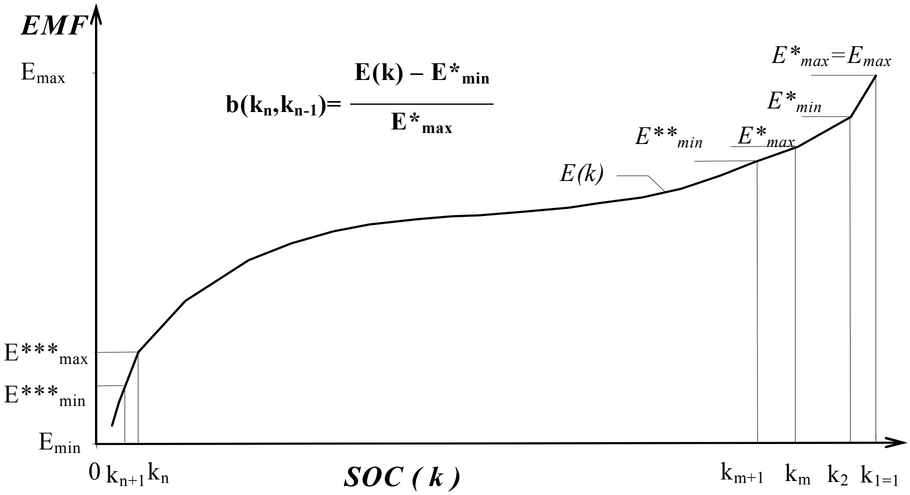


Fig. 2. Linearization method of EMF versus SOC (k)

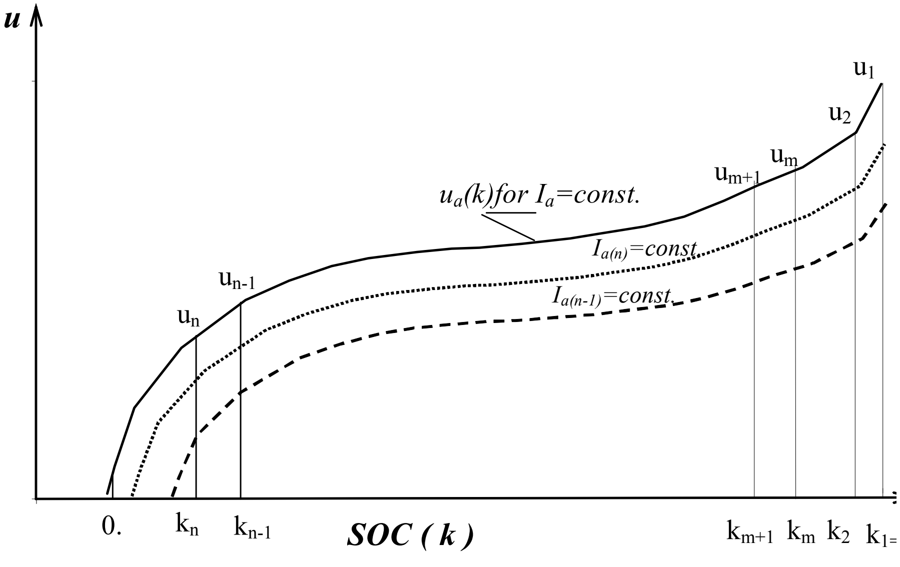


Fig. 3. Linearization method of voltage versus SOC (k)

Similarly as in the case of Fig.3, the following equations are generated:

$$\begin{cases} u(k_n) = E(k_n) \pm I_a R_w(k_n) \\ u(k_{n-1}) = E(k_{n-1}) \pm I_a R_w(k_{n-1}) \end{cases} \tag{16}$$

$u(k_n)$ and $u(k_{n-1})$ are known from the family of voltage characteristics that are obtained by laboratory tests. $I_{a(n)}$ is also known because $u(k_n)$ is determined for $I_{a(n)} = const.$

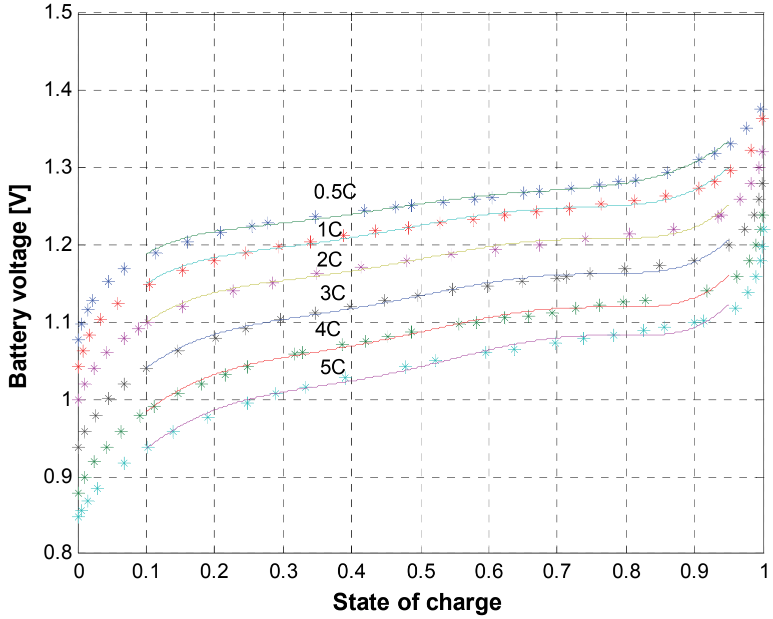


Fig. 4. Discharging data of a 14-Ah NiMH battery

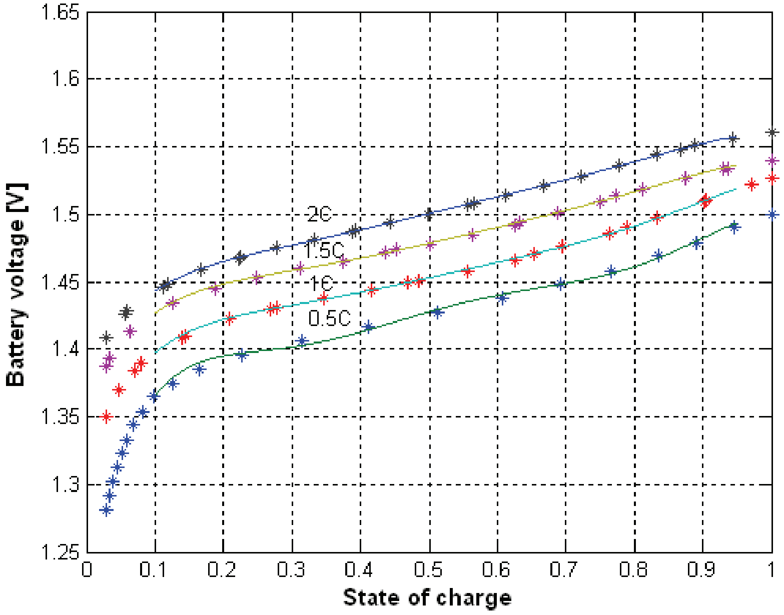


Fig. 5. Charging data of a 14-Ah NiMH battery

+ is for discharge
- is for charge
$k \in < 0,1 >$

Using the above-presented approach, based on experimental data (shown in Figs.4 and 5), it's possible to construct a proper equation set as in the shape of (15) and (16) and resolve it in an iterative way.

Last, the equations of R_w and EMF take the shape of the following polynomial:

$$\begin{aligned} R_w(k) &= A_r k^6 + B_r k^5 + C_r k^4 + D_r k^3 + E_r k^2 + F_r k + G_r \\ E(k) &= A_e k^6 + B_e k^5 + C_e k^4 + D_e k^3 + E_e k^2 + F_e k + G_e \\ b(k) &= A_b k^6 + B_b k^5 + C_b k^4 + D_b k^3 + E_b k^2 + F_b k + G_b \\ l(k) &= A_l k^7 + B_l k^6 + C_l k^5 + D_l k^4 + E_l k^3 + F_l k^2 + G_l k + H_l \end{aligned} \tag{17}$$

3. Battery modeling results

The basic elements that are used to formulate the mathematical model of a NiMH battery are the described iteration-approximation method and the approximations based on the battery discharging and charging characteristics that are obtained by an experiment. Experimental data are approximated to enable determination of the internal resistance in a small-enough range k=0.001. The modeling results (Figs. 6-8) in the battery SOC operating range of 0.1-0.95 show a small deviation (less than 1%) from the experimental data (Figs.9 and 10). The NiMH battery that is used in the experiment and the modeling is an HP battery for HEV application. The nominal voltage of the battery is 1.2V, and the rated capacity 14Ah.

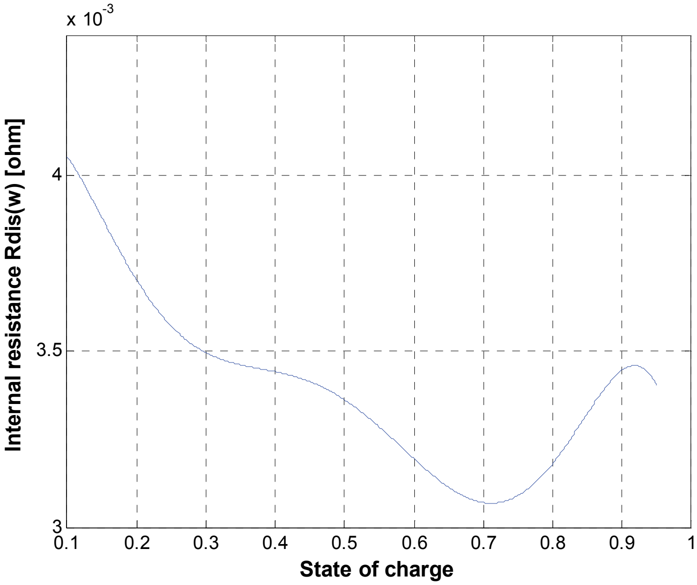


Fig. 6. Computed internal resistance characteristics of a 14-Ah NiMH battery for discharging

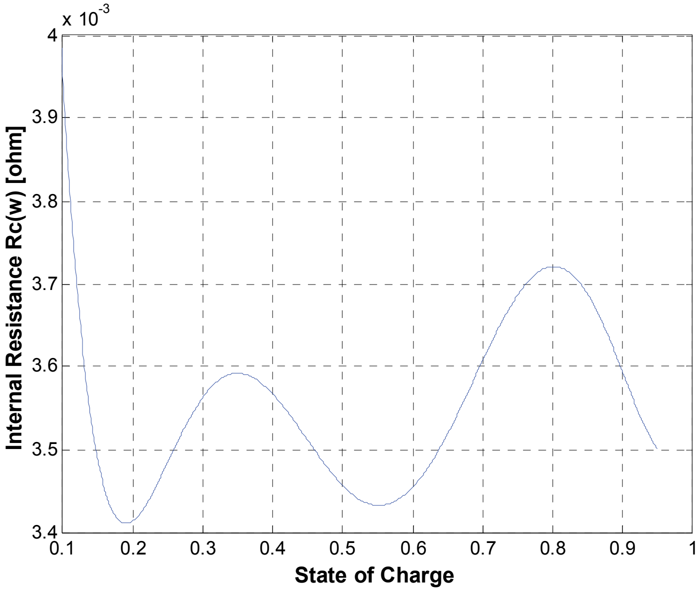


Fig. 7. Computed internal resistance characteristics of a 14-Ah NiMH battery for charging

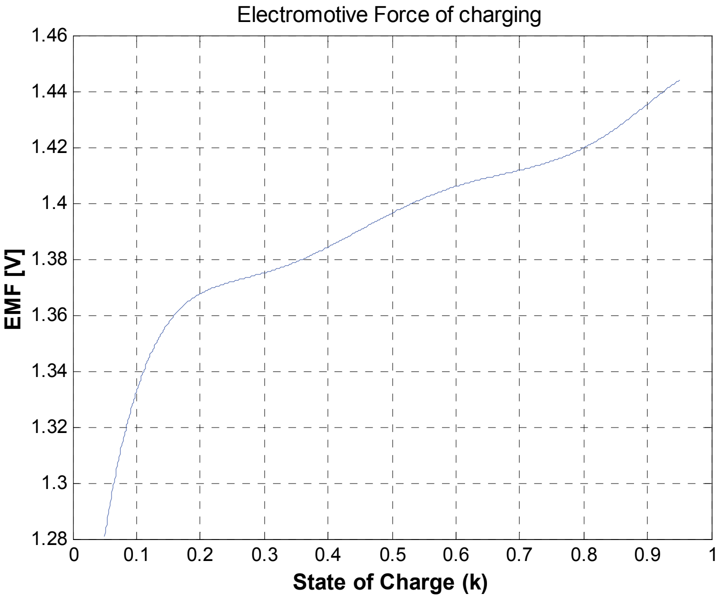


Fig. 8. Computed EMF of a 14-Ah NiMH battery

After approximation according to the computed results, approximated equations of (17) for 14-Ah NiMH battery can be obtained. These factors of equations (17) are available in Table 1.

Factors of Equation (17)	Internal resistance R(w) during discharging	Internal resistance Rd(w) during charging	Electromotive Force	Coefficient b	Coefficient l
				Discharging	Discharging
				Charging	Charging
A	0.65917	0.42073	13.504	-0.015363	0.65917
				0.015341	0.42073
B	-2.0397	-1.4434	-36.406	0.10447	-2.0528
				-0.10661	-1.4376
C	2.4684	1.9362	36.881	-0.18433	2.4978
				0.22702	1.9195
D	-1.4711	-1.2841	-17.198	0.13578	-1.495
				-0.21788	-1.2661
E	0.44578	0.43809	3.5264	-0.045129	0.45416
				0.10346	0.42896
F	-0.065274	-0.071757	-0.10793	0.0059814	-0.066422
				-0.023367	-0.06961
G	0.0099109	0.0078518	1.234	-9.416e-005	0.0099289
				0.0020389	0.0076585
H					-1.2154e-015
					1.9984e-008

Table 1. Factors of Eq. (17) for 14-Ah NiMH battery

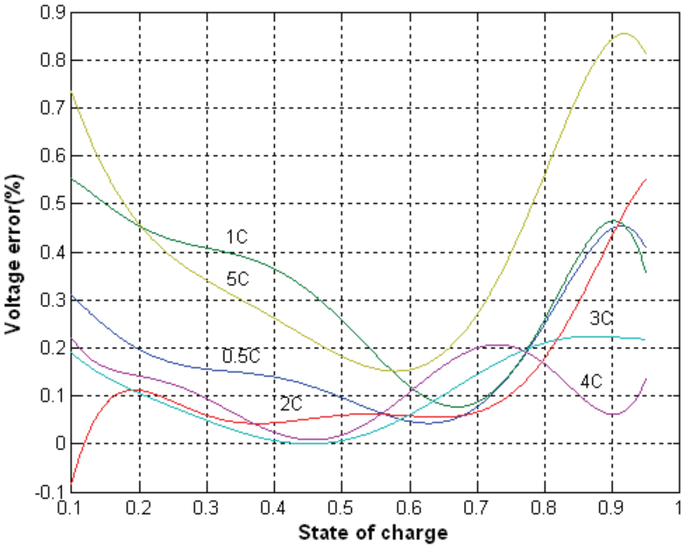


Fig. 9. Error of experiment data and the computed voltage at different discharge currents

The basic element used to formulate the mathematical model of Li-ion battery module from SAFT Company is the earlier described iteration-approximation method and the approximated based on the battery discharging characteristics obtained by experiment. The experimental data is approximated to enable determining the internal resistance in an enough small range k = 0.001. The analyses, in the operating range SOC between 0.01~0.95, gives us a small deviation (less than 2%) by using the iteration-approximation method from the experimental data. The VL30P-12S module has 30Ah rated capacity and it's special designed for HEV application.

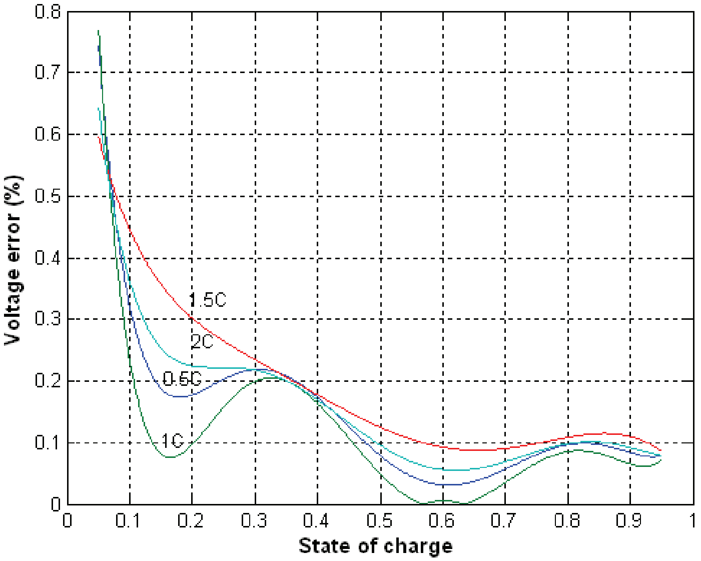


Fig. 10. Error of experiment data and computed voltage at different charge currents

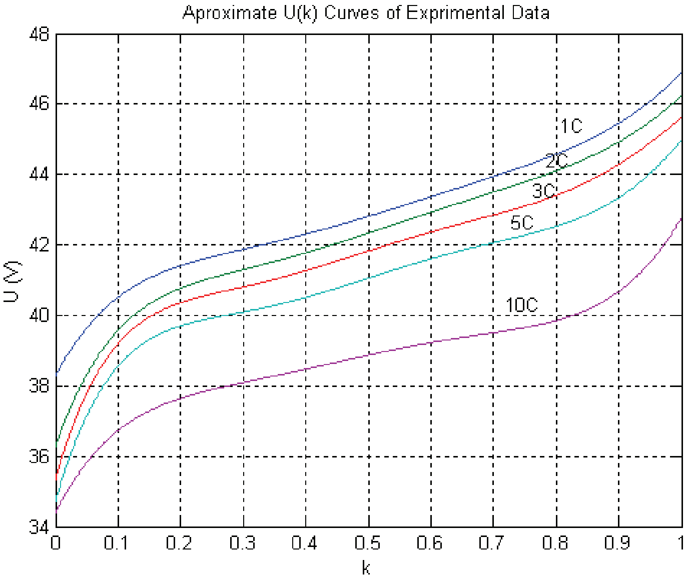


Fig. 11. The discharging voltage characteristics of SAFT 30Ah Li-ion module

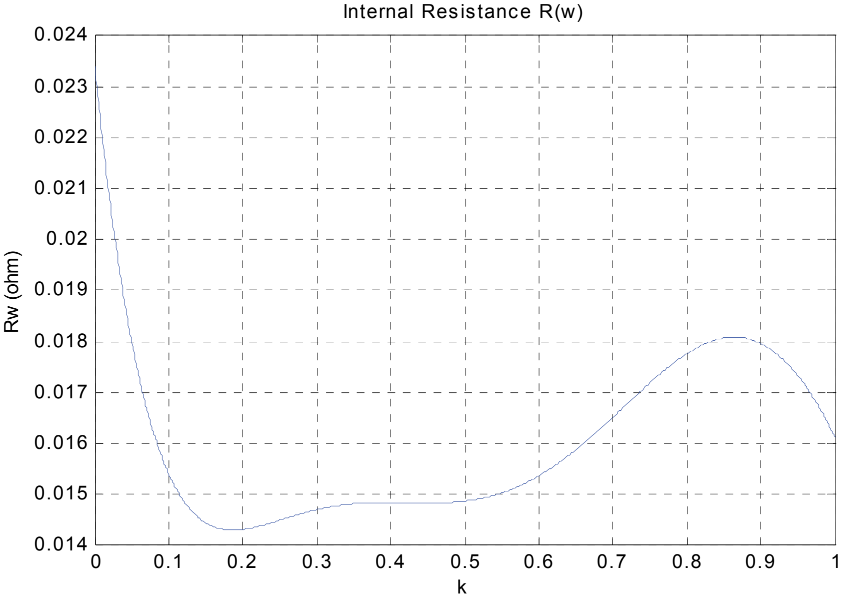


Fig. 12. The computed internal resistance of SAFT 30Ah Li-ion module

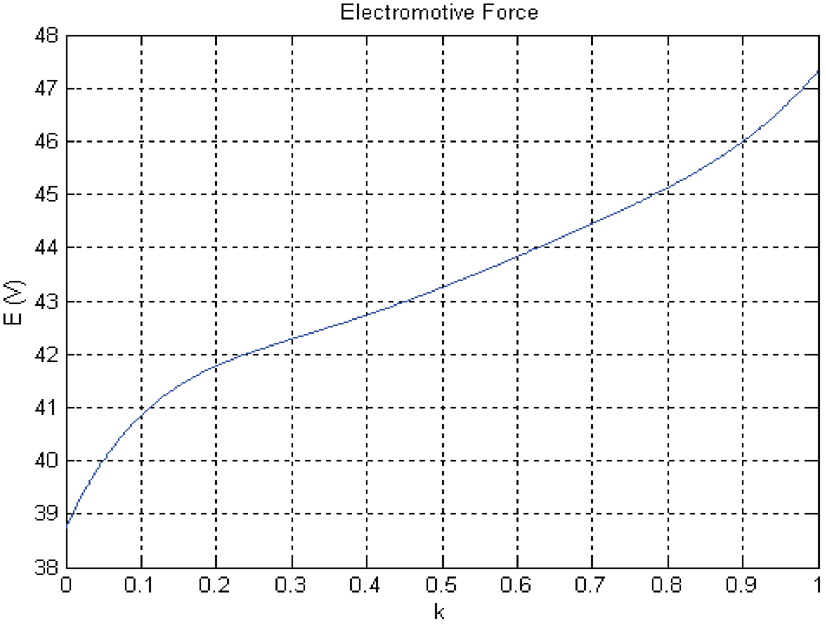


Fig. 13. The computed EMF of SAFT 30Ah Li-ion module

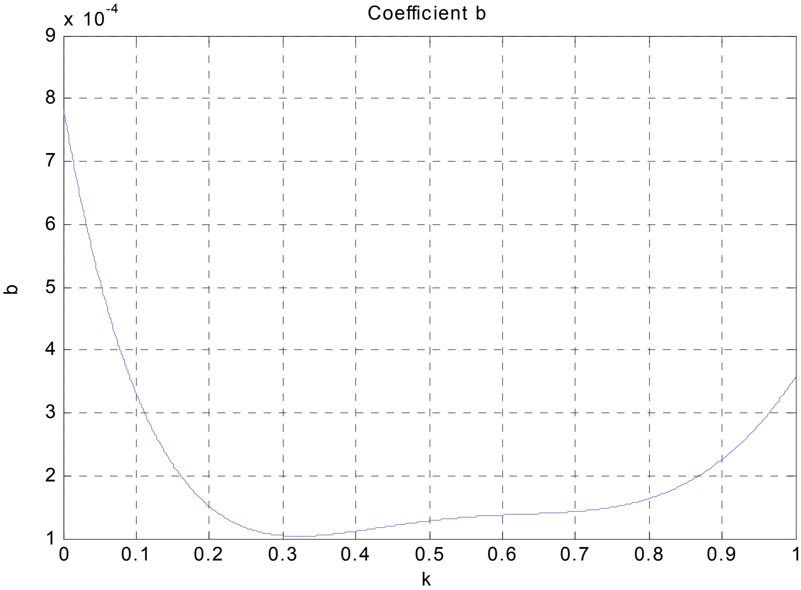


Fig. 14. The computed coefficient b of SAFT 30Ah Li-ion module

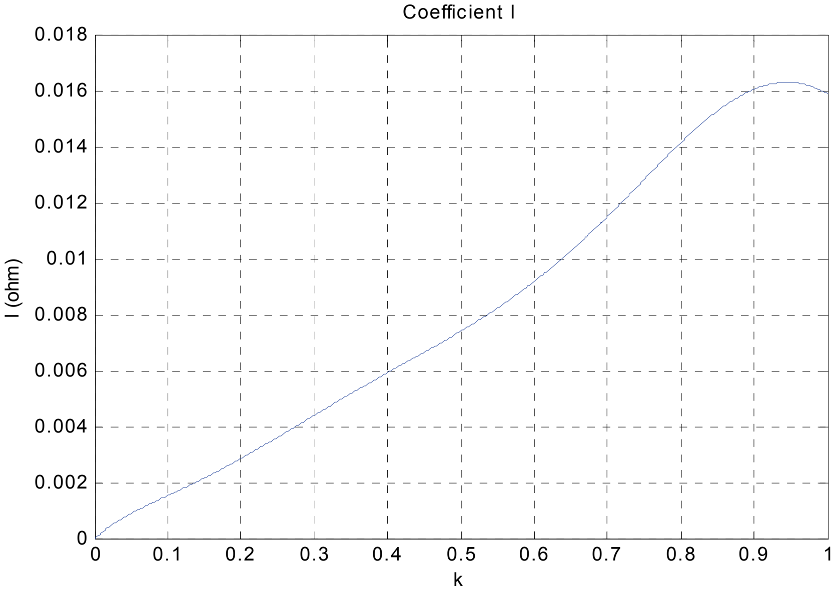


Fig. 15. The computed coefficient l of SAFT 30Ah Li-ion module

After approximation according to the computed results, approximated equations of (17) for 30-Ah Li-ion module can be obtained. These factors of equations (17) are available in Table 1.

Factors of equations (6.56)	Internal resistance R_w	Electromotive force E	Coefficient b	Coefficient l
A	0.71806	-28.091	0.0032193	0.71806
B	-2.6569	157.05	-0.016116	-2.6545
C	3.7472	-296.92	0.036184	3.736
D	-2.5575	265.34	-0.040738	-2.5406
E	0.8889	-119.29	0.023539	0.87755
F	-0.14693	30.476	-0.0065159	-0.14352
G	0.023413	38.757	0.00078501	0.022978
H				-1.7916e-015

Table 2. Factors of Eq. (17) for 30-Ah Li-ion module

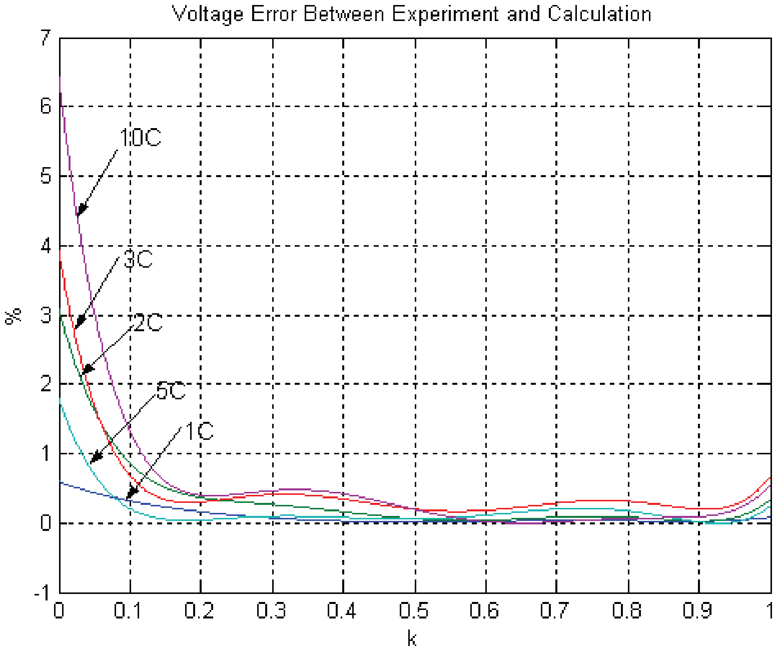


Fig. 16. Errors between testing data and computed result of SAFT 30Ah Li-ion module

4. Temperature influence analysis on battery performance

The determination of the battery EMF and internal resistance gives unlimited possibilities of calculating the battery's voltage versus SOC (k) relation for a different value of discharge-charge current. For a real driving condition, the battery discharge or charge depends on the drive architecture influencing the respective power distribution. In majority, battery charging takes place during vehicle regenerative braking, which means that this situation lasts for a relatively short time with a significant peak-current value. A discharging current that is too high results in a rapid increase in the battery temperature.

The main role of this study is to find a theoretical background for calculating the temperature influence on the battery SOC. The presented method is more accurate and complicated compared with other methods, which doesn't mean that it is more difficult to apply. First of all, it is necessary to make the following assumptions:

The considered battery is fully charged in nominal conditions: nominal current, nominal temperature and nominal capacity (i_b =1C, τ_b =20°C, the capacity is designed for nominal parameters, respectively).

The EMF for the considered battery is defined as its nominal condition in the nominal SOC alteration range $k \in <1,0>$. The assumption is taken that the EMF value of k=0.15 is the minimum EMF. For k=0, the EMF is defined as the "minimum-minimorum", in practice which should not be obtained. The same assumption is recommended for a value that is different from the nominal temperature for the k_τ (SOC) definition. As shown in Fig.19, the starting point value of the EMF for a different value from the nominal temperature can be higher or lower, which means that the extension alteration of the SOC could be longer or shorter. For instance (see Fig.17), in the case of the NiMH battery for a value that is higher than the nominal temperature, the discharge capacity is smaller than the nominal, which means that for a certain temperature, the battery capacity corresponding to this temperature is also changed in file $k_\tau \in <1,0>$. However, the full k_τ doesn't mean the same discharge capacity as in the case of nominal temperature but does mean the maximum discharge capacity at this temperature. For this reason, in fact, k_τ for this temperature is only $k(t) > k_\tau(t)$, [in some case, $k(t) < k_\tau(t)$, where $k(t)$ is connected only with nominal conditions].

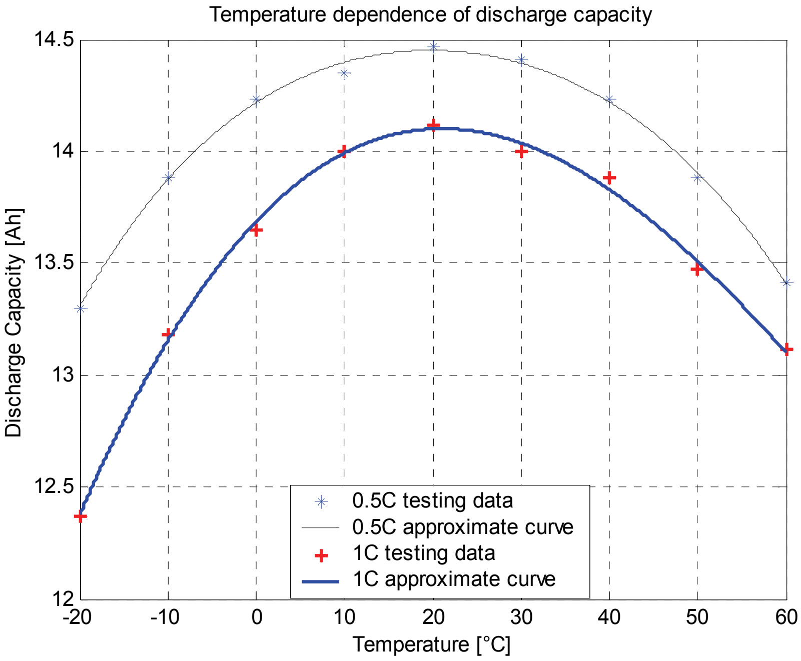


Fig. 17. Temperature dependence of the discharge capacity of the NiMH battery

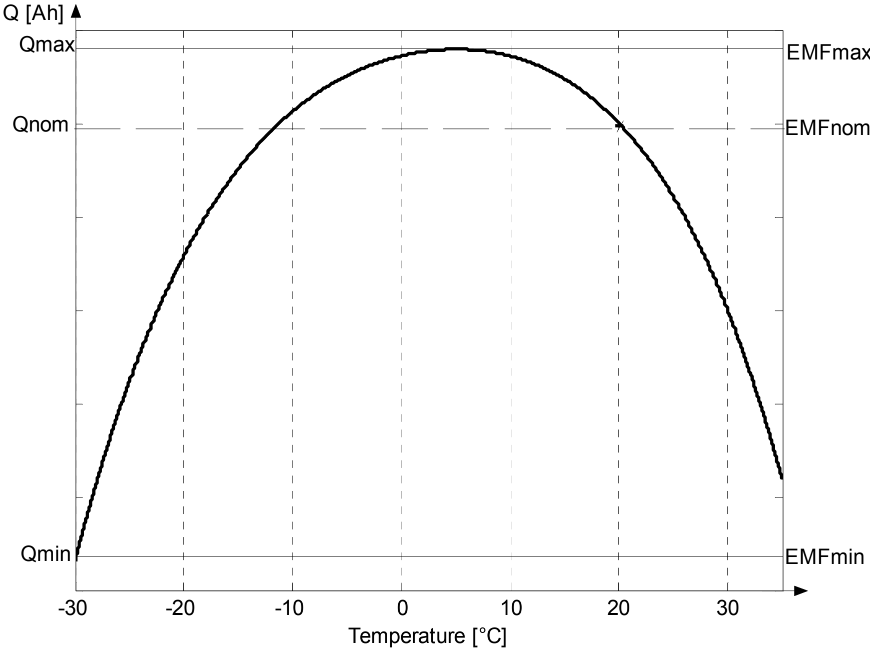


Fig. 18. Temperature dependence of the battery usable discharge capacity and the EMF starting point

From Fig.18, it is easy to note that the EMF (in the case of this battery type) value in the nominal conditions is smaller than the EMF value for a temperature that is lower than 20°C (the nominal temperature), which means that for a maximum EMF value, the available battery capacity is higher than in the case of the nominal terms. For nominal conditions, the SOC can be defined by a k factor ($k \in <1,0>$). If the EMF for the non-nominal conditions reaches its highest value, the available charge (in ampere-hours) will be also greater. It is easy to note the relation $Q_\tau = Q_{\max}$ and Q_{nom} is defined as follows:

$\frac{Q_\tau = Q_{\max}}{Q_{nom}} > 1$ [If $Q_\tau < Q_{nom} \rightarrow \frac{Q_\tau}{Q_{nom}} < 1$, correspondingly, $EMF_\tau < EMF_{nom} \rightarrow \frac{EMF_\tau}{EMF_{\max}} < 1$]

This corresponds to:

$\frac{EMF_\tau = EMF_{\max}}{EMF_{nom}} > 1$. On the other hand, for Q_{nom} , $k \in <1,0>$, but relating it to $Q_\tau > Q_{nom}$ in τ condition, the file $<1,0>$ means file $<0,Q_{\max}>$. Transforming k in nominal terms to k_τ is necessary to use the general relation $\frac{Q_\tau}{Q_{nom}}$. Theoretically, the product $k_{nom}\frac{Q_\tau}{Q_{nom}}$ transfers the SOC factor into other than nominal temperature conditions. The same transformation can be obtained for $k_{nom}\frac{EMF_\tau}{EMF_{nom}}$, where $k_{nom} \in <1,0>$.

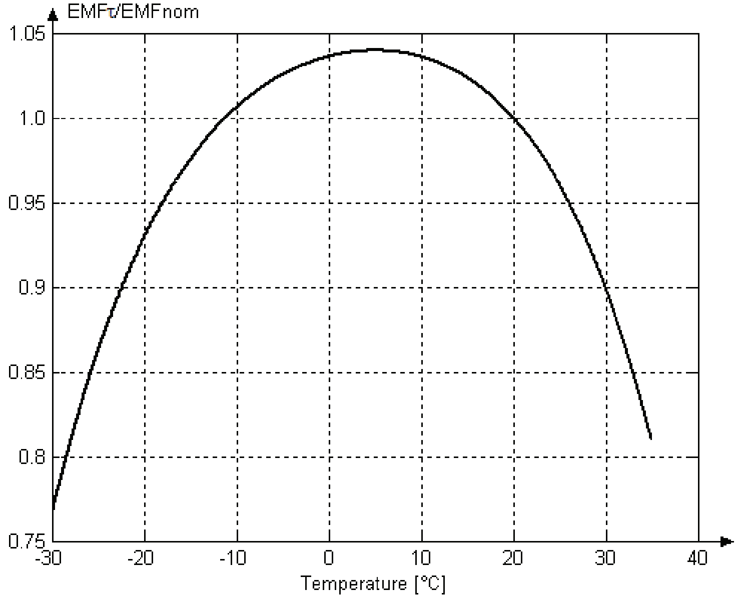


Fig. 19. Relation of $\frac{EMF_{\tau}}{EMF_{nom}}$ and temperature

Using the transformation factor $k_{nom}\frac{EMF_{\tau}}{EMF_{nom}}$ or $k * s_{\tau}$ ($k_{nom} = k, s_{\tau} = \frac{EMF_{\tau}}{EMF_{nom}}$),it is possible to relate the SOC of the battery that is determined for the nominal temperature to other different temperatures.

5. Algorithm of battery SOC indication

The algorithm is given as follows.

1. By simulation, the family of $u_b(k)$ for different constant currents $i_b \in < 0.5C, 6C >$ and nominal temperature (e.g. 20°C) can be obtained according to battery modeling results (EMF and internal resistance as functions of SOC).
2. From Fig.19, $s_{\tau} = \frac{EMF_{\tau}}{EMF_{nom}}$ is defined for $\tau \in$ <-30°C, +35°C>
3. From Fig.20, for k=0.9, …0.2, the following lookup table can be obtained

$$k = 0.9 \Rightarrow \begin{bmatrix} u_{11}, i_{11}, E_1 \\ u_{12}, i_{12}, E_1 \\ \cdots \\ u_{1n}, i_{1n}, E_1 \end{bmatrix} \cdots\cdots k = 0.2 \Rightarrow \begin{bmatrix} u_{81}, i_{81}, E_8 \\ u_{82}, i_{82}, E_8 \\ \cdots \\ u_{8n}, i_{8n}, E_8 \end{bmatrix}$$

Because of the practical limitation of the SOC alteration of the battery that is applied in hybrid drives, the range of k changes can be expressed as $< 0.9, 0.2 >$ for the nominal temperature.

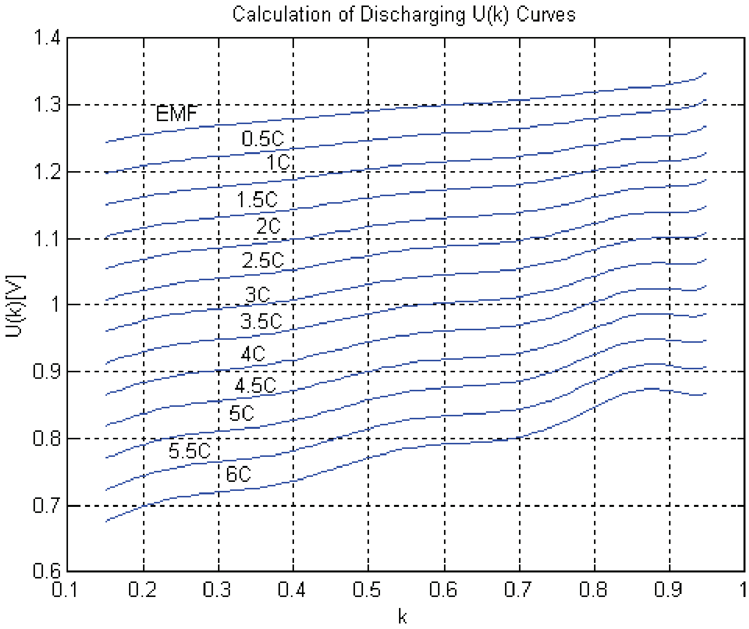


Fig. 20. EMF and calculated discharging voltage characteristics at different discharging current and nominal temperature

4. Considering the real temperatures, the SOC of the battery in relation to the nominal temperature can be defined as $s_{\tau}*k=k_{\tau}$ For instance, at a temperature of +5°C, $\frac{EMF_{\tau}}{EMF_{nom}}=1.06$; hence, $k_{+5°C}$=1.06k,which means that at this moment and this temperature, the available capacity is 1.06 times that of the nominal temperature. At a temperature +30°C, $\frac{EMF_{\tau}}{EMF_{nom}}=0.89$; hence, $k_{+30°C}$ =0.89k, which means that at this moment and this temperature, the available capacity is 0.89 times that of the nominal temperature.

A similar method and process can be used for the battery charging process (see Fig.21)

The above-depicted method can be used in design of battery management system (BMS) for the SOC determination and indication, especially in hybrid (HEV) and electric (EV) vehicle drives. Based on the aforementioned steps 1) - 4), the SOC indication algorithm can be depicted as is shown in Fig.22.

In HEV the battery SOC changes faster (because HP high power battery is used) but not so deep as in pure electric vehicles, equipped with high energy (HE) battery. It means that the SOC indication - display process may not be realized as frequently. It's not necessary to display the SOC of the battery every second. Certainly, the previous value of the SOC has to be remembered by a microprocessor.

High accuracy of determination of battery SOC is at first of all necessary for entire drive system control. In opposed to indication – display, the feedback SOC signals from battery must be available online.

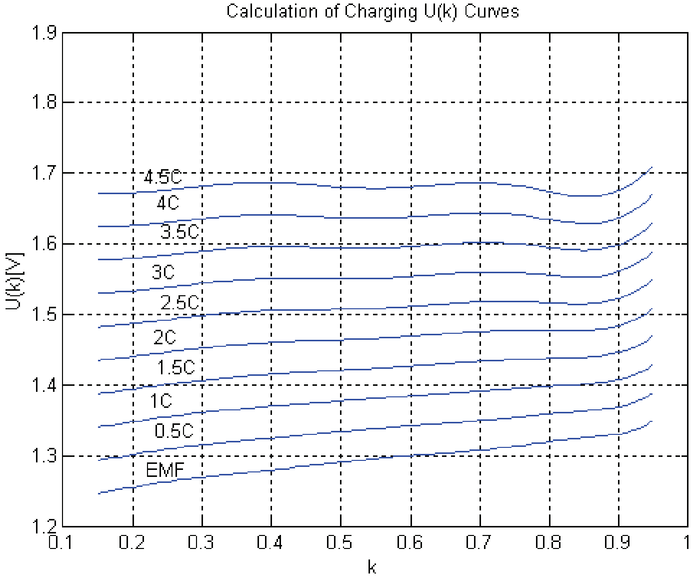
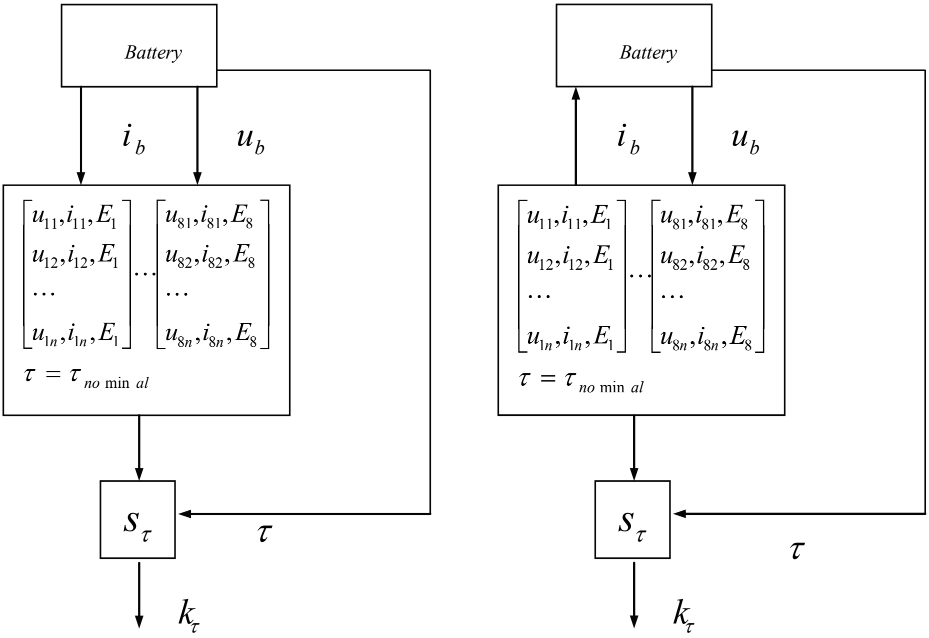


Fig. 21. EMF and calculated charging voltage characteristics at different charging current and nominal temperature



a) Discharging b) Charging

Fig. 22. SOC indication algorithm

The presented original method of EMF (as function of k - SOC) calculation is the background for constructing BMS. This procedure is easily adopted for control application in HEV and EV. Its high accuracy is very important for control drive systems (master controller) based on feedback signals from BMS.

The following equation is the background to determine accurate value of SOC (k) for dynamic conditions.

$$\begin{aligned} u(t) &= E(k) \pm R_w(k)i(t) \\ k &= k_{nom}s_\tau \end{aligned} \tag{18}$$

+ is for discharge; - is for charge; where E(k) and R_w(k) are taken from equation (17) for real battery module.

Based on equation (18), the SOC calculation can be obtained in a direct way in "online" dynamic battery voltage and current alterations. The solving (17) as high power factor polynomial is really possible "online" by using two procedures: look-up table (dividing polynomial function in shaped-line ranges) or "bisection" numerical iterative computation. In some cases, when the accuracy of SOC indication can be lower (about 5%) , which is accepted in HEV and EV drives, power factor of polynomial can be decreased by additional approximation E(k) and R_w(k). The accuracy of real time calculation is about 100 µs.

The second method is "bisection" iterative calculation.

The exemplary plots of battery voltage, current and SOC is shown in following figures 23, 24, 25. Because The SOC of battery is much slower changeable than its voltage and current, the SOC indication is computed and indicated by using "moving average" procedure.

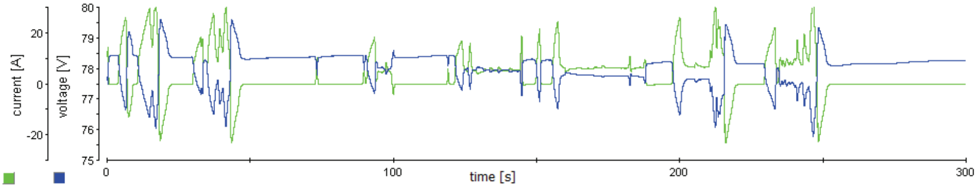


Fig. 23. Exemplary test of battery load in hybrid drive ; blue– battery current, green– battery voltage

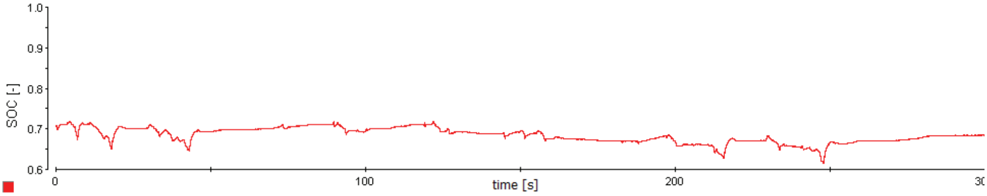
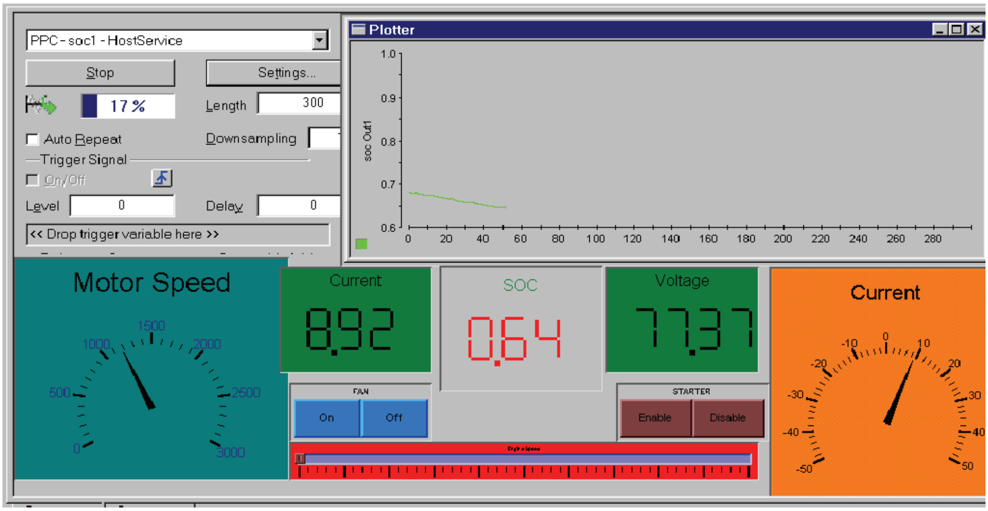


Fig. 24. Exemplary test of battery SOC indication in real drive conditions corresponding to battery load shown in Fig.23.



PPC - soc1 - HostService

□ Auto Repeat

Length 300

Level 0

Delay 0

17 %

Current 8.92

SOC 0.64

Voltage 77.37

Fig. 25. Screen of control system based on d'Space programming for SOC indication.

6. Conclusions

The assumed method and effective model are very accurate according to error checking results of the NiMH and Li-Ion batteries. The modeling method is valid for different types of batteries. The model can be conveniently used for vehicle simulation because the battery model is accurately approximated by mathematical equations. The model provides the methodology for designing a battery management system and calculating the SOC. The influence of temperature on battery performance is analyzed according to laboratory-tested data and the theoretical background for the SOC calculation is obtained. The algorithm of the battery SOC "online" indication considering the influence of temperature can be easily used in practice by a microprocessor

7. References

[1] K. L. Butler, M. Ehsani, and P. Kamath, "A matlab-based modeling and simulation package for electric and hybrid electric vehicle design," *IEEE Trans. Veh. Technol.*, vol. 48, no. 6, pp. 1770–1778, Nov. 1999.

[2] O. Caumont, P. L. Moigne, C. Rombaut, X. Muneret, and P. Lenain,"Energy gauge for lead acid batteries in electric vehicles," *IEEE Trans. Energy Convers.*, vol. 15, no. 3, pp. 354–360, 2000.

[3] M. Ceraol and G. Pede, "Techniques for estimating the residual range of an electric vehicle," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 109–115,Jan. 2001.

[4] C. C. Chan, "The state of the art of electric and hybrid vehicles," *Proc.IEEE*, vol. 90, no. 2, pp. 247–275, 2002.

[5] Valerie H. Johnson, Ahmad A. Pesaran, "Temperature-dependent battery models for high-power lithium-ion batteries", in Proc. *International Electric Vehicle Symposium,* vol. 2, 2000, pp. 1–6.

[6] W. Gu and C. Wang, "Thermal-electrochemical modeling of battery systems", *Journal of the Electrochemical Society* vol.147, No.8, (2000), pp. 2910-22.

[7] Szumanowski A. "Fundamentals of hybrid vehicle drives" Monograph Book, ISBN 83-7204-114-8, Warsaw-Radom 2000.

[8] Szumanowski A. "Hybrid electric vehicle drives design—edition based on urban buses" Monograph Book, ISBN 83-7204-456-2, Warsaw-Radom 2006.

[9] Robert F. Nelson, "Power requirements for battery in HEVs", *Journal of Power Sources,vol.* 91, pp.2-26, 2000.

[10] E. Karden, S. Buller, and R. W. De Doncker, "A frequency-domain approach to dynamical modeling of electrochemical power sources," Electrochimica Acta, vol. 47, no. 13–14, pp. 2347–2356, 2002.D.

[11] J. Marcos, A. Lago, C. M. Penalver, J. Doval, A. Nogueira, C. Castro, and J. Chamadoira, "An approach to real behaviour modeling for traction lead-acid batteries," in Proc. *Power Electronics Specialists Conference*, vol. 2, 2001, pp. 620–624.

[12] A. Salkind, T. Atwater, P. Singh, S. Nelatury, S. Damodar, C. Fennie, and D. Reisner, "Dynamic characterization of small lead-acid cells," *J. Power Sources*, vol. 96, no. 1, pp. 151–159, 2001.

[13] G. Plett "LiPB dynamic cell models for Kalman-Filter SOC estimation", *Proc. International Electric Vehicle Symposium*, 2003, CD-ROM.

[14] S. Pang, J. Farrell, J. Du, and M. Barth, "Battery state-of-charge estimation," in Proc. *American Control Conference*, vol. 2, 2001, pp. 1644–1649.

[15] S. Malkhandi, S. K. Sinha, and K. Muthukumar, "Estimation of state of charge of lead acid battery using radial basis function," in Proc. *Industrial Electronics Conference*, vol. 1, 2001, pp. 131–136.

[16] S. Rodrigues, N. Munichandraiah, A. Shukla, "A review of state-of-charge indication of batteries by means of a.c. impedance measurements", *Journal of Power Sources*, vol.87, No.1-2, 2000, pp.12-20.

[17] L. Jyunichi and T. Hiroya, "Battery state-of-charge indicator for electric vehicle," in Proc. *International Electric Vehicle Symposium,* vol. 2, 1996, pp. 229–234.

[18] S. Sato and A. Kawamura, "A new estimation method of state of charge using terminal voltage and internal resistance for lead acid battery," in Proc. *Power*, vol. 2, 2002, pp. 565–570.

[19] W. X. Shen, C. C. Chan, E. W. C. Lo, and K. T. Chau, "Estimation of battery available capacity under variable discharge currents," J. *Power Sources*, vol. 103, no. 2, pp. 180–187, 2002.

[20] W. X. Shen, K. T. Chau, C. C. Chan, Edward W. C. Lo, "Neural network-based residual capacity indicator for Nickel-Metal Hydride batteries in electric vehicles" *IEEE Trans. Veh. Technol.*,vol. 54, no. 5, pp. 1705–1712, Sep. 2005

[21] K. Morio, H. Kazuhiro, and P. Anil, "Battery SOC and distance to empty meter of the honda EV plus," in *Proc. International Electric Vehicle Symposium*, 1997, pp. 1–10.

[22] O. Caumont, P. L. Moigne, C. Rombaut, X. Muneret, and P. Lenain,"Energy gauge for lead-acid batteries in electric vehicles," *IEEE Trans.Energy Convers.*, vol. 15, no. 3, pp. 354–360, Sep. 2000.

[23] Sabine Piller, Marion Perrin, Andreas Jossen "Methods for state–of–charge determination and their applications", *Journal of Power Sources*, vol. 96 , pp.113-120, 2001.

[24] Antoni Szumanowski, Jakub Dębicki, Arkadiusz Hajduga, Piotr Piórkowski, Chang Yuhua, "Li-ion battery modeling and monitoring approach for hybrid electric vehicle applications", *Proc. International Electric Vehicle Symposium,* 2003, CD-ROM.

[25] Antoni Szumanowski, Yuhua Chang "Battery Management System Based on Battery Nonlinear Dynamics Modeling" IEEE Transactions on Vehicular Technology, Vol. 57 no.3 May 2008

10

Entropic Geometry of Crowd Dynamics

Vladimir G. Ivancevic and Darryn J. Reid
Land Operations Division, Defence Science & Technology Organisation
Australia

1. Introduction

In this Chapter we propose a nonlinear entropic model of crowd generic psycho–physical[1] dynamics. For this we use Feynman's action–amplitude formalism, operating on microscopic, mesoscopic and macroscopic synergetic levels, which correspond to individual, group (aggregate) and full crowd behavior dynamics, respectively. In all three levels, goal–directed behavior operates under entropy conservation, $\partial_t S = 0$, while naturally chaotic behavior operates under (monotonically) increasing entropy, $\partial_t S > 0$. Between these two distinct behavioral phases lies a topological phase transition with a chaotic inter-phase. We formulate a geometrical representation of this behavioral transition in terms of the Perelman-Ricci flow on the crowd's Riemannian configuration manifold.

Recall that in psychology the term *cognition*[2] refers to an information processing view of an individual psychological functions (see [3; 4; 68; 81; 88]). More generally, cognitive processes can be natural and artificial, conscious and not conscious; therefore, they are analyzed from different perspectives and in different contexts, e.g., anesthesia, neurology, psychology, philosophy, logic (both Aristotelian and mathematical), systemics, computer science, artificial intelligence (AI) and computational intelligence (CI). Both in psychology and in AI/CI, cognition refers to the mental functions, mental processes and states of intelligent entities (humans, human organizations, highly autonomous robots), with a particular focus toward the study of comprehension, inferencing, decision–making, planning and learning (see, e.g. [11]). The recently developed Scholarpedia, the free peer reviewed web encyclopedia of computational neuroscience is largely based on cognitive neuroscience (see, e.g. [79]). The concept of cognition is closely related to such abstract concepts as mind, reasoning, perception, intelligence, learning, and many others that describe numerous capabilities of the human mind and expected properties of AI/CI (see [51; 57] and references therein).

Yet disembodied cognition is a myth, albeit one that has had profound influence in Western science since Rene Descartes and others gave it credence during the Scientific Revolution. In fact, the mind-body separation had much more to do with explanation of method than with explanation of the mind and cognition, yet it is with respect to the latter that its impact is most widely felt. We find it to be an unsustainable assumption in the realm of crowd behavior.

[1] The new term "psychophysical" should not be confused with the reserved psychological term "psychophysics". By psycho-physical we mean cognitive–to–physical transition behavior: from mental idea to physical manifestation.

[2] Latin: "cognoscere = to know"

Mental intention is (almost immediately) followed by a physical action, that is, a human or animal movement [82]. In animals, this physical action would be jumping, running, flying, swimming, biting or grabbing. In humans, it can be talking, walking, driving, or shooting, etc. Mathematical description of human/animal movement in terms of the corresponding neuro-musculo-skeletal equations of motion, for the purpose of prediction and control, is formulated within the realm of biodynamics (see [43; 44; 45; 46; 47; 48; 49; 55]).

The crowd (or, collective) behavior is clearly formed by some kind of *superposition*, *contagion*, *emergence*, or *convergence* from the individual agents' behavior. Le Bon's 1895 contagion theory, presented in "The Crowd: A Study of the Popular Mind" influenced many 20th century figures. Sigmund Freud criticized Le Bon's concept of "collective soul," asserting that crowds do not have a soul of their own. The main idea of Freudian crowd behavior theory was that people who were in a crowd acted differently towards people than those who were thinking individually: the minds of the group would merge together to form a collective way of thinking. This idea was further developed in Jungian famous "collective unconscious" [63]. The term "collective behavior" [8] refers to social processes and events which do not reflect existing social structure (laws, conventions, and institutions), but which emerge in a "spontaneous" way. Collective behavior might also be defined as action which is neither conforming (in which actors follow prevailing norms) nor deviant (in which actors violate those norms). According to the emergence theory [86], crowds begin as collectivities composed of people with mixed interests and motives; especially in the case of less stable crowds (expressive, acting and protest crowds) norms may be vague and changing; people in crowds make their own rules as they go along. According to currently popular convergence theory, crowd behavior is not a product of the crowd itself, but is carried into the crowd by particular individuals, thus crowds amount to a convergence of like–minded individuals.

We propose that the contagion and convergence theories may be unified by acknowledging that both factors may coexist, even within a single scenario: we propose to refer to this third approach as *behavioral composition*. It represents a substantial philosophical shift from traditional analytical approaches, which have assumed either reduction of a whole into parts or the emergence of the whole from the parts. In particular, both contagion and convergence are related to social entropy, which is the natural decay of structure (such as law, organization, and convention) in a social system [16]. Thus, social entropy provides an entry point into realizing a behavioral–compositional theory of crowd dynamics.

Thus, while all mentioned psycho-social theories of crowd behavior are explanatory only, in this paper we attempt to formulate a geometrically predictive model–theory of crowd psychophysical behavior.

In this chapter we attempt to formulate a geometrically predictive model–theory of crowd behavioral dynamics, based on the previously formulated individual Life Space Foam concept [54].[3]

[3] General nonlinear stochastic dynamics, developed in a framework of Feynman path integrals, have recently [54] been applied to Lewinian field–theoretic psychodynamics [67], resulting in the development of a new concept of life–space foam (LSF) as a natural medium for motivational and cognitive psychodynamics. According to the LSF–formalism, the classic Lewinian life space can be macroscopically represented as a smooth manifold with steady force–fields and behavioral paths, while at the microscopic level it is more realistically represented as a collection of wildly fluctuating force–fields, (loco)motion paths and local geometries (and topologies with holes).

It is today well known that massive crowd movements can be precisely observed/monitored from satellites and all that one can see is crowd physics. Therefore, all involved psychology of individual crowd agents: cognitive, motivational and emotional – is only a

---

A set of least–action principles is used to model the smoothness of global, macro–level LSF paths, fields and geometry, according to the following prescription. The action $S[\Phi]$, with dimensions of *Energy* ×*Time* = *Effort* and depending on macroscopic paths, fields and geometries (commonly denoted by an abstract field symbol Φ^i) is defined as a temporal integral from the initial time instant t_{ini} to the final time instant t_{fin},

$$S[\Phi]=\int_{t_{ini}}^{t_{fin}}\mathfrak{L}[\Phi]dt, \tag{1}$$

with Lagrangian density given by

$$\mathfrak{L}[\Phi]=\int d^n x\,\mathcal{L}(\Phi_i,\partial_{x^j}\Phi^i),$$

where the integral is taken over all n coordinates $x^j = x^j\,(t)$ of the LSF, and $\partial_{x^j}\Phi^i$ are time and space partial derivatives of the Φ^i-variables over coordinates. The standard least action principle

$$\delta S[\Phi]=0, \tag{2}$$

gives, in the form of the so–called Euler–Lagrangian equations, a shortest (loco)motion path, an extreme force–field, and a life–space geometry of minimal curvature (and without holes). In this way, we have obtained macro–objects in the global LSF: a single path described by Newtonian–like equation of motion, a single force–field described by Maxwellian–like field equations, and a single obstacle–free Riemannian geometry (with global topology without holes).

To model the corresponding local, micro–level LSF structures of rapidly fluctuating MD & CD, an adaptive path integral is formulated, defining a multi–phase and multi–path (multi–field and multi– geometry) transition amplitude from the motivational state of *Intention* to the cognitive state of *Action*,

$$\langle Action\,|\,Intention\rangle_{total} := \int\mathcal{D}[w\Phi]\mathrm{e}^{\mathrm{i}S[\Phi]}, \tag{3}$$

where the Lebesgue integration is performed over all continuous $\Phi^i_{con}=paths+fields+geometries$, while summation is performed over all discrete processes and regional topologies Φ^j_{dis}. The symbolic differential $\mathcal{D}[w\Phi]$ in the general path integral (24), represents an adaptive path measure, defined as a weighted product

$$\mathcal{D}[w\Phi]=\lim_{N\to\infty}\prod_{s=1}^{N}w_s d\Phi^i_s,(i=1,...,n=con+dis). \tag{4}$$

The adaptive path integral (3)–(11) represents an ∞–dimensional neural network, with weights w updating by the general rule [57]

$$new\ value(t+1) = old\ value(t) + innovation(t).$$

non-transparent input (a hidden initial switch) for the fully observable crowd physics. In this paper we will label this initial switch as 'mental preparation' or 'loading', while the manifested physical action is labeled 'hitting'.

We propose the entropy formulation of crowd dynamics as a three–step process involving individual behavioral dynamics and collective behavioral dynamics. The chaotic behavioral phase transitions embedded in crowd dynamics may give a formal description for a phenomenon called *crowd turbulence* by D. Helbing, depicting crowd disasters caused by the panic stampede that can occur at high pedestrian densities and which is a serious concern during mass events like soccer championship games or annual pilgrimage in Makkah (see [37; 38; 39; 62]).

In this paper we propose the entropy formulation of crowd dynamics as a three–step process involving individual dynamics and collective dynamics.

2. Generic three–step crowd psycho–physical behavior

In this section we model a generic crowd dynamics (see e.g., [36; 69]) as a three–step process based on a general partition function formalism. Note that the number of variables X_i in the standard partition function from statistical mechanics (see equation (59) in Appendix) need not be countable, in which case the set of coordinates $\{x^i\}$ becomes a field $\phi = \phi(x)$, so the sum is to be replaced by the *Euclidean path integral* (that is a Wick–rotated Feynman transition amplitude in imaginary time, see subsection 3.4), as

$$Z(\phi) = \int \mathcal{D}[\phi] \exp[-H(\phi)],$$

More generally, in quantum field theory, instead of the field Hamiltonian $H(\phi)$ we have the action $S(\phi)$ of the theory. Both Euclidean path integral,

$$Z(\phi) = \int \mathcal{D}[\phi] \exp[-S(\phi)], \qquad \text{real path integral in imaginary time} \tag{5}$$

and Lorentzian one,

$$Z(\phi) = \int \mathcal{D}[\phi] \exp[iS(\phi)], \qquad \text{complex path integral in real time} \tag{6}$$

–r epresent quantum field theory (QFT) partition functions. We will give formal definitions of the above path integrals (i.e., general partition functions) in section 3. For the moment, we only remark that the Lorentzian path integral (6) represents a QFT generalization of the (nonlinear) Schrödinger equation, while the Euclidean path integral (5) in the (rectified) real time represents a statistical field theory (SFT) generalization of the Fokker–Planck equation.

Now, following the framework of the Extended Second Law of Thermodynamics (see Appendix), $\partial_t S \geq 0$, for entropy S in any complex system described by its partition function, we formulate a generic crowd dynamics, based on above partition functions, as the following three–step process:

1. Individual dynamics ($\mathcal{ID}$) is a transition process from an entropy–growing "loading" phase of mental preparation, to the entropy–conserving "hitting/executing" phase of physical action. Formally, $\mathcal{ID}$ is given by the phase–transition map:

$$\mathcal{ID} : \overbrace{\text{MENTAL PREPARATION}}^{\text{"LOADING"}:\partial_t S>0} \Rightarrow \overbrace{\text{PHYSICAL ACTION}}^{\text{"HITTING"}:\partial_t S=0} \tag{7}$$

defined by the individual (chaotic) phase–transition amplitude

$$\left\langle \overset{\partial_t S=0}{\text{PHYS. ACTION}} \middle| \mathit{CHAOS} \middle| \overset{\partial_t S>0}{\text{MENTAL PREP.}} \right\rangle_{\text{ID}} := \int \mathcal{D}[\Phi] \mathrm{e}^{iS_{\text{ID}}[\Phi]},$$

where the right-hand-side is the Lorentzian path-integral (or complex path-integral in real time, see Appendix), with the individual action

$$S_{\text{ID}}[\Phi] = \int_{t_{ini}}^{t_{fin}} L_{\text{ID}}[\Phi]dt,$$

where $L_{ID}[\Phi]$ is the behavioral Lagrangian, consisting of mental cognitive potential and physical kinetic energy.

2. Aggregate dynamics ($\mathcal{AD}$) represents the behavioral composition–transition map:

$$\mathcal{AD}: \sum_{i\in\text{AD}} \overbrace{\text{MENTAL PREPARATION}}^{\text{"LOADING"}:\partial_t S>0} \Rightarrow \sum_{i\in\text{AD}} \overbrace{\text{PHYSICAL ACTION}_i}^{\text{"HITTING"}:\partial_t S=0} \tag{8}$$

where the (weighted) aggregate sum is taken over all individual agents, assuming equipartition of the total energy. It is defined by the aggregate (chaotic) phase–transition amplitude

$$\left\langle \overset{\partial_t S=0}{\text{PHYS. ACTION}} \middle| \mathit{CHAOS} \middle| \overset{\partial_t S>0}{\text{MENTAL PREP.}} \right\rangle_{\text{AD}} := \int \mathcal{D}[\Phi] \mathrm{e}^{-S_{\text{AD}}[\Phi]},$$

with the Euclidean path-integral in real time, that is the SFT–partition function, based on the aggregate behavioral action

$$S_{\text{AD}}[\Phi] = \int_{t_{ini}}^{t_{fin}} L_{\text{AD}}[\Phi]dt, \qquad \text{with} \qquad L_{\text{AD}}[\Phi] = \sum_{i\in\text{AD}} L_{\text{ID}}^{i}[\Phi].$$

3. Crowd dynamics ($\mathcal{CD}$) represents the cumulative transition map:

$$\mathcal{CD}: \sum_{i\in\text{CD}} \overbrace{\text{MENTAL PREPARATION}}^{\text{"LOADING"}:\partial_t S>0} \Rightarrow \sum_{i\in\text{CD}} \overbrace{\text{PHYSICAL ACTION}_i}^{\text{"HITTING"}:\partial_t S=0} \tag{9}$$

where the (weighted) cumulative sum is taken over all individual agents, assuming equipartition of the total behavioral energy. It is defined by the crowd (chaotic) phase–transition amplitude

$$\left\langle \overset{\partial_t S=0}{\text{PHYS. ACTION}} \middle| \mathit{CHAOS} \middle| \overset{\partial_t S>0}{\text{MENTAL PREP.}} \right\rangle_{\text{CD}} := \int \mathcal{D}[\Phi] \mathrm{e}^{iS_{\text{CD}}[\Phi]},$$

with the general Lorentzian path-integral, that is, the QFT–partition function), based on the crowd behavioral action

$$S_{\text{CD}}[\Phi] = \int_{t_{ini}}^{t_{fin}} L_{\text{CD}}[\Phi]dt, \qquad \text{with} \qquad L_{\text{CD}}[\Phi] = \sum_{i\in\text{CD}} L_{\text{ID}}^{i}[\Phi] = \sum_{k=\#\text{ofADsinCD}} L_{\text{AD}}^{k}[\Phi].$$

All three entropic phase–transition maps, $\mathcal{ID}$, $\mathcal{AD}$ and $\mathcal{CD}$, are spatio–temporal biodynamic cognition systems, evolving within their respective configuration manifolds (i.e., sets of their respective degrees-of-freedom with equipartition of energy), according to biphasic action–functional formalisms with behavioral Lagrangian functions L_{ID}, L_{AD} and L_{CD}, each consisting of:

1. Cognitive mental potential (which is a mental preparation for the physical action), and
2. Physical kinetic energy (which describes the physical action itself).

To develop $\mathcal{ID}$, $\mathcal{AD}$ and $\mathcal{CD}$ formalisms, we extend into a physical (or, more precisely, biodynamic) crowd domain a purely–mental individual Life–Space Foam (LSF) framework for motivational cognition [54], based on the quantum–probability concept.[4]

[4] The quantum probability concept is based on the following physical facts [58; 59]

1. The time–dependent Schrödinger equation represents a complex–valued generalization of the real–valued Fokker–Planck equation for describing the spatio–temporal probability density function for the system exhibiting continuous–time Markov stochastic process.
2. The Feynman path integral (including integration over continuous spectrum and summation over discrete spectrum) is a generalization of the time–dependent Schrödinger equation, including both continuous–time and discrete–time Markov stochastic processes.
3. Both Schrödinger equation and path integral give 'physical description' of any system they are modelling in terms of its physical energy, instead of an abstract probabilistic description of the Fokker–Planck equation.

Therefore, the Feynman path integral, as a generalization of the (nonlinear) time–dependent Schrödinger equation, gives a unique physical description for the general Markov stochastic process, in terms of the physically based generalized probability density functions, valid both for continuous–time and discrete–time Markov systems. Its basic consequence is this: a different way for calculating probabilities. The difference is rooted in the fact that *sum of squares is different from the square of sums*, as is explained in the following text. Namely, in Dirac–Feynman quantum formalism, each possible route from the initial system state A to the final system state B is called a history. This history comprises any kind of a route, ranging from continuous and smooth deterministic (mechanical–like) paths to completely discontinues and random Markov chains (see, e.g., [23]). Each history (labelled by index i) is quantitatively described by a complex number.

In this way, the overall probability of the system's transition from some initial state A to some final state B is given not by adding up the probabilities for each history–route, but by 'head–to–tail' adding up the sequence of amplitudes making–up each route first (i.e., performing the sum–over–histories) – to get the total amplitude as a 'resultant vector', and then squaring the total amplitude to get the overall transition probability.

Here we emphasize that the domain of validity of the 'quantum' is not restricted to the microscopic world [87]. There are macroscopic features of classically behaving systems, which cannot be explained without recourse to the quantum dynamics. This field theoretic model leads to the view of the phase transition as a condensation that is comparable to the formation of fog and rain drops from water vapor, and that might serve to model both the gamma and beta phase transitions. According to such a model, the production of activity with long–range correlation in the brain takes place through the mechanism of spontaneous

The behavioral dynamics approach to $\mathcal{ID}$, $\mathcal{AD}$ and $\mathcal{CD}$ is based on *entropic motor control* [41; 42], which deals with neuro-physiological feedback information and environmental uncertainty. The probabilistic nature of human motor action can be characterized by entropies at the level of the organism, task, and environment. Systematic changes in motor adaptation are characterized as task–organism and environment–organism tradeoffs in entropy. Such compensatory adaptations lead to a view of goal–directed motor control as the product of an underlying conservation of entropy across the task–organism–environment system. In particular, an experiment conducted in [42] examined the changes in entropy of the coordination of isometric force output under different levels of task demands and feedback from the environment. The goal of the study was to examine the hypothesis that human motor adaptation can be characterized as a process of entropy conservation that is reflected in the compensation of entropy between the task, organism motor output, and environment. Information entropy of the coordination dynamics relative phase of the motor output was made conditional on the idealized situation of human movement, for which the goal was always achieved. Conditional entropy of the motor output decreased as the error tolerance and feedback frequency were decreased. Thus, as the likelihood of meeting the task demands was decreased increased task entropy and/or the amount of information from the environment is reduced increased environmental entropy, the subjects of this experiment employed fewer coordination patterns in the force output to achieve the goal. The conservation of entropy supports the view that context dependent adaptations in human goal–directed action are guided fundamentally by natural law and provides a novel means of examining human motor behavior. This is fundamentally related to the *Heisenberg uncertainty principle* [59] and further supports the argument for the primacy of a probabilistic approach toward the study of biodynamic cognition systems.[5]

breakdown of symmetry (SBS), which has for decades been shown to describe longrange correlation in condensed matter physics. The adoption of such a field theoretic approach enables modelling of the whole cerebral hemisphere and its hierarchy of components down to the atomic level as a fully integrated macroscopic quantum system, namely as a macroscopic system which is a quantum system not in the trivial sense that it is made, like all existing matter, by quantum components such as atoms and molecules, but in the sense that some of its macroscopic properties can best be described with recourse to quantum dynamics (see [22] and references therein). Also, according to Freeman and Vitielo, *many–body quantum field theory* appears to be the only existing theoretical tool capable to explain the dynamic origin of long-range correlations, their rapid and efficient formation and dissolution, their interim stability in ground states, the multiplicity of coexisting and possibly non–interfering ground states, their degree of ordering, and their rich textures relating to sensory and motor facets of behaviors. It is historical fact that many–body quantum field theory has been devised and constructed in past decades exactly to understand features like ordered pattern formation and phase transitions in condensed matter physics that could not be understood in classical physics, similar to those in the brain.

[5] Our entropic action–amplitude formalism represents a kind of a generalization of the Haken-Kelso- Bunz (HKB) model of self-organization in the individual's motor system [24; 65], including: multistability, phase transitions and hysteresis effects, presenting a contrary view to the purely feedback driven systems. HKB uses the concepts of synergetics (order

On the other hand, it is well known that humans possess more degrees of freedom than are needed to perform any defined motor task, but are required to co-ordinate them in order to reliably accomplish high-level goals, while faced with intense motor variability. In an attempt to explain how this takes place, Todorov and Jordan have formulated an alternative theory of human motor co-ordination based on the concept of stochastic optimal feedback control [84]. They were able to conciliate the requirement of goal achievement (e.g., grasping an object) with that of motor variability (biomechanical degrees of freedom). Moreover, their theory accommodates the idea that the human motor control mechanism uses internal 'functional synergies' to regulate task–irrelevant (redundant) movement.

Also, a developing field in coordination dynamics involves the theory of social coordination, which attempts to relate the DC to normal human development of complex social cues following certain patterns of interaction. This work is aimed at understanding how human social interaction is mediated by meta-stability of neural networks. fMRI and EEG are particularly useful in mapping thalamocortical response to social cues in experimental studies. In particular, a new theory called the *Phi complex* has been developed by S. Kelso and collaborators, to provide experimental results for the theory of social coordination dynamics (see the recent nonlinear dynamics paper discussing social coordination and EEG dynamics [85]). According to this theory, a pair of phi rhythms, likely generated in the mirror neuron system, is the hallmark of human social coordination. Using a dual–EEG recording system, the authors monitored the interactions of eight pairs of subjects as they moved their fingers with and without a view of the other individual in the pair.

Finally, the chaotic behavioral phase transitions embedded in $\mathcal{CD}$ may give a formal description for a phenomenon called *crowd turbulence* by D. Helbing, depicting crowd disasters caused by the panic stampede that can occur at high pedestrian densities and

parameters, control parameters, instability, etc) and the mathematical tools of nonlinearly coupled (nonlinear) dynamical systems to account for self-organized behavior both at the cooperative, coordinative level and at the level of the individual coordinating elements. The HKB model stands as a building block upon which numerous extensions and elaborations have been constructed. In particular, it has been possible to derive it from a realistic model of the cortical sheet in which neural areas undergo a reorganization that is mediated by intra- and inter-cortical connections. Also, the HKB model describes phase transitions ('switches') in coordinated human movement as follows: (i) when the agent begins in the anti-phase mode and speed of movement is increased, a spontaneous switch to symmetrical, in-phase movement occurs; (ii) this transition happens swiftly at a certain critical frequency; (iii) after the switch has occurred and the movement rate is now decreased the subject remains in the symmetrical mode, i.e. she does not switch back; and (iv) no such transitions occur if the subject begins with symmetrical, in-phase movements. The HKB dynamics of the order parameter relative phase as is given by a nonlinear first-order ODE:

$$\dot{\phi} = (\alpha + 2\beta r^2)\sin\phi - \beta r^2 \sin 2\phi,$$

where ϕ is the phase relation (that characterizes the observed patterns of behavior, changes abruptly at the transition and is only weakly dependent on parameters outside the phase transition), r is the oscillator amplitude, while α, β are coupling parameters (from which the critical frequency where the phase transition occurs can be calculated).

which is a serious concern during mass events like soccer championship games or annual pilgrimage in Makkah (see [37; 38; 39; 62]).

3. Formal crowd dynamics

In this section we formally develop a three-step crowd behavioral dynamics, conceptualized by transition maps (7)-(8)-(9), in agreement with Haken's synergetics [25; 26]. We first develop a macro-level individual behavioral dynamics $\mathcal{ID}$. Then we generalize $\mathcal{ID}$ into an 'orchestrated' behavioral-compositional crowd dynamics $\mathcal{CD}$, using a quantum-like micro-level formalism with individual agents representing 'crowd quanta'. Finally we develop a meso-level aggregate statistical-field dynamics $\mathcal{AD}$, such that composition of the aggregates $\mathcal{AD}$ makes-up the crowd.

3.1 Individual behavioral dynamics ($\mathcal{ID}$)

$\mathcal{ID}$ transition map (7) is developed using the following action-amplitude formalism (see [53; 54]):

1. Macroscopically, as a smooth Riemannian n-manifold M_{ID} (see Appendix) with steady force-fields and behavioral paths, modelled by a real-valued classical action functional $S_{ID}[\Phi]$, of the form

$$S_{\mathrm{ID}}[\Phi] = \int_{t_{ini}}^{t_{fin}} L_{\mathrm{ID}}[\Phi]dt,$$

(where macroscopic paths, fields and geometries are commonly denoted by an abstract field symbol Φ^i) with the potential-energy based Lagrangian L given by

$$L_{\mathrm{ID}}[\Phi] = \int d^n x \mathcal{L}_{\mathrm{ID}}(\Phi_i, \partial_{x^j}\Phi^i),$$

where $\mathcal{L}$ is Lagrangian density, the integral is taken over all n local coordinates $x^j = x^j(t)$ of the ID, and $\partial_{x^j}\Phi^i$ are time and space partial derivatives of the Φ^i -variables over coordinates. The standard least action principle

$$\delta S_{\mathrm{ID}}[\Phi] = 0,$$

gives, in the form of the Euler-Lagrangian equations, a shortest path, an extreme force-field, with a geometry of minimal curvature and topology without holes. We will see below that high Riemannian curvature generates chaotic behavior, while holes in the manifold produce topologically induced phase transitions.

2. Microscopically, as a collection of wildly fluctuating and jumping paths (histories), force-fields and geometries/topologies, modelled by a complex-valued adaptive path integral, formulated by defining a multi-phase and multi-path (multi-field and multi-geometry) transition amplitude from the entropy-growing state of Mental Preparation to the entropy-conserving state of Physical Action,

$$\langle \text{Physical Action} | \text{Mental Preparation} \rangle_{\mathrm{ID}} := \int_{\mathrm{ID}} \mathcal{D}[\Phi] e^{iS_{\mathrm{ID}}[\Phi]} \qquad (10)$$

where the functional ID-measure $\mathcal{D}[w\Phi]$ is defined as a weighted product

$$\mathcal{D}[w\Phi] = \lim_{N\to\infty}\prod_{s=1}^{N} w_s d\Phi_s^i, \qquad (i = 1,...,n = con + dis), \tag{11}$$

representing an ∞–dimensional neural network [54], with weights w_s updating by the general rule

$$new\ value(t+1) = old\ value(t) + innovation(t).$$

More precisely, the weights $w_s = w_s(t)$ in (11) are updated according to one of the two standard neural learning schemes, in which the micro–time level is traversed in discrete steps, i.e., if $t = t_0, t_1, ..., t_s$ then $t + 1 = t_1, t_2, ..., t_{s+1}$: [6]

a. A *self–organized, unsupervised* (e.g., Hebbian–like [35]) learning rule:

$$w_s(t+1) = w_s(t) + \frac{\sigma}{\eta}(w_s^d(t) - w_s^a(t)), \tag{12}$$

where $\sigma = \sigma(t)$, $\eta = \eta(t)$ denote *signal* and *noise*, respectively, while superscripts d and a denote *desired* and *achieved* micro–states, respectively; or

b. A certain form of a *supervised gradient descent learning*:

$$w_s(t+1) = w_s(t) - \eta\nabla J(t), \tag{13}$$

where η is a small constant, called the *step size*, or the *learning rate,* and $\nabla J(n)$ denotes the gradient of the 'performance hyper–surface' at the t–th iteration.

(Note that we could also use a reward–based, reinforcement learning rule [83], in which system learns its optimal policy: $innovation(t) = |reward(t) - penalty(t)|$.)

In this way, we effectively derive a unique and globally smooth, causal and entropic phase-transition map (7), performed at a macroscopic (global) time–level from some initial time t_{ini} to the final time t_{fin}. Thus, we have obtained macro–objects in the ID: a single path described by Newtonian–like equation of motion, a single force–field described by Maxwellian–like field equations, and a single obstacle–free Riemannian geometry (with global topology without holes).

In particular, on the macro–level, we have the ID–paths, that is biodynamical trajectories generated by the Hamilton action principle

$$\delta S_{\mathrm{ID}}[x] = 0,$$

with the Newtonian action $S_{ID}[x]$ given by (Einstein's summation convention over repeated indices is always assumed)

$$S_{\mathrm{ID}}[x] = \int_{t_{ini}}^{t_{fin}} [\varphi + \frac{1}{2} g_{ij}\dot{x}^i\dot{x}^j]dt, \tag{14}$$

[6] The traditional neural networks approaches are known for their classes of functions they can represent. Here we are talking about functions in an *extensional* rather than merely *intensional* sense; that is, function can be read as input/output behavior [5; 6; 19; 34]. This limitation has been attributed to their low-dimensionality (the largest neural networks are limited to the order of 10^5 dimensions [61]). The proposed path integral approach represents a new family of function-representation methods, which potentially offers a basis for a fundamentally more expansive solution.

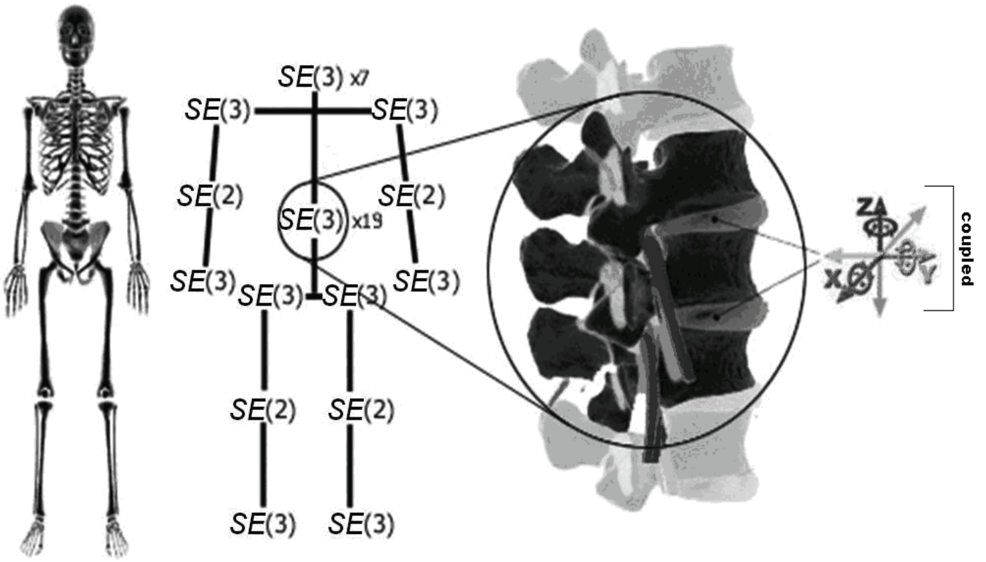


Fig. 1. Riemannian configuration manifold M_{ID} of human biodynamics is defined as a topological product $M = \prod_i SE(3)^i$ of constrained Euclidean $SE(3)$–groups of rigid body motion in 3D Euclidean space (see [49; 52]), acting in all major (synovial) human joints. The manifold M is a dynamical structure activated/controlled by potential covariant forces (16) produced by a synergetic action of about 640 skeletal muscles [47].

where $\varphi = \varphi(t, x^i)$ denotes the mental LSF–potential field, while the second term,

$$T = \frac{1}{2} g_{ij} \dot{x}^i \dot{x}^j ,$$

represents the physical (biodynamic) kinetic energy generated by the Riemannian inertial metric tensor g_{ij} of the configuration biodynamic manifold M_{ID} (see Figure 1). The corresponding Euler–Lagrangian equations give the Newtonian equations of human movement

$$\frac{d}{dt} T_{\dot{x}^i} - T_{x^i} = F_i , \tag{15}$$

where subscripts denote the partial derivatives and we have defined the covariant muscular forces $F_i = F_i(t, x^i, \dot{x}^i)$ as negative gradients of the mental potential $\varphi(x^i)$,

$$F_i = -\varphi_{x^i} . \tag{16}$$

Equation (15) can be put into the standard Lagrangian form as

$$\frac{d}{dt} L_{\dot{x}^i} = L_{x^i} , \qquad \text{with} \qquad L = T - \varphi(x^i), \tag{17}$$

or (using the Legendre transform) into the forced, dissipative Hamiltonian form [44; 47]

$$\dot{x}^i = \partial_{p_i} H + \partial_{p_i} R, \qquad \dot{p}_i = F_i - \partial_{x^i} H + \partial_{x^i} R, \tag{18}$$

where p_i are the generalized momenta (canonically–conjugate to the coordinates x^i), $H = H(p, x)$ is the Hamiltonian (total energy function) and $R = R(p, x)$ is the general dissipative function.

The human motor system possesses many independently controllable components that often allow for more than a single movement pattern to be performed in order to achieve a goal.

Hence, the motor system is endowed with a high level of adaptability to different tasks and also environmental contexts [42]. The multiple SE(3)–dynamics applied to human musculo–skeletal system gives the fundamental law of biodynamics, which is the *covariant force law*:

$$\text{Force co}-\text{vector field} = \text{Mass distribution} \times \text{Acceleration vector}-\text{field}, \tag{19}$$

which is formally written:

$$F_i = g_{ij} a^j, \qquad (i, j = 1, \ldots, n = \dim(M))$$

where F_i are the covariant force/torque components, g_{ij} is the inertial metric tensor of the configuration Riemannian manifold $M = \prod_i SE(3)^i$ (g_{ij} defines the mass–distribution of the human body), while a^j are the contravariant components of the linear and angular acceleration vector-field. (This fundamental biodynamic law states that contrary to common perception, acceleration and force are not quantities of the same nature: while acceleration is a non-inertial vector-field, force is an inertial co-vector-field. This apparently insignificant difference becomes crucial in injury prediction/prevention, especially in its derivative form in which the 'massless jerk' (= $\dot{a}$) is relatively benign, while the 'massive jolt' (= $\dot{F}$) is deadly.) Both Lagrangian and (topologically equivalent) Hamiltonian development of the covariant force law is fully elaborated in [47; 48; 49; 52]. This is consistent with the postulation that human action is guided primarily by natural law [66].

On the micro–ID level, instead of each single trajectory defined by the Newtonian equation of motion (15), we have an ensemble of fluctuating and crossing paths on the configuration manifold M with weighted probabilities (of the unit total sum). This ensemble of micro–paths is defined by the simplest instance of our adaptive path integral (10), similar to the Feynman's original sum over histories,

$$\langle \text{P}\mathit{hysical\ Action} | \mathit{Mental\ Preparation} \rangle_M = \int_{ID} \mathcal{D}[wx] \mathrm{e}^{\mathrm{i}S[x]}, \tag{20}$$

where $\mathcal{D}[wx]$ is the functional ID–measure on the space of all weighted paths, and the exponential depends on the action $S_{\text{ID}}[x]$ given by (14).

3.2 Crowd behavioral–compositional dynamics ($\mathcal{CD}$)

In this subsection we develop a generic crowd $\mathcal{CD}$, as a unique and globally smooth, causal and entropic phase–transition map (9), in which agents (or, crowd's individual entities) can be both humans and robots. This crowd behavioral action takes place in a crowd smooth Riemannian 3*n*-manifold *M*. Recall from Figure 1 that each individual segment of a human body moves in the Euclidean 3–space $\mathbb{R}^3$ according to its own constrained SE(3)–group. Similarly, each individual agent's trajectory, $x^i = x^i(t)$, $i = 1, ...n$, is governed by the Euclidean SE(2)–group of rigid body motions in the plane. (Recall that a Lie group $SE(2) \equiv SO(2) \times \mathbb{R}$ is a set of all 3 × 3– matrices of the form:

$$\begin{bmatrix} \cos\theta & \sin\theta & x \\ -\sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix},$$

including both rigid translations (i.e., Cartesian *x,y*–coordinates) and rotation matrix $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ in Euclidean plane $\mathbb{R}^2$ (see [49; 52]). The crowd configuration manifold *M* is defined as a union of Euclidean SE(2)–groups for all *n* individual agents in the crowd, that is crowd's configuration 3*n*–manifold is defined as a set

$$M = \sum_{k=1}^{n} SE(2)^k \equiv \sum_{k=1}^{n} SO(2)^k \times \mathbb{R}^k, \tag{21}$$

$$\text{coordinated by } \mathbf{x}^k = \{x^k, y^k, \theta^k\}, \qquad (\text{for } k = 1,2,...,n).$$

In other words, the crowd configuration manifold *M* is a *dynamical planar graph* with individual agents' SE(2)–groups of motion in the vertices and time-dependent inter-agent distances $I_{ij} = \left[x^i(t_i) - x^j(t_j)\right]$ as edges.

Similarly to the individual case, the crowd action functional includes mental cognitive potential and physical kinetic energy, formally given by (with *i*, *j* = 1, ..., 3*n*):

$$A[x^i, x^j; t_i, t_j] = \frac{1}{2}\int_{t_i}\int_{t_j} \delta(I_{ij}^2)\, \dot{x}^i(t_i)\dot{x}^j(t_j)\, dt_i dt_j + \frac{1}{2}\int_t g_{ij}\, \dot{x}^i(t)\dot{x}^j(t) dt, \tag{22}$$

$$\text{with } I_{ij}^2 = \left[x^i(t_i) - x^j(t_j)\right]^2, \qquad \text{where } IN \le t_i, t_j, t \le OUT.$$

The first term in (22) represents the mental potential for the interaction between any two agents x^i and x^i within the total crowd matrix x^{ij}. (Although, formally, this term contains cognitive velocities, it still represents 'potential energy' from the physical point of view.) It is defined as a double integral over a delta function of the square of interval I^2 between two points on the paths in their individual cognitive LSFs. Interaction occurs only when this LSF– distance between the two agents x^i and x^j vanishes. Note that the cognitive intentions of any two agents generally occur at different times t_i and t_j unless $t_i = t_j$, when cognitive synchronization occurs. This term effectively represents the *crowd cognitive controller* (see [53]).

The second term in (22) represents kinetic energy of the physical interaction of agents. Namely, after the above cognitive synchronization is completed, the second term of physical kinetic energy is activated in the common CD manifold, reducing it to just one of the agents' individual manifolds, which is equivalent to the center-of-mass segment in the human musculo-skeletal system. Therefore, from (22) we can derive a generic Euler–Lagrangian dynamics that is a composition of (17), which also means that we have in place a generic Hamiltonian dynamics that is a amalgamate of (18), as well as the crowd covariant force law (19), the governing law of crowd biodynamics:

$$\text{Crowd force co}-\text{vector field} = \text{Crowd mass distribution} \times \text{Crowd acceleration vector}-\text{field},$$

$$\text{formally: } F_i = g_{ij}a^j, \quad \text{where } g_{ij} \text{ is the inertial metric tensor of crowd manifold } M. \tag{23}$$

The left-hand side of this equation defines forces acting on the crowd, while right-hand defines its mass distribution coupled to the crowd kinematics ($\mathcal{CK}$, described in the next subsection).

At the slave level, the adaptive path integral, representing an ∞–dimensional neural network, corresponding to the crowd behavioral action (22), reads

$$\langle \text{P}hysical\ Action|Mental\ Preparation\rangle_{\text{CD}} = \int_{\text{CD}} \mathcal{D}[w,x,y]\text{e}^{iA[x,y;t_i,t_j]}, \tag{24}$$

where the Lebesgue-type integration is performed over all continuous paths $x^i = x^i(t_i)$ and $y^j = y^j(t_j)$, while summation is performed over all associated discrete Markov fluctuations and jumps. The symbolic differential in the path integral (24) represents an adaptive path measure, defined as the weighted product

$$\mathcal{D}[w,x,y] = \lim_{N\to\infty}\prod_{s=1}^{N} w_{ij}^s dx^i dy^j, \qquad (i,j = 1,...,n). \tag{25}$$

The quantum–field path integral (24)–(25) defines the microstate $\mathcal{CD}$–level, an ensemble of fluctuating and crossing paths on the crowd 3n–manifold M.

The crowd manifold M itself has quite a sophisticated topological structure defined by its macrostate Euler–Lagrangian dynamics. As a Riemannian smooth n–manifold, M gives rise to its fundamental n–groupoid, or n–category $\prod_n(M)$ (see ([49; 52]). In $\prod_n(M)$, 0–cells are points in M; 1–cells are paths in M(i.e., parameterized smooth maps $f : [0,1] \to M$); 2–cells are smooth homotopies (denoted by $\simeq$) of paths relative to endpoints (i.e., parameterized smooth maps $h : [0,1] \times [0,1] \to M$); 3–cells are smooth homotopies of homotopies of paths in M (i.e., parameterized smooth maps $j : [0,1] \times [0,1] \times [0,1] \to M$). Categorical composition is defined by pasting paths and homotopies. In this way, the following recursive homotopy dynamics emerges on the crowd 3n–manifold M:

$\texttt{0}-\texttt{cell}: x_0\bullet \qquad x_0 \in M; \qquad$ in the higher cells below: $t,s \in [0,1]$;

$$\begin{tikzcd} \texttt{1}-\texttt{cell}: x_0\bullet \arrow[r, "f"] & \bullet x_1 \end{tikzcd} \qquad f : x_0 \simeq x_1 \in M,$$

$f : [0,1] \to M, f : x_0 \mapsto x_1, x_1 = f(x_0), f(0) = x_0, f(1) = x_1;$

e.g., linear path: $f(t) = (1-t)\,x_0 + t\,x_1;$ or

Euler–Lagrangian $f-$ dynamics with endpoint conditions (x_0, x_1) :

$\frac{d}{dt} f_{\dot{x}^i} = f_{x^i}$, with $x(0) = x_0, \quad x(1) = x_1, \quad (i = 1, ..., n);$

$$\begin{tikzcd} \texttt{2}-\texttt{cell}: x_0\bullet \arrow[r, bend left=50, "f"{name=U}] \arrow[r, bend right=50, "g"'{name=D}] & \bullet x_1 \arrow[Rightarrow, from=U, to=D, "h"] \end{tikzcd} \qquad h : f \simeq g \in M,$$

$h : [0,1] \times [0,1] \to M, h : f \mapsto g, g = h(f(x_0)),$

$h(x_0, 0) = f(x_0), h(x_0, 1) = g(x_0), h(0, t) = x_0, h(1, t) = x_1$

e.g., linear homotopy: $h(x_0, t) = (1-t)\, f(x_0) + t\, g(x_0);$ or

homotopy between two Euler–Lagrangian $(f, g)-$ dynamics

with the same endpoint conditions (x_0, x_1) :

$\frac{d}{dt} f_{\dot{x}^i} = f_{x^i}$, and $\frac{d}{dt} g_{\dot{x}^i} = g_{x^i}$ with $x(0) = x_0, \quad x(1) = x_1;$

$$\begin{tikzcd} \texttt{3}-\texttt{cell}: x_0\bullet \arrow[r, bend left=60, "f"{name=U}] \arrow[r, bend right=60, "g"'{name=D}] & \bullet x_1 \arrow[Rightarrow, from=U, to=D, bend right=40, "h"'{name=L}] \arrow[Rightarrow, from=U, to=D, bend left=40, "i"{name=R}] \arrow[Rrightarrow, from=L, to=R, "j"] \end{tikzcd} \qquad j : h \simeq i \in M,$$

$j : [0,1] \times [0,1] \times [0,1] \to M, j : h \mapsto i, i = j(h(f(x_0)))$

$j(x_0, t, 0) = h(f(x_0)), j(x_0, t, 1) = i(f(x_0)),$

$j(x_0, 0, s) = f(x_0), j(x_0, 1, s) = g(x_0),$

$j(0, t, s) = x_0, j(1, t, s) = x_1$

e.g., linear composite homotopy: $j(x_0, t, s) = (1-t)\, h(f(x_0)) + t\, i(f(x_0));$

or, homotopy between two homotopies between above two Euler-Lagrangian $(f, g)-$ dynamics with the same endpoint conditions (x_0, x_1).

3.3 Dissipative crowd kinematics ($\mathcal{CD}$)

The crowd action (22) with its amalgamate Lagrangian dynamics (17) and amalgamate Hamiltonian dynamics (18), as well as the crowd force law (23) define the macroscopic crowd dynamics, $\mathcal{CD}$. Suppose, for a moment, that $\mathcal{CD}$ is force–free and dissipation free, therefore conservative. Now, the basic characteristic of the conservative Lagrangian/Hamiltonian systems evolving in the phase space spanned by the system coordinates and their velocities/momenta, is that their *flow* φ_t^L (explained below) preserves the phase–space volume, as proposed by the Liouville theorem, which is the well known fact in statistical mechanics. However, the preservation of the phase volume causes structural instability of the conservative system, i.e., the phase–space spreading effect by which small phase regions R_t will tend to get distorted from the initial one R_o during the conservative system evolution. This problem, governed by entropy growth ($\partial_t S > 0$), is much more serious in higher dimensions than in lower dimensions, since there are so many 'directions' in which the region can locally spread (see [49; 74]). This phenomenon is related to *conservative Hamiltonian chaos* (see section 4 below).

However, this situation is not very frequent in case of 'organized' human crowd. Its self-organization mechanisms are clearly much stronger than the conservative statistical mechanics effects, which we interpret in terms of Prigogine's dissipative structures (see Appendix). Formally, if dissipation of energy in a system is much stronger then its inertial characteristics, then instead of the second-order Newton–Lagrangian dynamic equations of motion, we are actually dealing with the first-order driftless (non-acceleration, non-inertial) kinematic equations of motion (see Appendix, eq. (64)), which is related to *dissipative chaos* [71]. Briefly, the dissipative crowd flow can be depicted like this: from the set of initial conditions for individual agents, the crowd evolves in time towards the set of the corresponding *entangled attractors,*[7] which are mutually separated by fractal (non-integer dimension) separatrices.

In this subsection we elaborate on the dissipative crowd kinematics ($\mathcal{CK}$), which is self–controlled and dominates the $\mathcal{CD}$ if the crowd's inertial forces are much weaker then the crowd's dissipation of energy, presented here in the form of nonlinear velocity controllers.

[7] Recall that quantum entanglement is a quantum mechanical phenomenon in which the quantum states of two or more objects are linked together so that one object can no longer be adequately described without full mention of its counterpart – even though the individual objects may be spatially separated. This interconnection leads to correlations between observable physical properties of remote systems. The related phenomenon of wave-function collapse gives an impression that measurements performed on one system instantaneously influence the other systems entangled with the measured system, even when far apart.

Entanglement has many applications in quantum information theory. Mixed state entanglement can be viewed as a resource for quantum communication. A common measure of entanglement is the entropy of a mixed quantum state (see, e.g. [59]). Since a mixed quantum state ρ is a probability distribution over a quantum ensemble, this leads naturally to the definition of the *von Neumann entropy*, $S(\rho) = -\mathrm{Tr}\,(\rho \log_2 \rho)$, which is obviously similar to the classical *Shannon entropy* for probability distributions ($p_1, \dots , p_n$), defined as $S(p_1, \dots , p_n) = -\Sigma_i\, p_i \log_2 p_i$. As in statistical mechanics, one can say that the more uncertainty (number of microstates) the system should possess, the larger is its entropy. Entropy gives a tool which can be used to quantify entanglement. If the overall system is pure, the entropy of one subsystem can be used to measure its degree of entanglement with the other subsystems.

The most popular issue in a research on dissipative quantum brain modelling has been *quantum entanglement* between the *brain* and its *environment* [77; 78], where the brain–environment system has an entangled 'memory' state, identified with the ground (vacuum) state $|0>_{\mathcal{N}}$, that cannot be factorized into two single–mode states. (In the Vitiello–Pessa dissipative quantum brain model [77; 78], the evolution of the $\mathcal{N}$–coded memory system was represented as a trajectory of given initial condition running over time–dependent states $|0(t)>_{\mathcal{N}}$, each one minimizing the free energy functional.) Similar to this microscopic brain–environment entanglement, we propose a kind of *macroscopic entanglement* between the operating modes of the crowd behavioral controller and its biodynamics, which can be considered as a 'long–range correlation'.

Applied externally to the dimension of the crowd $3n$–manifold M, entanglement effectively reduces the number of active degrees of freedom in (21).

Recall that the essential concept in dynamical systems theory is the notion of a *vector–field* (that we will denote by a boldface symbol), which assigns a tangent vector to each point p in the manifold in case. In particular, **v** is a gradient vector–field if it equals the gradient of some scalar function. A *flow–line* of a vector–field **v** is a path **fl**(t) satisfying the vector ODE, $\dot{\mathbf{fl}}\,(t) = \mathbf{v}(\mathbf{fl}(t))$, that is, **v** yields the velocity field of the path **fl**(t). The set of all flow lines of a vector–field **v** comprises its flow φ_t that is (technically, see e.g., [49; 52]) a one–parameter Lie group of diffeomorphisms (smooth bijective functions) generated by a vector-field **v** on M, such that

$$\varphi_t \circ \varphi_s = \varphi_{t+s}, \qquad \varphi_0 = \text{identity}, \qquad \text{which gives:} \quad \gamma(t) = \varphi_t(\gamma(0)).$$

Analytically, a vector-field **v** is defined as a set of autonomous ODEs. Its solution gives the flow φ_t, consisting of integral curves (or, flow lines) **fl**(t) of the vector–field, such that all the vectors from the vector-field are tangent to integral curves at different representative points $p \in M$. In this way, through every representative point $p \in M$ passes both a curve from the flow and its tangent vector from the vector-field. Geometrically, vector-field is defined as a cross-section of the tangent bundle TM of the manifold M.

In general, given an nD frame $\{\partial_i\} \equiv \{\partial/\partial x^i\}$ on a smooth n–manifold M (that is, a basis of tangent vectors in a local coordinate chart $x^i = (x^1, ..., x^n) \subset M$), we can define any vector-field **v** on M by its components $v^i = v^i(t)$ as

$$\mathbf{v} = v^i\partial_i = v^i\frac{\partial}{\partial x^i} = v^1\frac{\partial}{\partial x^1} + ... + v^n\frac{\partial}{\partial x^n}.$$

Thus, a vector-field $\mathbf{v} \in \mathcal{X}(M)$ (where $\mathcal{X}(M)$ is the set of all smooth vector-fields on M) is actually a differential operator that can be used to differentiate any smooth scalar function $f = f(x^1, ..., x^n)$ on M, as a *directional derivative* of f in the direction of **v**. This is denoted simply **v**f, such that

$$\mathbf{v}f = v^i\partial_i f = v^i\frac{\partial f}{\partial x^i} = v^1\frac{\partial f}{\partial x^1} + ... + v^n\frac{\partial f}{\partial x^n}.$$

In particular, if $\mathbf{v} = \dot{\gamma}\,(t)$ is a velocity vector-field of a space curve $\gamma(t) = (x^1(t), ..., x^n(t))$, defined by its components $v^i = \dot{x}^i\,(t)$, directional derivative of $f\,(x^i)$ in the direction of **v** becomes

$$\mathbf{v}f = \dot{x}^i\partial_i f = \frac{dx^i}{dt}\frac{\partial f}{\partial x^i} = \frac{df}{dt} = \dot{f},$$

which is a rate-of-change of f along the curve $\gamma(t)$ at a point $x^i(t)$.

Given two vector-fields, $\mathbf{u} = u^i\partial_i, \mathbf{v} = v^i\partial_i \in \mathcal{X}(M)$, their Lie bracket (or, commutator) is another vector-field $[\mathbf{u},\mathbf{v}] \in \mathcal{X}(M)$, defined by

$$[\mathbf{u},\mathbf{v}] = \mathbf{uv} - \mathbf{vu} = u^i\partial_i v^j\partial_j - v^j\partial_j u^i\partial_i,$$

which, applied to any smooth function f on M, gives

$$[\mathbf{u},\mathbf{v}](f) = \mathbf{u}\big(\mathbf{v}(f)\big) - \mathbf{v}\big(\mathbf{u}(f)\big).$$

The Lie bracket measures the failure of 'mixed directional derivatives' to commute. Clearly, mixed partial derivatives *do* commute, $[\partial_i, \partial_j] = 0$, while in general it is *not* the case, $[\mathbf{u},\mathbf{v}] \neq 0$. In addition, suppose that **u** generates the flow φ_t and **v** generates the flow φ_s. Then, for any smooth function *f* on *M*, we have at any point *p* on *M*,

$$[\mathbf{u},\mathbf{v}](f)(p) = \frac{\partial^2}{\partial t \partial s}\left(f(\varphi_s(\varphi_t(p))) - f(\varphi_t(\varphi_s(p))\right),$$

which means that in $f(\varphi_s(\varphi_t(p)))$ we are starting at *p*, flowing along **v** a little bit, then along **u** a little bit, and then evaluating *f*, while in $f(\varphi_t(\varphi_s(p)))$ we are flowing first along **u** and then **v**. Therefore, the Lie bracket infinitesimally measures how these flows fail to commute.

The Lie bracket satisfies the following three properties (for any three vector-fields $\mathbf{u},\mathbf{v},\mathbf{w} \in M$ and two constants *a*, *b* – thus forming a Lie algebra on the crowd manifold *M*):

i. $[\mathbf{u},\mathbf{v}] = -[\mathbf{v},\mathbf{u}]$ skew-symmetry;

ii. $[\mathbf{u}, a\mathbf{v} + b\mathbf{w}] = a[\mathbf{u},\mathbf{v}] + b[\mathbf{u},\mathbf{w}] -$ bilinearity; and

iii. $[\mathbf{u},[\mathbf{v},\mathbf{w}]] + [\mathbf{v},[\mathbf{w},\mathbf{u}]] + [\mathbf{w},[\mathbf{u},\mathbf{v}]] -$ Jacobi identity.

A new set of vector-fields on *M* can be generated by repeated Lie brackets of **u**, **v**, $\mathbf{w} \in M$.

The Lie bracket is a standard tool in geometric nonlinear control theory (see, e.g. [49; 52]). Its action on vector-fields can be best visualized using the popular car parking example, in which the driver has two different vector–field transformations at his disposal. They can turn the steering wheel, or they can drive the car forward or backward. Here, we specify the state of a car by four coordinates: the (x, y) coordinates of the center of the rear axle, the direction θ of the car, and the angle ϕ between the front wheels and the direction of the car. *l* is the constant length of the car. Therefore, the 4D configuration manifold of a car is a set $M \equiv SO(2) \times \mathbb{R}^2$, coordinated by $\mathbf{x} \equiv \{x, y, \theta, \phi\}$, which is slightly more complicated than the individual crowd agent's 3D configuration manifold $SE(2) \equiv SO(2) \times \mathbb{R}$, coordinated by $\mathbf{x} = \{x, y, \theta\}$. The driftless car kinematics can be defined as a vector ODE:

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x})c_1 + \mathbf{v}(\mathbf{x})c_2, \tag{26}$$

with two vector–fields, $\mathbf{u},\mathbf{v} \in \mathcal{X}(M)$, and two scalar control inputs, c_1 and c_2. The infinitesimal car–parking transformations will be the following vector–fields

$$\mathbf{u}(\mathbf{x}) \equiv \text{DRIVE} = \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y} + \frac{\tan\phi}{l}\frac{\partial}{\partial \theta} \equiv \begin{pmatrix} \cos\theta \\ \sin\theta \\ \frac{1}{l}\tan\phi \\ 0 \end{pmatrix},$$

$$\text{and} \qquad \mathbf{v}(\mathbf{x}) \equiv \text{STEER} = \frac{\partial}{\partial \phi} \equiv \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

The car kinematics (26) therefore expands into a matrix ODE:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\phi} \end{pmatrix} = \text{DRIVE} \cdot c_1 + \text{STEER} \cdot c_2 \equiv \begin{pmatrix} \cos\theta \\ \sin\theta \\ \frac{1}{l}\tan\phi \\ 0 \end{pmatrix} \cdot c_1 + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \cdot c_2.$$

However, STEER and DRIVE do not commute (otherwise we could do all your steering at home before driving of on a trip). Their combination is given by the Lie bracket

$$[\mathbf{v}, \mathbf{u}] \equiv [\text{STEER}, \text{DRIVE}] = \frac{1}{l\cos^2\phi} \frac{\partial}{\partial\theta} \equiv \text{WRIGGLE}.$$

The operation [**v**,**u**] ≡ WRIGGLE ≡ [STEER,DRIVE] is the infinitesimal version of the sequence of transformations: steer, drive, steer back, and drive back, i.e.,

$$\{\text{STEER}, \text{DRIVE}, \text{STEER}^{-1}, \text{DRIVE}^{-1}\}.$$

Now, WRIGGLE can get us out of some parking spaces, but not tight ones: we may not have enough room to WRIGGLE out. The usual tight parking space restricts the DRIVE transformation, but not STEER. A truly tight parking space restricts STEER as well by putting your front wheels against the curb.

Fortunately, there is still another commutator available:

$$[-\mathbf{u}, [\mathbf{v}, \mathbf{u}]] \equiv [\text{DRIVE}, [\text{STEER}, \text{DRIVE}]] = [[\mathbf{u}, \mathbf{v}], \mathbf{u}] \equiv$$

$$[\text{DRIVE}, \text{WRIGGLE}] = \frac{1}{l\cos^2\phi}\left(\sin\theta \frac{\partial}{\partial x} - \cos\theta \frac{\partial}{\partial y}\right) \equiv \text{SLIDE}$$

The operation [[**u**,**v**],**u**] ≡ SLIDE ≡ [DRIVE,WRIGGLE] is a displacement at right angles to the car, and can get us out of any parking place. We just need to remember to steer, drive, steer back, drive some more, steer, drive back, steer back, and drive back:

$$\{\text{STEER}, \text{DRIVE}, \text{STEER}^{-1}, \text{DRIVE}, \text{STEER}, \text{DRIVE}^{-1}, \text{STEER}^{-1}, \text{DRIVE}^{-1}\}.$$

We have to reverse steer in the middle of the parking place. This is not intuitive, and no doubt is part of a common problem with parallel parking.

Thus, from only two controls, c_1 and c_2, we can form the vector–fields DRIVE ≡ **u**, STEER ≡ **v**, WRIGGLE ≡ [**v**,**u**], and SLIDE ≡ [[**u**,**v**],**u**], allowing us to move anywhere in the car configuration manifold $M \equiv SO(2) \times \mathbb{R}^2$. All above computations are straightforward in *Mathematica*TM[8] if we define the following three symbolic functions:

1. Jacobian matrix: JacMat[v_List, x_List] := Outer[D, v, x];
2. Lie bracket: LieBrc[u_List, v_List, x_List] := JacMat[v, x] . u - JacMat[u, x] . v;
3. Repeated Lie bracket: Adj[u_List, v_List, x_List, k_] :=
   If[k == 0, v, LieBrc[u, Adj[u, v, x, k - 1], x]];

[8] The above computations could instead be done in other available packages, such as Maple, by suitably translating the provided example code.

In case of the human crowd, we have a slightly simpler, but multiplied problem, i.e., superposition of n individual agents' motions. So, we can define the dissipative crowd kinematics as a system of n vector ODEs:

$$\dot{\mathbf{x}}^k = \mathbf{u}^k(\mathbf{x})c_1^k + \mathbf{v}^k(\mathbf{x})c_2^k, \qquad \text{where} \tag{27}$$

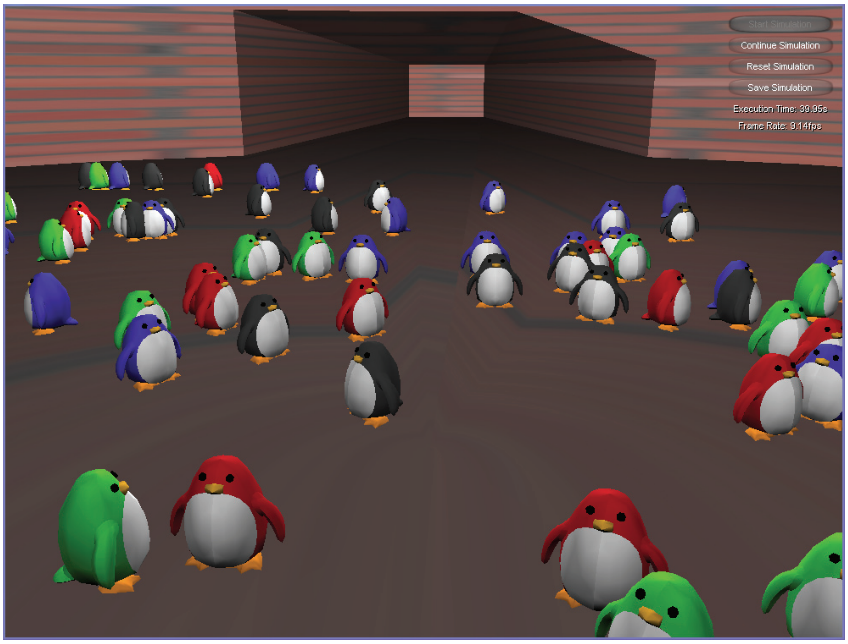
$$\mathbf{u}^k(\mathbf{x}) \equiv \mathrm{DRIVE}^k = \cos^k\theta\frac{\partial}{\partial x^k} + \sin^k\theta\frac{\partial}{\partial y^k} \equiv \begin{pmatrix} \cos^k\theta \\ \sin^k\theta \\ 0 \end{pmatrix}, \qquad \text{and}$$

$$\mathbf{v}^k(\mathbf{x}) \equiv \mathrm{STEER}^k = \frac{\partial}{\partial\theta^k} \equiv \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad \text{while } c_1^k \text{ and } c_2^k \text{ are crowd controls.}$$

Thus, the crowd kinematics (27) expands into the matrix ODE:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = \mathrm{DRIVE}^k \cdot c_1^k + \mathrm{STEER}^k \cdot c_2^k \equiv \begin{pmatrix} \cos^k\theta \\ \sin^k\theta \\ 0 \end{pmatrix} \cdot c_1^k + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot c_2^k. \tag{28}$$

A 3D simulation of random, dissipative crowd kinematics (27)–(28) of 120 penguin-like $SE(2)$–robots, developed in C++/DirX is presented in Figure 2.



Execution Time: 39.95s
Frame Rate: 9.14fps

Fig. 2. Driving and steering random $SE(2)$–dynamics of 120 penguin-like robots (with embedded collision-detection). Compare with [2].

The dissipative crowd kinematics (27)–(28) obeys the set of n-tuple integral rules of motion that are similar (though slightly simpler) to the above rules of the car kinematics, including the following derived vector-fields:

$$\mathrm{WRIGGLE}^k \equiv [\mathrm{STEER}^k,\mathrm{DRIVE}^k] \equiv [\mathbf{v}^k,\mathbf{u}^k] \text{ and}$$
$$\mathrm{SLIDE}^k \equiv [\mathrm{DRIVE}^k,\mathrm{WRIGGLE}^k] \equiv [[\mathbf{u}^k,\mathbf{v}^k],\mathbf{u}^k]$$

Thus, controlled by the two vector controls c_1^k and c_2^k, the crowd can form the vector–fields: DRIVE ≡ $\mathbf{u}^k$, STEER ≡ $\mathbf{v}^k$, WRIGGLE ≡ $[\mathbf{v}^k,\mathbf{u}^k]$, and SLIDE ≡ $[[\mathbf{u}^k,\mathbf{v}^k],\mathbf{u}^k]$, allowing it to move anywhere within its configuration manifold M given by (21). Solution of the dissipative crowd kinematics (27)–(28) defines the dissipative crowd flow, ϕ_t^K.

Now, the general $\mathcal{CD}$–$\mathcal{CK}$ crowd behavior can be defined as a amalgamate flow (behavioral Lagrangian flow, ϕ_t^L, plus dissipative kinematic flow, ϕ_t^K) on the crowd manifold M defined by (21),

$$C_t = \phi_t^L + \phi_t^K : t \mapsto (M(t), g(t)),$$

which is a one-parameter family of homeomorphic (topologically equivalent) Riemannian manifolds[9] (M, $g = g_{ij}$), parameterized by a 'time' parameter t. That is, C_t can be used for

[9] Proper differentiation of vector and tensor fields on a smooth Riemannian manifold (like the crowd $3n$–manifold M) is performed using the *Levi–Civita covariant derivative* (see, e.g., [49; 52]). Formally, let M be a Riemannian N–manifold with the tangent bundle TM and a local coordinate system $\{x^i\}_{i=1}^N$ defined in an open set $U \subset M$. The covariant derivative operator, $\nabla_X : C^\infty(TM) \to C^\infty(TM)$, is the unique linear map such that for any vector-fields X,Y,Z, constant c, and scalar function f the following properties are valid:

$$\nabla_{X+cY} = \nabla_X + c\nabla_Y, \qquad \nabla_X(Y + fZ) = \nabla_X Y + (Xf)Z + f\nabla_X Z, \qquad \nabla_X Y - \nabla_Y X = [X,Y],$$

where $[X,Y]$ is the Lie bracket of X and Y. In local coordinates, the metric g is defined for any orthonormal basis ($\partial_i = \partial/\partial x^i$) in $U \subset M$ by $g_{ij} = g(\partial_i, \partial_j) = \delta_{ij}$, $\partial_k g_{ij} = 0$. Then the affine *Levi–Civita connection* is defined on M by

$$\nabla_{\partial_i}\partial_j = \Gamma_{ij}^k \partial_k, \quad \text{where} \quad \Gamma_{ij}^k = \frac{1}{2} g^{kl}\left(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}\right) \text{are the Christoffel symbols.}$$

Now, using the covariant derivative operator ∇_X we can define the *Riemann curvature* (3,1)–tensor $\mathfrak{Rm}$ by

$$\mathfrak{Rm}(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{X,Y]} Z,$$

which measures the curvature of the manifold by expressing how noncommutative covariant differentiation is. The (3,1)–components R_{ijk}^l of $\mathfrak{Rm}$ are defined in $U \subset M$ by

$$\mathfrak{Rm}\left(\partial_i,\partial_j\right)\partial_k = R_{ijk}^l \partial_l, \quad \text{or} \quad R_{ijk}^l = \partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \Gamma_{jk}^m \Gamma_{im}^l - \Gamma_{ik}^m \Gamma_{jm}^l.$$

Also, the Riemann (4,0)–tensor $R_{ijk}^l = g_{lm} R_{ijk}^m$ is defined as the g–based inner product on M,

$$R_{ijkl} = \left\langle \mathfrak{Rm}\left(\partial_i,\partial_j\right)\partial_k, \partial_l \right\rangle.$$

The first and second Bianchi identities for the Riemann (4,0)–tensor R_{ijkl} hold, describing smooth deformations of the crowd manifold M over time. The manifold family $(M(t), g(t))$ at time t determines the manifold family $(M(t + dt), g(t + dt))$ at an infinitesimal time $t + dt$ into the future, according to some presecribed geometric flow, like the celebrated *Ricci flow* [30; 31; 32; 33] (that was an instrument for a proof of a 100–year old Poincaré conjecture),

$$\partial_t g_{ij}(t) = -2R_{ij}(t), \tag{29}$$

where R_{ij} is the Ricci curvature tensor (see Appendix) of the crowd manifold M and $\partial_t g(t)$ is defined as

$$\partial_t g(t) \equiv \frac{d}{dt}g(t) := \lim_{dt\to 0}\frac{g(t+dt)-g(t)}{dt}. \tag{30}$$

3.4 Aggregate behavioral–compositional dynamics ($\mathcal{AD}$)

To formally develop the meso-level aggregate behavioral–compositional dynamics ($\mathcal{AD}$), we start with the crowd path integral (24), which can be redefined if we Wick–rotate the time variable t to imaginary values, $t \to \tau = \mathrm{i}t$, thereby transforming the Lorentzian path integral in real time into the Euclidean path integral in imaginary time. Furthermore, if we rectify the time axis back to the real line, we get the adaptive SFT–partition function as our proposed $\mathcal{AD}$ –model:

$$\langle \text{Physical Action}|\text{Mental Preparation}\rangle_{\text{AD}} = \int_{CD}\mathcal{D}[w,x,y]\mathrm{e}^{-A[x,y;t_i,t_j]}. \tag{31}$$

The adaptive $\mathcal{AD}$ –transition amplitude ⟨Physical Action | Mental Preparation⟩$_{\text{AD}}$ as defined by the SFT–partition function (31) is a general model of a *Markov stochastic process*. Recall that Markov process is a random process characterized by a *lack of memory*, i.e., the statistical properties of the immediate future are uniquely determined by the present, regardless of the past (see, e.g. [23; 49]). The N–dimensional Markov process can be defined by the Ito stochastic differential equation,

$$dx_i(t) = A_i[x^i(t),t]dt + B_{ij}[x^i(t),t]dW^j(t), \tag{32}$$

$$R_{ijkl} + R_{jkil} + R_{kijl} = 0, \qquad \nabla_i R_{jklm} + \nabla_j R_{kilm} + \nabla_k R_{ijlm} = 0,$$

while the twice contracted second Bianchi identity reads: $2\nabla_j R_{ij} = \nabla_i R$.
The (0,2) *Ricci tensor* $\mathfrak{Rc}$ is the trace of the Riemann (3,1) –tensor $\mathfrak{Rm}$,

$$\mathfrak{Rc}(Y,Z) + \mathrm{tr}(X \to \mathfrak{Rm}(X,Y)Z), \quad \text{so that} \quad \mathfrak{Rc}(X,Y) = g(\mathfrak{Rm}(\partial_i, X)\partial_i, Y),$$

Its components $R_{jk} = \mathfrak{Rc}(\partial_j, \partial_k)$are given in $U \subset M$ by the contraction

$$R_{jk} = R^i_{ijk}, \quad \text{or} \quad R_{jk} = \partial_i \Gamma^i_{jk} - \partial_k \Gamma^i_{ji} + \Gamma^i_{mi}\Gamma^m_{jk} - \Gamma^i_{mk}\Gamma^m_{ji}.$$

Finally, the scalar curvature R is the trace of the Ricci tensor $\mathfrak{Rc}$, given in $U \subset M$ by: $R = g^{ij}R_{ij}$.

$$x^i(0) = x_{i0}, \qquad (i, j = 1, \dots, N) \tag{33}$$

or corresponding *Ito stochastic integral equation*

$$x^i(t) = x^i(0) + \int_0^t ds A_i[x^i(s), s] + \int_0^t dW^j(s) B_{ij}[x^i(s), s], \tag{34}$$

in which $x^i(t)$ is the variable of interest, the vector $A_i[x(t), t]$ denotes deterministic *drift*, the matrix $B_{ij}[x(t), t]$ represents continuous stochastic *diffusion fluctuations*, and $W^j(t)$ is an N–variable *Wiener process* (i.e., generalized Brownian motion [23]) and

$$dW^j(t) = W^j(t + dt) - W^j(t).$$

The two Ito equations (33)–(34) are equivalent to the general *Chapman–Kolmogorov probability equation* (see equation (35) below). There are three well known special cases of the Chapman– Kolmogorov equation (see [23]):

1. When both $B_{ij}[x(t), t]$ and $W(t)$ are zero, i.e., in the case of pure deterministic motion, it reduces to the *Liouville equation*

$$\partial_t P(x', t' \,|\, x'', t'') = -\sum_i \frac{\partial}{\partial x^i}\left\{A_i[x(t), t] P(x', t' \,|\, x'', t'')\right\}.$$

2. When only $W(t)$ is zero, it reduces to the *Fokker–Planck equation*

$$\partial_t P(x', t' \,|\, x'', t'') = -\sum_i \frac{\partial}{\partial x^i}\left\{A_i[x(t), t] P(x', t' \,|\, x'', t'')\right\}$$

$$+\frac{1}{2}\sum_{ij} \frac{\partial^2}{\partial x^i \partial x^j}\left\{B_{ij}[x(t), t] P(x', t' \,|\, x'', t'')\right\}.$$

3. When both $A_i\,[x(t), t]$ and $B_{ij}[x(t), t)$ are zero, i.e., the state–space consists of integers only, it reduces to the *Master equation* of discontinuous jumps

$$\partial_t P(x', t' \,|\, x'', t'') = \int dx W(x' \,|\, x'', t) P(x', t' \,|\, x'', t'') - \int dx W(x'' \,|\, x', t) P(x', t' \,|\, x'', t'').$$

The *Markov assumption* can now be formulated in terms of the conditional probabilities $P(x^i, t_i)$: if the times t_i increase from right to left, the conditional probability is determined entirely by the knowledge of the most recent condition. Markov process is generated by a set of conditional probabilities whose probability–density $P = P(x', t' \,|\, x'', t'')$ evolution obeys the general *Chapman–Kolmogorov integro–differential equation*

$$\partial_t P = -\sum_i \frac{\partial}{\partial x^i}\left\{A_i[x(t), t] P\right\} + \frac{1}{2}\sum_{ij} \frac{\partial^2}{\partial x^i \partial x^j}\left\{B_{ij}[x(t), t] P\right\}$$

$$+\int dx \left\{W(x' \,|\, x'', t) P - W(x'' \,|\, x', t) P\right\}$$

including *deterministic drift*, *diffusion fluctuations* and *discontinuous jumps* (given respectively in the first, second and third terms on the r.h.s.). This general Chapman–Kolmogorov integro-differential equation (35), with its conditional probability density evolution, $P = P(x', t' \,|\, x'', t'')$, is represented by our SFT–partition function (31).

Furthermore, discretization of the adaptive SFT–partition function (31) gives the standard *partition function* (see Appendix)

$$Z = \sum_j \mathrm{e}^{-w_j E^j / T}, \tag{35}$$

where E^j is the motion energy eigenvalue (reflecting each possible motivational energetic state), T is the temperature–like environmental control parameter, and the sum runs over all ID energy eigenstates (labelled by the index j). From (35), we can calculate the *transition entropy*, as $S = k_B \ln Z$ (see the next section).

4. Entropy, chaos and phase transitions in the crowd manifold

Recall that nonequilibrium phase transitions [25; 26; 27; 28; 29] are phenomena which bring about qualitative physical changes at the macroscopic level in presence of the same microscopic forces acting among the constituents of a system. In this section we extend the $\mathcal{CD}$ formalism to incorporate both algorithmic and geometrical entropy as well as dynamical chaos [50; 58; 60] between the entropy–growing phase of Mental Preparation and the entropy– conserving phase of Physical Action, together with the associated topological phase transitions.

4.1 Algorithmic entropy

The Boltzmann and Shannon (hence also Gibbs entropy, which is Shannon entropy scaled by k ln 2, where k is the Bolzmann constant) entropy definitions involve the notion of *ensembles*. Membership of microscopic states in ensembles defines the probability density function that underpins the entropy function; the result is that the entropy of a definite and completely known microscopic state is precisely zero. Bolzmann entropy defines the probabilistic model of the system by effectively discarding part of the information about the system, while the Shannon entropy is concerned with measuring the ignorance of the observer – the amount of missing information – about the system.

Zurek proposed a new physical entropy measure that can be applied to individual microscopic system states and does not use the ensemble structure. This is based on the notion of a fixed individually random object provided by Algorithmic Information Theory and Kolmogorov Complexity: put simply, the randomness $K(x)$ of a binary string x is the length in terms of number of bits of the smallest program p on a universal computer that can produce x.

While this is the basic idea, there are some important technical details involved with this definition. The randomness definition uses the prefix complexity $K(.)$ rather than the older Kolmogorov complexity measure $C(.)$: the prefix complexity $K(x\,|\,y)$ of x given y is the Kolmogorov complexity $C_{\phi_u}(x\,|\,y)= \min\{p\,|\,x= \phi_u(\langle y,\ p\rangle)\}$ (with the convention that $C_{\phi_u}(x\,|\,y)= \infty$ if there is no such p) that is taken with respect to a reference universal partial recursive function ϕ_u that is a universal prefix function. Then the prefix complexity $K(x)$ of x is just $K(x\,|\,\varepsilon)$ where ε is the empty string. A partial recursive prefix function $\phi : M \to \mathbb{N}$ is a partial recursive function such that if $\phi(p) < \infty$ and $\phi(q) < \infty$ then p is not a proper prefix of q: that is, we restrict the complexity definition to a set of strings (which are descriptions of effective procedures) such that none is a proper prefix of any other. In this way, all effective procedure descriptions are *self-delimiting*: the total length of the description is given within the description itself. A universal prefix function ϕ_u is a prefix function such that $\forall n \in \mathbb{N}$ ϕ_u $(\langle y, \langle n, p\rangle\rangle = \phi_n(\langle y, p\rangle$, where ϕ_n is numbered n according to some Godel numbering of the partial recursive functions; that is, a universal prefix function is a partial recursive function that simulates any partial recursive function. Here, $\langle x,y\rangle$ stands for a total recusive one-one mapping from $\mathbb{N}\times\mathbb{N}$ into $\mathbb{N}$, $\langle x_1, x_2, \ldots, x_n\rangle = \langle x_1, \langle x_2, \ldots, x_n\rangle\rangle$,$\mathbb{N}$ is the set of natural numbers, and $M = \{0,1\}^*$ is the set of all binary strings.

This notion of entropy circumvents the use of probability to give a concept of entropy that can be applied to a fully specified macroscopic state: the algorithmic randomness of the state is the length of the shortest possible effective description of it. To illustrate, suppose for the moment that the set of microscopic states is countably infinite, with each state identified with some natural number. It is known that the discrete version of the Gibbs entropy (and hence of Shannon's entropy) and the algorithmic entropy are asymptotically consistent under mild assumptions. Consider a system with a countably infinite set of microscopic states X supporting a probability density function $P(.)$ so that $P(x)$ is the probability that the system is in microscopic state $x \in X$. Then the Gibbs entropy is $S_G(P) = -(k\ln 2)\sum_{x\in X} P(x)\log P(x)$ (which is Shannon's information-theoretic entropy $H(P)$ scaled by k ln 2). Supposing that $P(.)$ is recursive, then $S_G(P) = (k\ln 2)\sum_{x\in X} P(x)K(x) + C$, where C_ϕ is a constant depending only on the choice of the reference universal prefix function ϕ. Hence, as a measure of entropy, the function $K(.)$ manifests the same kind of behavior as Shannon's and Gibbs entropy measures.

Zurek's proposal was of a new physical entropy measure that includes contributions from both the randomness of a state and ignorance about it. Assume now that we have determined the macroscopic parameters of the system, and encode this as a string - which can always be converted into an equivalent binary string, which is just a natural number under a standard encoding. It is standard to denote the binary string and its corresponding natural number interchangeably; here let x be the encoded macroscopic parameters. Zurek's definition of *algorithmic entropy* of the macroscopic state is then $K(x) + H_x$, where $H_x = S_B(x)/(k \ln 2)$, where $S_B(x)$ is the Bolzmann entropy of the system constrained by x and k is Bolzmann's constant; the physical version of the algorithmic entropy is therefore defined as $S_A(x) = (k \ln 2)(K(x) + H_x)$. Here H_x represents the level of ignorance about the microscopic state, given the parameter set x; it can decrease towards zero as knowledge about the state of the system increases, at which point the algorithmic entropy reduces to the Bolzmann entropy.

4.2 Ricci flow and Perelman entropy–action on the crowd manifold

Recall that the inertial metric crowd flow, $C_t : t \to (M(t), g(t))$ on the crowd $3n$–mani-fold (21) is a one-parameter family of homeomorphic Riemannian manifolds (M, g), evolving by the Ricci flow (29)–(30).

Now, given a smooth scalar function $u : M \to \mathbb{R}$ on the Riemannian crowd $3n$–manifold M, its Laplacian operator Δ is locally defined as

$$\Delta u = g^{ij}\nabla_i\nabla_j u,$$

where ∇_i is the covariant derivative (or, Levi–Civita connection, see Appendix). We say that a smooth function $u : M\times [0,T) \to \mathbb{R}$, where $T \in (0,\infty]$, is a solution to the heat equation (see Appendix, eq. (60)) on M if

$$\partial_t u = \Delta u. \tag{36}$$

One of the most important properties satisfied by the heat equation is the maximum principle, which says that for any smooth solution to the heat equation, whatever point-wise bounds hold at $t = 0$ also hold for $t > 0$ [13]. This property exhibits the smoothing behavior of the heat diffusion (36) on M.

Closely related to the heat diffusion (36) is the (the Fields medal winning) Perelman entropy–action functional, which is on a $3n$–manifold M with a Riemannian metric g_{ij} and a (temperature-like) scalar function f given by [75]

$$\mathcal{E} = \int_M (R + |\nabla f|^2) \mathrm{e}^{-f} d\mu \tag{37}$$

where R is the scalar Riemann curvature on M, while $d\mu$ is the volume $3n$–form on M, defined as

$$d\mu = \sqrt{\det(g_{ij})}\, dx^1 \wedge dx^2 \wedge \ldots \wedge dx^{3n}. \tag{38}$$

During the Ricci flow (29)–(30) on the crowd manifold (21), that is, during the inertial metric crowd flow, $C_t : t \to (M(t), g(t))$, the Perelman entropy functional (37) evolves as

$$\partial_t \mathcal{E} = 2\int | R_{ij} + \nabla_i \nabla_j f |^2 \mathrm{e}^{-f} d\mu. \tag{39}$$

Now, the *crowd breathers* are solitonic crowd behaviors, which could be given by localized periodic solutions of some nonlinear soliton PDEs, including the exactly solvable sine–Gordon equation and the focusing nonlinear Schrödinger equation. In particular, the time-dependent crowd inertial metric $g_{ij}(t)$, evolving by the Ricci flow $g(t)$ given by (29)–(30) on the crowd $3n$–manifold M is the *Ricci crowd breather*, if for some $t_1 < t_2$ and $\alpha > 0$ the metrics $\alpha g_{ij}(t_1)$ and $g_{ij}(t_2)$ differ only by a diffeomorphism; the cases $\alpha = 1$, $\alpha < 1$, $\alpha > 1$ correspond to steady, shrinking and expanding crowd breathers, respectively. Trivial crowd breathers, for which the metrics $g_{ij}(t_1)$ and $g_{ij}(t_2)$ on M differ only by diffeomorphism and scaling for each pair of t_1 and t_2, are the *crowd Ricci solitons*. Thus, if we consider the Ricci flow (29)–(30) as a biodynamical system on the space of Riemannian metrics modulo diffeomorphism and scaling, then crowd breathers and solitons correspond to periodic orbits and fixed points respectively. At each time the Ricci soliton metric satisfies on M an equation of the form [75]

$$R_{ij} + c g_{ij} + \nabla_i b_j + \nabla_j b_i = 0,$$

where c is a number and b_i is a 1–form; in particular, when $b_i = \frac{1}{2}\nabla_i a$ for some function a on M, we get a gradient Ricci soliton.

Define $\lambda(g_{ij}) = \inf \mathcal{E}(g_{ij}, f)$, where infimum is taken over all smooth f, satisfying

$$\int_M \mathrm{e}^{-f} d\mu = 1. \tag{40}$$

$\lambda(g_{ij})$ is the lowest eigenvalue of the operator $-4\Delta + R$. Then the entropy evolution formula (39) implies that $\lambda(g_{ij}(t))$ is non-decreasing in t, and moreover, if $\lambda(t_1) = \lambda(t_2)$, then for $t \in [t_1, t_2]$ we have $R_{ij} + \nabla_i\nabla_j f = 0$ for f which minimizes $\mathcal{E}$ on M [75]. Therefore, a steady breather on M is necessarily a steady soliton.

If we define the conjugate heat operator on M as

$$\Box^* = -\partial / \partial t - \Delta + R$$

then we have the conjugate heat equation: $\Box^* u = 0$.

The entropy functional (37) is nondecreasing under the coupled Ricci–diffusion flow on M [56]

$$\partial_t g_{ij} = -2R_{ij}, \qquad \partial_t u = -\Delta u + \frac{R}{2} u - \frac{|\nabla u|^2}{u}, \tag{41}$$

where the second equation ensures $\int_M u^2 d\mu = 1$, to be preserved by the Ricci flow $g(t)$ on M.

If we define $u = \mathrm{e}^{-\frac{f}{2}}$, then (41) is equivalent to f–evolution equation on M (the nonlinear backward heat equation),

$$\partial_t f = -\Delta f + |\nabla f|^2 - R,$$

which instead preserves (40). The coupled Ricci–diffusion flow (41) is the most general biodynamic model of the crowd reaction–diffusion processes on M. In a recent study [1] this general model has been implemented for modelling a generic perception–action cycle with applications to robot navigation in the form of a dynamical grid.

Perelman's functional $\mathcal{E}$ is analogous to negative thermodynamic entropy [75]. Recall (see Appendix) that thermodynamic partition function for a generic canonical ensemble at temperature β^{-1} is given by

$$Z = \int \mathrm{e}^{-\beta E} d\omega(E), \tag{42}$$

where $\omega(E)$ is a 'density measure', which does not depend on β. From it, the average energy is given by $\langle \mathrm{E} \rangle = -\partial_\beta \ln Z$, the entropy is $S = \beta \langle E \rangle + \ln Z$, and the fluctuation is $\sigma = \langle (E - \langle E \rangle)^2 \rangle = \partial_{\beta^2} \ln Z$.

If we now fix a closed $3n$–manifold M with a probability measure m and a metric $g_{ij}(\tau)$ that depends on the temperature τ, then according to equation

$$\partial_\tau g_{ij} = 2(R_{ij} + \nabla_i \nabla_j f),$$

the partition function (42) is given by

$$\ln Z = \int (-f + \frac{n}{2}) dm. \tag{43}$$

From (43) we get (see [75])

$$\langle E \rangle = -\tau^2 \int_M (R + |\nabla f|^2 - \frac{n}{2\tau}) dm, \qquad S = -\int_M (\tau(R + |\nabla f|^2) + f - n) dm,$$

$$\sigma = 2\tau^4 \int_M |R_{ij} + \nabla_i \nabla_j f - \frac{1}{2\tau} g_{ij}|^2 dm, \qquad \text{where } dm = u dV, \; u = (4\pi\tau)^{-\frac{n}{2}} \mathrm{e}^{-f}.$$

From the above formulas, we see that the fluctuation σ is nonnegative; it vanishes only on a gradient shrinking soliton. $\langle E\rangle$ is nonnegative as well, whenever the flow exists for all sufficiently small $\tau > 0$. Furthermore, if the heat function u: (a) tends to a δ–function as $\tau \to 0$, or (b) is a limit of a sequence of partial heat functions u_i, such that each u_i tends to a δ–function as $\tau \to \tau_i > 0$, and $\tau_i \to 0$, then the entropy S is also nonnegative. In case (a), all the quantities $\langle E\rangle$, S, σ tend to zero as $\tau \to 0$, while in case (b), which may be interesting if $g_{ij}(\tau)$ becomes singular at $\tau = 0$, the entropy S may tend to a positive limit.

4.3 Chaotic inter-phase in crowd dynamics induced by its Riemannian geometry change

Recall that $\mathcal{CD}$ transition map (9) is defined by the chaotic crowd phase–transition amplitude

$$\left\langle \overset{\partial_t S=0}{\text{PHYS. ACTION}} \middle| CHAOS \middle| \overset{\partial_t S>0}{\text{MENTAL PREP.}} \right\rangle := \int_M \mathcal{D}[x] \mathrm{e}^{iA[x]},$$

where we expect the inter-phase chaotic behavior (see [53]). To show that this chaotic interphase is caused by the change in Riemannian geometry of the crowd $3n$–manifold M, we will first simplify the $\mathcal{CD}$ action functional (22) as

$$A[x] = \frac{1}{2}\int_{t_{ini}}^{t_{fin}} [g_{ij}\dot{x}^i\dot{x}^j - V(x,\dot{x})]dt, \tag{44}$$

with the associated standard Hamiltonian, corresponding to the amalgamate version of (18),

$$H(p,x) = \sum_{i=1}^{N} \frac{1}{2} p_i^2 + V(x,\dot{x}), \tag{45}$$

where p_i are the SE(2)–momenta, canonically conjugate to the individual agents' SE(2)–coordinates x_i, (i = 1, ...,$3n$). Biodynamics of systems with action (44) and Hamiltonian (45) are given by the set of *geodesic equations* [49; 52]

$$\frac{d^2x^i}{ds^2} + \Gamma^i_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds} = 0, \tag{46}$$

where Γ^i_{jk} are the Christoffel symbols of the affine Levi–Civita connection of the Riemannian $\mathcal{CD}$ manifold M (see Appendix). In this geometrical framework, the instability of the trajectories is the instability of the geodesics, and it is completely determined by the curvature properties of the $\mathcal{CD}$ manifold M according to the Jacobi equation of geodesic deviation [49; 52]

$$\frac{D^2J^i}{ds^2} + R^i_{jkm}\frac{dx^j}{ds}J^k\frac{dx^m}{ds} = 0, \tag{47}$$

whose solution J, usually called Jacobi variation field, locally measures the distance between nearby geodesics; D/ds stands for the covariant derivative along a geodesic and R^i_{jkm} are the components of the Riemann curvature tensor of the $\mathcal{CD}$ manifold M.

The relevant part of the Jacobi equation (47) is given by the tangent dynamics equation [12; 15]

$$\ddot{J}^i + R^i_{0k0} J^k = 0, \qquad (i,k = 1,\ldots,3n), \tag{48}$$

where the only non-vanishing components of the curvature tensor of the $\mathcal{CD}$ manifold M are

$$R^i_{0k0} = \partial^2 V / \partial x^i \partial x^k. \tag{49}$$

The tangent dynamics equation (48) can be used to define Lyapunov exponents in dynamical systems given by the Riemannian action (44) and Hamiltonian (45), using the formula [14]

$$\lambda_1 = \lim_{t\to\infty} 1/2t \log(M^N_{i=1}[J^2_i(t) + \dot{J}^2_i(t)] / M^N_{i=1}[J^2_i(0) + \dot{J}^2_i(0)]). \tag{50}$$

Lyapunov exponents measure the strength of dynamical chaos in the crowd behavior. The sum of positive Lyapunov exponents defines the *Kolmogorov–Sinai entropy* (see Appendix).

4.4 Crowd nonequilibrium phase transitions induced by manifold topology change

Now, to relate these results to topological phase transitions within the $\mathcal{CD}$ manifold M given by (21), recall that any two high–dimensional manifolds M_v and $M_{v'}$ have the same topology if they can be continuously and differentiably deformed into one another, that is if they are diffeomorphic. Thus by topology change the 'loss of diffeomorphicity' is meant [80]. In this respect, the so–called topological theorem [21] says that non–analyticity is the 'shadow' of a more fundamental phenomenon occurring in the system's configuration manifold (in our case the $\mathcal{CD}$ manifold): a topology change within the family of equipotential hypersurfaces

$$M_v = \{(x^1,\ldots,x^{3n}) \in \mathbb{R}^{3n} \mid V(x^1,\ldots,x^{3n}) = v\},$$

where V and x^i are the microscopic interaction potential and coordinates respectively. This topological approach to PTs stems from the numerical study of the dynamical counterpart of phase transitions, and precisely from the observation of discontinuous or cuspy patterns displayed by the largest Lyapunov exponent λ_1 at the transition energy [14]. Lyapunov exponents cannot be measured in laboratory experiments, at variance with thermodynamic observables, thus, being genuine dynamical observables they are only be estimated in numerical simulations of the microscopic dynamics. If there are critical points of V in configuration space, that is points $x_c = [\bar{x}_1,\ldots,\bar{x}_{3n}]$ such that $\nabla V(x)\big|_{x=x_c} = 0$, according to the Morse Lemma [40], in the neighborhood of any critical point x_c there always exists a coordinate system $x(t) = [x^1(t), \ldots, x^{3n}(t)]$ for which [14]

$$V(x) = V(x_c) - x_1^2 - \ldots - x_k^2 + x_{k+1}^2 + \ldots + x_{3n}^2, \tag{51}$$

where k is the index of the critical point, i.e., the number of negative eigenvalues of the Hessian of the potential energy V. In the neighborhood of a critical point of the $\mathcal{CD}$–manifold M, equation (51) yields the simplified form of (49), $\partial^2 V/\partial x^i \partial x^j = \pm\delta_{ij}$, giving j unstable directions that contribute to the exponential growth of the norm of the tangent vector J.

This means that the strength of dynamical chaos within the $\mathcal{CD}$–manifold M, measured by the largest Lyapunov exponent λ_1 given by (50), is affected by the existence of critical points x_c of the potential energy $V(x)$. However, as $V(x)$ is bounded below, it is a good Morse function, with no vanishing eigenvalues of its Hessian matrix. According to Morse theory [40], the existence of critical points of V is associated with topology changes of the hypersurfaces $\{M_v\}_{v\in\mathbb{R}}$. The topology change of the $\{M_v\}_{v\in\mathbb{R}}$ at some v_c is a necessary condition for a phase transition to take place at the corresponding energy value [21]. The topology changes implied here are those described within the framework of Morse theory through 'attachment of handles' [40] to the $\mathcal{CD}$–manifold M.

In our path–integral language this means that suitable topology changes of equipotential submanifolds of the $\mathcal{CD}$–manifold M can entail thermodynamic–like phase transitions [25; 26; 27], according to the general formula:

$$\langle \text{phase out}|\text{phase in}\rangle := \int_{\text{top-ch}} \mathcal{D}[w\Phi]\mathrm{e}^{iS[\Phi]}.$$

The statistical behavior of the crowd biodynamics system with the action functional (44) and the Hamiltonian (45) is encompassed, in the canonical ensemble, by its partition function, given by the Hamiltonian path integral [52]

$$Z_{3n} = \int_{\text{top-ch}} \mathcal{D}[p]\mathcal{D}[x]\exp\{\mathrm{i}\int_t^{t'}[p_i\dot{x}^i - H(p,x)]d\tau\}, \tag{52}$$

where we have used the shorthand notation

$$\int_{\text{top-ch}} \mathcal{D}[p]\mathcal{D}[x] \equiv \int\prod_\tau \frac{dx(\tau)dp(\tau)}{2\pi}.$$

The path integral (52) can be calculated as the partition function [20],

$$Z_{3n}(\beta) = \int\prod_{i=1}^{3n} dp_i\,dx^i \mathrm{e}^{-\beta H(p,x)} = \left(\frac{\pi}{\beta}\right)^{\frac{3n}{2}} \int\prod_{i=1}^{3n} dx^i \mathrm{e}^{-\beta V(x)}$$

$$= \left(\frac{\pi}{\beta}\right)^{\frac{3n}{2}} \int_0^\infty dv\,\mathrm{e}^{-\beta v}\int_{M_v}\frac{d\sigma}{\|\nabla V\|}, \tag{53}$$

where the last term is written using the so–called co–area formula [18], and v labels the equipotential hypersurfaces M_v of the $\mathcal{CD}$ manifold M,

$$M_v = \{(x^1,\ldots,x^{3n})\in\mathbb{R}^{3n}\,|\,V(x^1,\ldots,x^{3n}) = v\}.$$

Equation (53) shows that the relevant statistical information is contained in the canonical configurational partition function

$$Z_{3n}^C = \int\prod dx^i\,V(x)\mathrm{e}^{-\beta V(x)}.$$

Note that Z_{3n}^C is decomposed, in the last term of (53), into an infinite summation of geometric integrals,

$$\int_{M_v} d\sigma/\|\nabla V\|,$$

defined on the $\{M_v\}_{v\in\mathbb{R}}$. Once the microscopic interaction potential $V(x)$ is given, the configuration space of the system is automatically foliated into the family $\{M_v\}_{v\in\mathbb{R}}$ of these equipotential hypersurfaces. Now, from standard statistical mechanical arguments we know that, at any given value of the inverse temperature β, the larger the number $3n$, the closer to $M_v \equiv M_{u_\beta}$ are the microstates that significantly contribute to the averages, computed through $Z_{3n}(\beta)$, of thermodynamic observables. The hypersurface M_{u_β} is the one associated with

$$u_\beta = (Z_{3n}^C)^{-1}\int\prod dx^i V(x)\mathrm{e}^{-\beta V(x)},$$

the average potential energy computed at a given β. Thus, at any β, if $3n$ is very large the effective support of the canonical measure shrinks very close to a single $M_v = M_{u_\beta}$. Hence, the basic origin of a phase transition lies in a suitable topology change of the $\{M_v\}$, occurring at some v_c [20]. This topology change induces the singular behavior of the thermodynamic observables at a phase transition. It is conjectured that the counterpart of a phase transition is a breaking of diffeomorphicity among the surfaces M_v, it is appropriate to choose a diffeomorphism invariant to probe if and how the topology of the M_v changes as a function of v. Fortunately, such a topological invariant exists, the Euler characteristic of the crowd manifold M, defined by [49; 52]

$$\chi(M) = \sum_{k=0}^{3n}(-1)^k b_k(M), \tag{54}$$

where the Betti numbers $b_k(M)$ are diffeomorphism invariants (b_k are the dimensions of the de Rham's cohomology groups $H^k(M;\mathbb{R})$; therefore the b_k are integers). This homological formula can be simplified by the use of the Gauss–Bonnet theorem, that relates $\mathcal{X}(M)$ with the total Gauss–Kronecker curvature K_G of the $\mathcal{CD}$–manifold M given by [52; 58]

$$\chi(M) = \int_M K_G\, d\mu, \qquad \text{where } d\mu \text{ is given by (38).}$$

5. Conclusion

Our understanding of crowd dynamics is presently limited in important ways; in particular, the lack of a geometrically *predictive* theory of crowd behavior restricts the ability for authorities to intervene appropriately, or even to recognize when such intervention is needed. This is not merely an idle theoretical investigation: given increasing population sizes and thus increasing opportunity for the formation of large congregations of people, death and injury due to trampling and crushing – even within crowds that have not formed under common malicious intent – is a growing concern among police, military and emergency services. This paper represents a contribution towards the understanding of crowd behavior for the purpose of better informing decision–makers about the dangers and likely consequences of different intervention strategies in particular circumstances.

In this chapter, we have proposed an entropic geometrical model of crowd dynamics, with dissipative kinematics, that operates across macro–, micro– and meso–levels. This proposition is motivated by the need to explain the dynamics of crowds across these levels simultaneously: we contend that only by doing this can we expect to adequately characterize the geometrical properties of crowds with respect to regimes of behavior and the changes of state that mark the boundaries between such regimes.

In pursuing this idea, we have set aside traditional assumptions with respect to the separation of mind and body. Furthermore, we have attempted to transcend the long–running debate between contagion and convergence theories of crowd behavior with our multi-layered approach: rather than representing a reduction of the whole into parts or the emergence of the whole from the parts, our approach is build on the supposition that the direction of logical implication can and does flow in both directions simultaneously. We refer to this third alternative, which effectively unifies the other two, as *behavioral composition.*

The most natural statistical descriptor is crowd entropy, which satisfies the extended second thermodynamics law applicable to open systems comprised of many components. Similarities between the configuration manifolds of individual (micro–level) and crowds (macro–level) motivate our claim that goal–directed movement operates under entropy conservation, while natural crowd dynamics operates under monotonically increasing entropy functions. Of particular interest is what happens between these distinct topological phases: the phase transition is marked by chaotic movement.

We contend that backdrop gives us a basis on which we can build a geometrically predictive model–theory of crowd behavior dynamics. This contrasts with previous approaches, which are explanatory only (explanation that is really narrative in nature). We propose an entropy formulation of crowd dynamics as a three step process involving individual and collective psycho-dynamics, and – crucially – non-equilibrium phase transitions whereby the forces operating at the microscopic level result in geometrical change at the macroscopic level. Here we have incorporated both geometrical and algorithmic notions of entropy as well as chaos in studying the topological phase transition between the entropy conservation of physical action and the entropy increase of mental preparation.

6. Appendix

6.1 Extended second law of thermodynamics

According to Boltzmann's interpretation of the second law of thermodynamics, there exists a function of the state variables, usually chosen to be the *physical entropy S* of the system that varies monotonically during the approach to the unique final state of thermodynamic equilibrium:

$$\partial_t S \geq 0 \qquad \text{(for any isolated system).} \tag{55}$$

It is usually interpreted as a *tendency to increased disorder*, i.e., an irreversible trend to maximum disorder. The above interpretation of entropy and a second law is fairly obvious for systems of *weakly interacting particles*, to which the arguments developed by Boltzmann referred.

However, according to Prigogine [70], the above interpretation of entropy and a second law is fairly obvious *only* for systems of *weakly interacting particles*, to which the arguments developed by Boltzmann referred. On the other hand, for strongly interacting systems like the crowd, the above interpretation does not apply in a straightforward manner since, we know that for such systems there exists the possibility of evolving to more ordered states through the mechanism of *phase transitions*.

Let us now turn to nonisolated systems (like a human crowd), which exchange energy/matter with the environment. The entropy variation will now be the sum of two terms. One, entropy flux, d_eS, is due to these exchanges; the other, entropy production, d_iS, is due to the phenomena going on within the system. Thus the entropy variation is

$$\partial_t S = \frac{d_i S}{dt} + \frac{d_e S}{dt}. \tag{56}$$

For an isolated system $d_eS = 0$, and (56) together with (55) reduces to $dS = d_iS \geq 0$, the usual statement of the second law. But even if the system is nonisolated, d_iS will describe those (irreversible) processes that would still go on even in the absence of the flux term d_eS. We thus require the following extended form of the second law:

$$\partial_t S \geq 0 \qquad \text{(for any nonisolated system).} \tag{57}$$

As long as d_iS is strictly positive, irreversible processes will go on continuously within the system.[10] Thus, $d_iS > 0$ is equivalent to the condition of dissipativity as time irreversibility. If, on the other hand, d_iS reduces to zero, the process will be reversible and will merely join neighboring states of equilibrium through a slow variation of the flux term d_eS.

From a computational perspective, we have a related *algorithmic entropy*. Suppose we have a universal machine capable of simulating any effective procedure (i.e., a universal machine that can compute any computable function). There are several models to choose from, classically we would use a Universal Turing Machine but for technical reasons we are more interested in Lambda–type Calculi or Combinatory Logics. Let us describe the system of interest through some encoding as a combinatorial structure (classically this would be a

[10] Among the most common irreversible processes contributing to d_iS are chemical reactions, heat conduction, diffusion, viscous dissipation, and relaxation phenomena in electrically or magnetically polarized systems. For each of these phenomena two factors can be defined: an appropriate internal *flux*, J_i, denoting essentially its rate, and a driving *force*, X_i, related to the maintenance of the nonequilibrium constraint. A most remarkable feature is that d_iS becomes a *bilinear form* of J_i and X_i. The following table summarizes the fluxes and forces associated with some commonly observed irreversible phenomena (see [48; 70])

Phenomenon	**Flux**	**Force**	**Rank**
Heat conduction	Heat flux, $\mathbf{J}_{th}$	$grad(1/T)$	Vector
Diffusion	Mass flux, $\mathbf{J}_d$	$-[grad(\mu/T) - \mathbf{F}]$	Vector
Viscous flow	Pressure tensor, $\mathbf{P}$	$(1/T)\ grad\,\mathbf{v}$	Tensor
Chemical reaction	Rate of reaction, ω	Affinity of reaction	Scalar

In general, the fluxes J_k are very complicated functions of the forces X_i. A particularly simple situation arises when their relation is linear, then we have the celebrated *Onsager relations*,

$$J_i = L_{ik} X_k, \qquad (i, k = 1, \ldots, n) \tag{58}$$

in which L_{ik} denote the set of *phenomenological coefficients*. This is what happens near equilibrium where they are also symmetric, $L_{ik} = L_{ki}$. Note, however, that certain states far from equilibrium can still be characterized by a linear dependence of the form of (58) that occurs either accidentally or because of the presence of special types of regulatory processes.

binary string, but again I prefer for technical reasons Normal Forms with respect to alpha/beta/eta, weak, strong reduction, which are basically the Lambda-type Calculi and Combinatory Logic notions roughly akin to a "computational" step). In other words, we have states of our system now represented as sentences in some language. The entropy is simply the minimum effective procedure against our computational model that generates the description of the system state. This is a universal and absolute notion of compression of our data – the entropy is the strongest compression over all possible compression schemes, in effect. Now here is the 'magic': this minimum is absolute in the sense that it does not vary (except by a constant) with respect to our reference choice of machine.

6.2 Thermodynamic partition function

Recall that the partition function Z is a quantity that encodes the statistical properties of a system in thermodynamic equilibrium. It is a function of temperature and other parameters, such as the volume enclosing a gas. Other thermodynamic variables of the system, such as the total energy, free energy, entropy, and pressure, can be expressed in terms of the partition function or its derivatives.

A canonical ensemble is a statistical ensemble representing a probability distribution of microscopic states of the system. Its probability distribution is characterized by the proportion p_i of members of the ensemble which exhibit a measurable macroscopic state i, where the proportion of microscopic states for each macroscopic state i is given by the Boltzmann distribution,

$$p_i = \tfrac{1}{Z}\mathrm{e}^{-E_i/(kT)} = \mathrm{e}^{-(E_i-A)/(kT)},$$

where E_i is the energy of state i. It can be shown that this is the distribution which is most likely, if each system in the ensemble can exchange energy with a heat bath, or alternatively with a large number of similar systems. In other words, it is the distribution which has *maximum entropy* for a given average energy $\langle E_i \rangle$.

The partition function of a *canonical ensemble* is defined as a sum $Z(\beta) = \sum_j \mathrm{e}^{-\beta E_j}$,

where $\beta = 1/(k_BT)$ is the 'inverse temperature', where T is an ordinary temperature and k_B is the Boltzmann's constant. However, as the position x^i and momentum p_i variables of an ith particle in a system can vary continuously, the set of microstates is actually uncountable. In this case, some form of *coarse-graining* procedure must be carried out, which essentially amounts to treating two mechanical states as the same microstate if the differences in their position and momentum variables are 'small enough'. The partition function then takes the form of an integral. For instance, the partition function of a gas consisting of N molecules is proportional to the $6N$-dimensional phase-space integral,

$$Z(\beta) \sim \int_{\mathbb{R}^{6N}} d^3p_i\, d^3x^i \exp[-\beta H(p_i, x^i)],$$

where $H = H(p_i, x^i)$, $(i = 1, ..., N)$ is the classical Hamiltonian (total energy) function.

More generally, the so-called *configuration integral*, as used in probability theory, information science and dynamical systems, is an abstraction of the above definition of a partition function in statistical mechanics. It is a special case of a normalizing constant in probability theory, for the Boltzmann distribution. The partition function occurs in many problems of probability theory because, in situations where there is a natural symmetry, its associated probability measure, the *Gibbs measure* (see below), which generalizes the notion of the canonical ensemble, has the *Markov property*.

Given a set of random variables X_i taking on values x_i, and purely potential Hamiltonian function $H(x_i)$, ($i = 1, ...,N$), the partition function is defined as

$$Z(\beta)=\sum_{x^i}\exp\left[-\beta H(x^i)\right]. \tag{59}$$

The function H is understood to be a real-valued function on the space of states $\{X_1,X_2 \ldots\}$ while β is a real-valued free parameter (conventionally, the inverse temperature). The sum over the x^i is understood to be a sum over all possible values that the random variable X_i may take. Thus, the sum is to be replaced by an integral when the X_i are continuous, rather than discrete. Thus, one writes

$$Z(\beta)=\int dx^i \exp\left[-\beta H(x^i)\right],$$

for the case of continuously-varying random variables X_i.

The Gibbs measure of a random variable X_i having the value x^i is defined as the probability density function

$$P(X_i=x^i)=\frac{1}{Z(\beta)}\exp\left[-\beta E(x^i)\right]=\frac{\exp\left[-\beta H(x^i)\right]}{\sum_{x^i}\exp\left[-\beta H(x^i)\right]}.$$

where $E(x^i) = H(x^i)$ is the energy of the configuration x^i. This probability, which is now properly normalized so that $0 \leq P(x^i) \leq 1$, can be interpreted as a likelihood that a specific configuration of values x^i, ($i = 1, 2, ...N$) occurs in the system. $P(x^i)$ is also closely related to Ω, the probability of a *random partial recursive function halting*.

As such, the partition function $Z(\beta)$ can be understood to provide the Gibbs measure on the space of states, which is the unique statistical distribution that maximizes the entropy for a fixed expectation value of the energy,

$$\langle H\rangle=-\frac{\partial\log(Z(\beta))}{\partial\beta}.$$

The associated entropy is given by

$$S=-\sum_{x^i}P(x^i)\ln P(x^i)=\beta\langle H\rangle+\log Z(\beta),$$

representing 'ignorance' + 'randomness'.

The principle of maximum entropy related to the expectation value of the energy $\langle H\rangle$, is a postulate about a universal feature of any probability assignment on a given set of propositions (events, hypotheses, indices, etc.). Let some testable information about a probability distribution function be given. Consider the set of all trial probability distributions which encode this information. Then the probability distribution which maximizes the information entropy is the true probability distribution, with respect to the testable information prescribed.

Applied to the crowd dynamics, the Boltzman's theorem of *equipartition of energy* states that the expectation value of the energy $\langle H \rangle$ is uniformly spread among all degrees-of-freedom of the crowd (that is, across the whole crowd manifold M).

6.3 Free energy, Landau's phase transitions and Haken's synergetics

All thermodynamic–like properties of a multi-component system like a human (or robot) crowd may be expressed in terms of its *free energy potential,* $\mathcal{F} = -k_B T \ln Z(\beta)$, and its partial derivatives. In particular, the physical entropy S of the crowd is defined as the (negative) first partial derivative of the free energy $\mathcal{F}$ with respect to the control parameter temperature T, i.e., $S = -\partial_T \mathcal{F}$, while the *specific heat capacity* C is the second derivative, $C = T\partial_T S$.

A *phase* of the crowd denotes a set of its states that have relatively uniform behavioral properties. A *crowd phase transition* represents the its transformation from one phase to another (see e.g., [48; 58]). In general, the crowd phase transitions are divided into two categories:

- The *first–order phase transitions*, or, *discontinuous phase transitions*, are those that involve a latent heat C. During such a transition, a crowd either absorbs or releases a fixed (and typically large) amount of energy. Because energy cannot be instantaneously transferred between the system and its environment, first–order crowd transitions are associated with *mixed–phase regimes* in which some parts of the crowd have completed the transition and others have not. This forms a turbulent spatioi-temporal chaotic interphase, difficult to study, because its dynamics can be violent and hard to control.
- The *second–order phase transitions* are the *continuous phase transitions*, in the entropy S is continuous, without any latent heat C. They are purely entropic crowd transitions, which are at the focus of the present study.

In Landau's theory od phase transitions (see [48; 58]), the probability density function P is exponentially related to the free energy potential $\mathcal{F}$, i.e., $P \approx e^{-\mathcal{F}(T)}$, if $\mathcal{F}$ is considered as a function of some order parameter o. Thus, the most probable order parameter is determined by the requirement $\mathcal{F} = \min$. Therefore, the most natural order parameter for the crowd dynamics would be its entropy S.

The following table gives the analogy between various systems in thermal equilibrium and the corresponding nonequilibrium systems analyzed in Haken's synergetics [25; 26; 27]:

System in thermal equilibrium	Nonequilibrium system
Free energy potential $\mathcal{F}$	Generalized potential V
Order parameters o_i	Order parameters o_i
$\dot{o}_i = -\frac{\partial \mathcal{F}}{\partial o_i}$	$\dot{o}_i = -\frac{\partial V}{\partial o_i}$
Temperature T	Control input u
Entropy S	System output y
Specific Heat c	System efficiency e

In particular, in case of human biodynamics [48; 58], natural control inputs u_i are muscular forces and torques, F_i, natural system outputs y_i are joint coordinates q^i and momenta p_i, while the system efficiencies e_i represent the changes of coordinates and momenta with changes of corresponding muscular torques for the ith active human joint, $e_i^q = \frac{\partial q^i}{\partial F_i}$, $e_i^p = \frac{\partial p_i}{\partial F_i}$.

6.4 Heat equation, Dirichlet action and gradient flow on a Riemannian manifold

The heat equation

$$\dot{u} = \Delta u, \tag{60}$$

on a compact Riemannian manifold M with static metric ($\partial_t g = 0$), where $u : [0,T] \times M \to \mathbb{R}$ is a scalar field, can be interpreted as the gradient flow for the *Dirichlet action*

$$E(u) := \frac{1}{2}\int_M |\nabla u|_g^2 \, d\mu, \tag{61}$$

using the inner product, $\langle u_1, u_2 \rangle_\mu := \int_M u_1 u_2 \, d\mu,$ associated to the volume measure $d\mu$. This can be proved if we evolve u in time at some arbitrary rate u, an application of integration by parts formula,

$$\int_M u \nabla_\alpha X^\alpha \, d\mu = -\int_M (\nabla_\alpha u) X^\alpha \, d\mu$$

(where $\mathrm{div}(X) := \nabla_\alpha X^\alpha$ is the divergence of the vector-field X^α, which validates the Stokes theorem, $\int_M \mathrm{div}(X) \, d\mu = 0$), gives

$$\partial_t E(u) = -\int_M (\Delta u)\dot{u} \, d\mu = \langle -\Delta u, \dot{u} \rangle_\mu, \tag{62}$$

from which we see that (60) is indeed the gradient flow for (62) with respect to the inner product. In particular, if u solves the heat equation (60), we see that the Dirichlet energy is decreasing in time,

$$\partial_t E(u) = -\int_M |\Delta u|^2 \, d\mu. \tag{63}$$

Thus we see that by representing the parabolic PDE (60) as a gradient flow, we automatically gain a controlled quantity of the evolution, namely the energy functional that is generating the gradient flow. This representation also strongly suggests that solutions of (60) should eventually converge to stationary points of the Dirichlet energy (61), which by (62) are harmonic functions (i.e., the functions u with $\Delta u = 0$). As an application of the gradient flow interpretation, we can assert that the only periodic (or, "breather") solutions to the heat equation (60) are the harmonic functions (which must be constant if the manifold M is compact). Indeed, if a solution u was periodic, then the monotone functional E must be constant, which by (63) implies that u is harmonic as claimed.

6.5 Lyapunov exponents and Kolmogorov–Sinai entropy

A branch of nonlinear dynamics has been developed with the aim of formalizing and quantitatively characterizing the general sensitivity to initial conditions. The *largest Lyapunov exponent* λ, together with the related *Kaplan–Yorke dimension* d_{KY} and the *Kolmogorov–Sinai entropy* h_{KS} are the three indicators for measuring the *rate of error growth* produced by a dynamical system [17; 50; 60].

The characteristic Lyapunov exponents are somehow an extension of the linear stability analysis to the case of aperiodic motions. Roughly speaking, they measure the typical rate of exponential divergence of nearby trajectories. In this sense they give information on the rate of growth of a very small error on the initial state of a system [9; 10].

Consider an nD dynamical system given by the set of ODEs of the form

$$\dot{x} = f(x), \tag{64}$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Recall that since the r.h.s of equation (64) does not depend on t explicitly, the system is called *autonomous*. We assume that f is smooth enough that the evolution is well defined for time intervals of arbitrary extension, and that the motion occurs in a bounded region R of the system phase space M. We intend to study the separation between two trajectories in M, $x(t)$ and $x'(t)$, starting from two close initial conditions, $x(0)$ and $x'(0) = x(0) + \delta x(0)$ in $R_0 \subset M$, respectively.

As long as the difference between the trajectories, $\delta x(t) = x'(t) - x(t)$, remains infinitesimal, it can be regarded as a vector, $z(t)$, in the tangent space T_xM of M. The time evolution of $z(t)$ is given by the linearized differential equations:

$$\dot{z}_i(t) = \left.\frac{\partial f_i}{\partial x_j}\right|_{x(t)} z_j(t).$$

Under rather general hypothesis, Oseledets [72] proved that for almost all initial conditions $x(0) \in R$, there exists an orthonormal basis $\{e_i\}$ in the tangent space T_xM such that, for large times,

$$z(t) = c_i e_i \exp(\lambda_i t), \tag{65}$$

where the coefficients $\{c_i\}$ depend on $z(0)$. The exponents $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ are called *characteristic Lyapunov exponents*. If the dynamical system has an ergodic invariant measure on M, the spectrum of LEs $\{\lambda_i\}$ does not depend on the initial conditions, except for a set of measure zero with respect to the natural invariant measure.

Equation (65) describes how an nD spherical region $R = S^n \subset M$, with radius ϵ centered in $x(0)$, deforms, with time, into an ellipsoid of semi–axes $\epsilon_i(t) = \epsilon \exp(\lambda_i t)$, directed along the e_i vectors. Furthermore, for a generic small perturbation $\delta x(0)$, the distance between the reference and the perturbed trajectory behaves as

$$|\delta x(t)| \sim |\delta x(0)| \exp(\lambda_1 t)\left[1 + O\left(\exp -(\lambda_1 - \lambda_2)t\right)\right].$$

If $\lambda_1 > 0$ we have a rapid (exponential) amplification of an error on the initial condition. In such a case, the system is chaotic and, unpredictable on the long times. Indeed, if the initial error amounts to $\delta_0 = |\delta x(0)|$, and we purpose to predict the states of the system with a certain tolerance Δ, then the prediction is reliable just up to a *predictability time* given by

$$T_p \sim \frac{1}{\lambda_1} \ln\left(\frac{\Delta}{\delta_0}\right).$$

This equation shows that T_p is basically determined by the *positive leading Lyapunov exponent*, since its dependence on δ_0 and Δ is logarithmically weak. Because of its preeminent role, λ_1 is often referred as 'the leading positive Lyapunov exponent', and denoted by λ.

Therefore, Lyapunov exponents are average rates of expansion or contraction along the principal axes. For the *i*th principal axis, the corresponding Lyapunov exponent is defined as

$$\lambda_i = \lim_{t\to\infty}\{(1/t)\ln[L_i(t)/L_i(0)]\}, \tag{66}$$

where $L_i(t)$ is the radius of the ellipsoid along the *i*th principal axis at time *t*.
An initial volume V_0 of the phase–space region R_0 evolves on average as

$$V(t) = V_0 e^{(\lambda_1+\lambda_2+\cdots+\lambda_{2n})t}, \tag{67}$$

and therefore the rate of change of $V(t)$ is simply

$$\dot{V}(t) = \sum_{i=1}^{2n} \lambda_i V(t).$$

In the case of a 2D phase area *A*, evolving as $A(t) = A_0 e^{(\lambda_1+\lambda_2)t}$, a *Lyapunov dimension* d_L is defined as

$$d_L = \lim_{\epsilon\to 0}\left[\frac{d(\ln(N(\epsilon)))}{d(\ln(1/\epsilon))}\right],$$

where $N(\epsilon)$ is the number of squares with sides of length ϵ required to cover $A(t)$, and *d* represents an ordinary *capacity dimension*,

$$d_c = \lim_{\epsilon\to 0}\left(\frac{\ln N}{\ln(1/\epsilon)}\right).$$

Lyapunov dimension can be extended to the case of *n*D phase–space by means of the *Kaplan–Yorke dimension* [64; 73; 89] as

$$d_{KY} = j + \frac{\lambda_1+\lambda_2+\cdots+\lambda_j}{|\lambda_{j+1}|},$$

where the λ_i are ordered (λ_1 being the largest) and *j* is the index of the smallest nonnegative Lyapunov exponent.
On the other hand, a state, initially determined with an error $\delta x(0)$, after a time enough larger than $1/\lambda$, may be found almost everywhere in the region of motion $R \in M$. In this respect, the *Kolmogorov–Sinai* (KS) *entropy*, h_{KS}, supplies a more refined information. The error on the initial state is due to the maximal resolution we use for observing the system. For simplicity, let us assume the same resolution ϵ for each degree of freedom. We build a partition of the phase space *M* with cells of volume ϵ^d, so that the state of the system at $t = t_0$ is found in a region R_0 of volume $V_0 = \epsilon^d$ around $x(t_0)$. Now we consider the trajectories starting from V_0 at t_0 and sampled at discrete times $t_j = j\ \tau$ (j =1, 2, 3, . . . , t). Since we are considering motions that evolve in a bounded region $R \subset M$, all the trajectories visit a finite number of different cells, each one identified by a symbol. In this way a unique sequence of symbols $\{s(0), s(1), s(2), \dots\}$ is associated with a given trajectory $x(t)$. In a chaotic system,

although each evolution $x(t)$ is univocally determined by $x(t_0)$, a great number of different symbolic sequences originates by the same initial cell, because of the divergence of nearby trajectories. The total number of the admissible symbolic sequences, $\tilde{N}$ (ϵ, t), increases exponentially with a rate given by the topological entropy

$$h_T = \lim_{\epsilon\to 0}\lim_{t\to\infty}\frac{1}{t}\ln \widetilde{N}(\epsilon,t).$$

However, if we consider only the number of sequences $N_{eff}(\epsilon, t) \leq \tilde{N}$ (ϵ, t) which appear with very high probability in the long time limit – those that can be numerically or experimentally detected and that are associated with the natural measure – we arrive at a more physical quantity called the Kolmogorov–Sinai (or metric) entropy, which is the key entropy notion in ergodic theory [17]:

$$h_{KS} = \lim_{\epsilon\to 0}\lim_{t\to\infty}\frac{1}{t}\ln N_{eff}(\epsilon,t) \leq h_T. \tag{68}$$

h_{KS} quantifies the long time exponential rate of growth of the number of the effective coarse-grained trajectories of a system. This suggests a link with information theory where the Shannon entropy measures the mean asymptotic growth of the number of the typical sequences – the ensemble of which has probability almost one – emitted by a source.

We may wonder what is the number of cells where, at a time $t > t_0$, the points that evolved from R_0 can be found, i.e., we wish to know how big is the coarse–grained volume $V(\epsilon, t)$, occupied by the states evolved from the volume V_0 of the region R_0, if the minimum volume we can observe is $V_{min} = \epsilon^d$. As stated above (67), we have

$$V(t) \sim V_0 \exp(t\sum_{i=1}^{d}\lambda_i).$$

However, this is true only in the limit $\epsilon\to 0$. In this (unrealistic) limit, $V(t) = V_0$ for a conservative system (where $\sum_{i=1}^{d}\lambda_i = 0$) and $V(t) < V_0$ for a dissipative system (where $\sum_{i=1}^{d}\lambda_i < 0$). As a consequence of limited resolution power, in the evolution of the volume $V_0 = \epsilon^d$ the effect of the contracting directions (associated with the negative Lyapunov exponents) is completely lost. We can experience only the effect of the expanding directions, associated with the positive Lyapunov exponents. As a consequence, in the typical case, the coarse grained volume behaves as

$$V(\epsilon,t) \sim V_0 \mathrm{e}^{(\Sigma_{\lambda_i>0}\lambda_i)t},$$

when V_0 is small enough. Since $N_{eff}(\epsilon, t) \propto V(\epsilon, t)/V_0$, one has: $h_{KS} = \sum_{\lambda_i>0}\lambda_i$. This argument can be made more rigorous with a proper mathematical definition of the metric entropy. In this case one derives the Pesin relation [17; 76]: $h_{KS} \leq \sum_{\lambda_i>0}\lambda_i$. Because of its relation with the Lyapunov exponents, or by the definition (68), it is clear that also h_{KS} is a fine-grained and global characterization of a dynamical system. $\sum_{\lambda_i>0}$

The metric entropy is an invariant characteristic quantity of a dynamical system, i.e., given two systems with invariant measures, their KS–entropies exist and they are equal iff the systems are isomorphic [7].

Finally, the *topological entropy* on the manifold M equals the supremum of the Kolmogorov-Sinai entropies,

$$h(u) = \sup\{h_{KS}(u) = h_{\mu}(u) : \mu \in P_u(M)\},$$

where u : $M \rightarrow M$ is a continuous map on M, and μ ranges over all u-invariant (Borel) probability measures on M. Dynamical systems of positive topological entropy are often considered topologically chaotic.

7. References

[1] Aidman, E., Ivancevic, V., Jennings, A. A Coupled Reaction–Diffusion Field Model for Perception–Action Cycle with Applications to Robot Navigation. *Int. J. Intel. Def. Sup. Sys.* 2008, 1(2), 93-115.

[2] Arizona State University. New Computer Model Predicts Crowd Behavior. *ScienceDaily.* 2007, May 22.

[3] Ashcraft M.H.Human Memory and Cognition (2nd ed.) Harper Collins: New York, 1994.

[4] Ashcraft, M.H. Cognition (4th ed.), Prentice Hall: New Jersey, 2005.

[5] Barendregt, H. The Lambda Calculus: Its syntax and semantics. Studies in Logic and the Foundations of Mathematics. North Holland: Amsterdam, 1984.

[6] van Benthem, J. Reflections on epistemic logic. *Logique & Analyse,* 1991, 133–134, 5 14.

[7] Billingsley, P. Ergodic theory and information. Wiley: New York, 1965.

[8] Blumer, H. Collective Behavior. In Principles of Sociology (A.M. Lee, ed.), Barnes & Noble: New York, 1951, pp 67–121.

[9] Boffetta, G., Lacorata, G., Vulpiani, A. (eds.) Introduction to chaos and diffusion. Chaos in geophysical flows. Proc. ISSAOS, 2001.

[10] Boffetta, G., Cencini, M., Falcioni, M., Vulpiani, A. Predictability: a way to characterize complexity. *Phys. Rep.* 2002, 356, 367–474.

[11] Busemeyer, J.R., Diederich A. Survey of decision field theory. *Math. Soc. Sci.* 2002, 43, 345–370.

[12] Caiani, L., Casetti, L., Clementi, C., Pettini, M. Geometry of Dynamics Lyapunov Exponents and Phase Transitions. *Phys. Rev. Lett.* 1997, 79, 4361–4364.

[13] Cao, H.D., Chow, B. Recent developments on the Ricci flow. *Bull. Amer. Math. Soc.* 1999, 36, 59–74.

[14] Casetti, L., Pettini, M., Cohen, E.G.D. Geometric Approach to Hamiltonian Dynamics and Statistical Mechanics. *Phys. Rep.* 2000, 337, 237–341.

[15] Casetti, L., Clementi, C., Pettini, M. Riemannian theory of Hamiltonian chaos and Lyapunov exponents. *Phys. Rev. E* 1996, 54, 5969.

[16] Downarowicz, T. Entropy. *Scholarpedia* 2007, 2(11), 3901.

[17] Eckmann, J.P., Ruelle, D. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* 1985, 57, 617–630.

[18] Federer, H. Geometric Measure Theory. Springer: New York, 1969.

[19] Forster, T., Logic, Induction and the Theory of Sets. London Math. Soc. Student Texts 56, Cambridge Univ. Press: Cambridge, 2003.

[20] Franzosi, R., Pettini, M., Spinelli, L. Topology and phase transitions: a paradigmatic evidence. *Phys. Rev. Lett.* 2000, 84, 2774–2777.

[21] Franzosi, R., Pettini, M. Theorem on the origin of Phase Transitions. *Phys. Rev. Lett.* 2004, 92, 060601.

[22] Freeman, W.J., Vitiello, G. Nonlinear brain dynamics as macroscopic manifestation of underlying many–body field dynamics. *Phys. Life Rev.* 2006, 3(2), 93–118.

[23] Gardiner, C.W. Handbook of Stochastic Methods for Physics Chemistry and Natural Sciences (2nd ed.). Springer, Berlin, 1985.

[24] Haken, H., Kelso, J.A.S., Bunz, H. A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* 1985, 51, 347–356.

[25] Haken, H. Synergetics: An Introduction (3rd ed.). Springer: Berlin, 1983.

[26] Haken, H. Advanced Synergetics: Instability Hierarchies of Self–Organizing Systems and Devices (3rd ed.) Springer: Berlin, 1993.

[27] Haken, H. Principles of Brain Functioning: A Synergetic Approach to Brain Activity, Behavior and Cognition, Springer: Berlin, 1996.

[28] Haken, H. Information and Self–Organization: A Macroscopic Approach to Complex Systems. Springer: Berlin, 2000.

[29] Haken, H. Brain Dynamics, Synchronization and Activity Patterns in Pulse–Codupled Neural Nets with Delays and Noise, Springer: Berlin, 2002.

[30] Hamilton, R.S. Three-manifolds with positive Ricci curvature. *J. Diff. Geom.* 1982, 17, 255– 306.

[31] Hamilton, R.S. Four-manifolds with positive curvature operator. *J. Dif. Geom.* 1986, 24, 153–179.

[32] Hamilton, R.S. The Ricci flow on surfaces. *Cont. Math.* 1988, 71, 237–261.

[33] Hamilton, R.S. The Harnack estimate for the Ricci flow. *J. Dif. Geom.* 1993, 37, 225 243.

[34] Hankin, C. An introduction to Lambda Calculi for Computer Scientists. King's College Pub. 2004.

[35] Hebb, D.O. The Organization of Behavior. Wiley: New York, 1949.

[36] Helbing, D., Molnar, P., Social force model for pedestrian dynamics. *Phys. Rev. E* 1995, 51(5), 4282–4286.

[37] Helbing, D., Farkas, I., Vicsek, T. Simulating dynamical features of escape panic. *Nature* 2000, 407, 487–490.

[38] Helbing, D., Johansson, A., Mathiesen, J., Jensen, M.H., Hansen, A. Analytical approach to continuous and intermittent bottleneck flows. *Phys. Rev. Lett.* 2006, 97, 168001.

[39] Helbing, D., Johansson, A., Zein Al-Abideen, H. The Dynamics of Crowd Disasters: An Empirical Study. *Phys. Rev. E* 2007, 75, 046109.

[40] Hirsch, M.W. Differential Topology. Springer: New York, 1976.

[41] Hong, S.L., Newell, K.M. Entropy conservation in the control of human action. *Nonl. Dyn. Psych. Life. Sci.* 2008, 12(2), 163–190.

[42] Hong, S.L., Newell, K.M. Entropy compensation in human motor adaptation. *Chaos* 2008, 18(1), 013108.

[43] Ivancevic, V., Snoswell, M. Fuzzy–stochastic functor machine for general humanoid–robot dynamics. *IEEE Trans. SMCB* 2001, 31(3), 319–330.

[44] Ivancevic, V. Symplectic Rotational Geometry in Human Biomechanics. *SIAM Rev.* 2004, 46(3), 455–474.

[45] Ivancevic, V. Beagley, N. Brain–like functor control machine for general humanoid biodynamics. *Int. J. Math. Math. Sci.* 2005, 11, 1759–1779.

[46] Ivancevic, V. Lie–Lagrangian model for realistic human bio-dynamics. *Int. J. Hum. Rob.* 2006, 3(2), 205–218.

[47] Ivancevic, V., Ivancevic, T., Human–Like Biomechanics. Springer: Dordrecht, 2006.

[48] Ivancevic, V., Ivancevic, T. Natural Biodynamics.World Scientific: Singapore, 2006.

[49] Ivancevic, V., Ivancevic, T. Geometrical Dynamics of Complex Systems: A Unified Modelling Approach to Physics Control Biomechanics Neurodynamics and Psycho–Socio– Economical Dynamics. Springer: Dordrecht, 2006.

[50] Ivancevic, V., Ivancevic, T., High–Dimensional Chaotic and Attractor Systems. Springer: Berlin, 2007.

[51] Ivancevic, V., Ivancevic, T. Computational Mind: A Complex Dynamics Perspective. Springer: Berlin, 2007.

[52] Ivancevic, V., Ivancevic, T., Applied Differential Geometry: A Modern Introduction. World Scientific: Singapore, 2007.

[53] Ivancevic, V., Aidman, E., Yen, L. Extending Feynman's Formalisms for Modelling Human Joint Action Coordination. *Int. J. Biomath.* 2008, (to appear).

[54] Ivancevic, V., Aidman, E. Life-space foam: A medium for motivational and cognitive dynamics. *Physica A* 2007, 382, 616–630.

[55] Ivancevic, V. Generalized Hamiltonian biodynamics and topology invariants of humanoid robots. *Int. J. Math. Math. Sci.* 2002, 31(9), 555–565.

[56] Ivancevic, V., Ivancevic, T. Ricci flow and bio–reaction–diffusion systems. *SIAM Rev.* 2008 (submitted).

[57] Ivancevic, V., Ivancevic, T. Neuro–Fuzzy Associative Machinery for Comprehensive Brain and Cognition Modelling. Springer: Berlin, 2007.

[58] Ivancevic, V., Ivancevic, T. Complex Nonlinearity: Chaos, Phase Transitions, Topology Change and Path Integrals. Springer: 2008.

[59] Ivancevic, V., Ivancevic, T. Quantum Leap: From Dirac and Feynman Across the Universe to Human Body and Mind.World Scientific: Singapore, 2008.

[60] Ivancevic, T., Jain, L., Pattison, J., Hariz, A. Nonlinear Dynamics and Chaos Methods in Neurodynamics and Complex Data Analysis. *Nonl. Dyn.* 2008 (Springer Online first).

[61] Izhikevich, E.M., Edelman, G.M. Large-Scale Model of Mammalian Thalamocortical Systems. *PNAS* 2008, 105, 3593–3598.

[62] Johansson, A., Helbing, D., Z. Al-Abideen, H., Al-Bosta, S. From Crowd Dynamics to Crowd Safety: A Video–Based Analysis. *Adv. Com. Sys.* 2008, 11(4), 497–527.

[63] Jung, C.J. Collected Works of C.G. Jung. Princeton Univ. Press: New Jersey, 1970.

[64] Kaplan, J.L., Yorke, J.A. Numerical Solution of a Generalized Eigenvalue Problem for Even Mapping. Peitgen, H.O.,Walther, H.O. (eds.). Functional Differential Equations and Approximations of Fixed Points, Lecture Notes in Mathematics, 730, Springer: Berlin, 1979, pp 228–256.

[65] Kelso, JAS. Dynamic Patterns: The Self Organization of Brain and Behavior. MIT Press: Cambridge, 1995.

[66] Kugler, P.N., Turvey, M.T. Information, Natural Law, and the Self–Assembly of Rhythmic Movement: Theoretical and Experimental Investigations, Erlbaum: Hillsdale, 1987.

[67] Lewin, K. Resolving Social Conflicts, and, Field Theory in Social Science. Am. Psych. Assoc.,Washington, 1997.

[68] Matlin, M.W. Cognition. (7th ed.), Wiley: New York, 2008.

[69] Nara, A., Torrens, P.M. Spatial and temporal analysis of pedestrian egress behavior and efficiency, In Association of Computing Machinery (ACM) Advances in Geographic Information Systems, Samet, H.; Shahabi, C.; Schneider, M.(Eds.) 2007, New York, ACM, 284-287.

[70] Nicolis, G., Prigogine, I. Self–Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations. Wiley: Europe, 1977.

[71] Nicolis, J.S. Dynamics of hierarchical systems: An evolutionary approach. Springer: Berlin, 1986.

[72] Oseledets, V.I. A Multiplicative Ergodic Theorem: Characteristic Lyapunov Exponents of Dynamical Systems. *Trans. Moscow Math. Soc.* 1968, 19, 197–231.

[73] Ott, E., Grebogi, C., Yorke, J.A. Controlling chaos. *Phys. Rev. Lett.* 1990, 64, 1196 1199.

[74] Penrose, R. The Emperor's New Mind. Oxford Univ. Press: Oxford, 1989.

[75] Perelman, G. The entropy formula for the Ricci flow and its geometric applications. arXiv:math.DG/0211159, 2002.

[76] Pesin, Ya.B. Lyapunov Characteristic Exponents and Smooth Ergodic Theory. *Russ. Math. Surveys* 1977, 32(4), 55–114.

[77] Pessa, E., Vitiello, G. Quantum noise, entanglement and chaos in the quantum field theory of mind/brain states. *Mind and Matter* 2003, 1, 59–79.

[78] Pessa, E., Vitiello, G. Quantum noise induced entanglement and chaos in the dissipative quantum model of brain. *Int. J. Mod. Phys.* 2004, 18B, 841–858.

[79] Pessoa, L. On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 2008, 9, 148–158.

[80] Pettini, M. Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics. Springer, New York, 2007.

[81] Reed, S.K. Cognition: Theory and Applications. (7th ed.) Wadsworth Pub. 2006.

[82] Schöner, G. Dynamical Systems Approaches to Cognition. In: Cambridge Handbook of Computational Cognitive Modeling. Cambridge Univ. Press: Cambridge, 2007.

[83] Sutton, R.S., Barto, A.G. Reinforcement Learning: An Introduction. MIT Press: Cambridge, MA, 1998.

[84] Todorov, E., Jordan, M.I. Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 2002, 5(11), 1226–1235.

[85] Tognoli, E., Lagarde, J., DeGuzman, G.C., Kelso, J.A.S. The phi complex as a neuromarker of human social coordination. *PNAS* 2007, 104(19), 8190–8195.

[86] Turner, R.H., Killian, L.M. Collective Behavior (4th ed.) Englewood Cliffs: New Jersey, 1993.

[87] Umezawa, H. Advanced field theory: micro macro and thermal concepts. Am. Inst. Phys.: New York, 1993.

[88] Willingham, D.T. Cognition: The Thinking Animal (3rd ed.) Prentice Hall: New York, 2006.

[89] Yorke, J.A., Alligood, K., Sauer, T. Chaos: An Introduction to Dynamical Systems. Springer: New York, 1996.

11

Nonlinear Dynamics and Probabilistic Behavior in Medicine: A Case Study

H. Nicolis
Unité RIMBAUD (adolescents), Service de Psychiatrie
CHU Brugman 4, place A. Van Gehuchten 1020 Bruxelles
Belgium

1. Introduction

Nonlinearity is ubiquitous in medicine and life sciences, from the molecular and cellular to the organismic and population levels, owing to the presence of a variety of interactions, feedbacks and other kinds of regulatory processes that ensure the harmonious coexistence of the multitude of simultaneously ongoing activities (Mosekilde, 1996).

Nonlinearities arising from the cooperative interactions between the subunits constituting a system in conjunction with appropriate environmental stimuli, give often rise to collective behaviors transcending the individual subunits. A striking example of such collective behavior is contagion, be it in the form of propagation of a disease, of a rumor or on a more microscopic scale of a mutation, whereby a previously unaffected unit becomes affected in its turn following an encounter with the information-carrying unit. In this chapter we will be concerned with a particularly dramatic instance of contagion arising in the context of adolescent psychiatry, namely, adolescent suicidal outbreaks.

Suicidal trends rank among the most serious disorders of adolescence. In most countries, mortality from suicide is the second or the third leading cause (depending on the surveys) of teenage deaths. The incidence of suicide attempts peaks during mid adolescence (Becker, Schmidt, 2004). It is estimated that 20% of adolescents have suicidal thoughts and among them as much as 5 to 8% have attempted to commit the act (Pommereau, 2001). Each of these suicidal acts leaves behind surviving family members, friends and acquaintances who must cope with the loss (Bridge et al, 2003).

A number of risk factors for adolescent suicidality have been identified. Among these the most important are depression and exposure to suicide, suicide attempts or suicidal thoughts by family and friends, suggesting that the adolescent can be considered at potential risk of contagion with suicidality stimulations. Here, suicide contagion refers to the link between adolescent's exposure to a suicide stimulus and subsequent rise in the frequency of suicide attempts or suicide rate and is considered most likely to occur in already suicidal adolescent and to be a time-limited risk. In this respect, it appears reasonable to view a suicidal trend as a behavioral attribute. If so, suicide contagion could be regarded as a particular manifestation of behavioral contagion whereby, much like in an infectious disease, an attitude or a mood passes from a person to the next. Jones and Jones (1994) provided statistical support of behavioral contagion in a number of situations, and the perspectives opened in their analysis constitute one of the principal motivations of the present work.

Generally speaking, if a behavior is contagious, its prevalence increases with the number of susceptible adolescents rather than the total number of individuals present. Wheeler (1970) identifies behavioral contagion by 4 criteria:

1. An observer is motivated to behave in a certain way;
2. The observer knows how to perform the behavior in question but is not performing it;
3. The observer sees a model perform the behavior;
4. The observer after observing the model performs the behavior.

The theory of contagion rests on three central concepts apart from contagion itself: susceptibility, mode of transmission and exposure. Susceptibility is necessary for contagious transmission.

One aspect of youth suicide of particular concern is the repeated reports of suicide outbreaks among young people. These outbreaks have been reported from as long ago as ancient Greece and from around the world. They have been called suicide clusters, a term that describes three or more suicides occurring within a defined space and time. The incidence of cluster suicides is highest among teenagers and young adults (Gould, 2001) and a growing concern has been that adolescents exposed to a peer's suicide may be at increased risk to engage in suicidal behavior (Brent et al, 1993 a, b). Many studies have also addressed the question of whether indirect exposure to suicide through media or Internet accounts contributes to subsequent suicide (Baume et al, 1997; Davidson et al, 1989).

The most common explanation for the above noted phenomena is that of imitation. This mechanism is consistent with reported epidemics of suicide involving unusual methods such as immolation etc…Imitation is also consistent with the short latency between publicity and the increased rate of suicide within 1 to 2 weeks. According to McKenzie et al (2005) there is indirect evidence that imitative suicide occurs among people with mental illnesses and may account for about 10% of suicides by current and recent patients.

One could argue that individuals are influenced in their suicidal thoughts mainly through their direct exposure to an actual suicidal attempt. If so, suicidal trend would be a spontaneous process occurring at a rate equal to the size of the population of concerned individuals multiplied by proportionality constant whose value depends on the exposure in question. In this context, Joiner (1999) wonders if the pernicious agent of the hypothetical contagion in suicide exists. He insists on the important role of exposure, external influence rather than contagion and suggests that the concept of imitation may be not needed. He emphasizes that the vulnerable people may become socially contagious via assortative relating and thus simultaneously susceptible to the effects of life stress. Other studies report that the predominant psychiatric sequelae observed in adolescents exposed to violent deaths are anxiety, depression and post traumatic stress disorder. It has been suggested that the degree to which the second person identifies with or feels similar to the deceased person may influence the degree to which he is affected by this exposure.

While these mechanisms are undoubtedly operating in a number of circumstances of interest, our main thesis here is that they cannot account properly for suicidal outbreaks, as they lack the necessary ingredient of feedback. The alternative we thus propose is that of cooperativity, when a population of susceptible individuals is mixed with a population of suicidal ones. The nature of the suicidal attempt is in this perspective completely different, as it now depends on the size of two subpopulations in close interaction. This double dependence calls for a nonlinear approach to the problem and opens the way to self-acceleration, abrupt transitions and other analogous behaviors concomitant to the well-established syndrome of outbreaks.

In this chapter, the propagation of suicidal trends is viewed as the result of encounters in the course of which a susceptible individual can change its mental state with some probability following interaction with a suicidal one. The encounters can be short ranged like e.g. a physical encounter in a hospital unit and a school class, or long ranged like e.g. communication though the Internet. Different contagion scenarios are explored and the main trends to be expected are identified. The results are confronted to the data available and different strategies for improving current prevention practices are suggested. Two types of methodologies are employed. In a first approach, the variability arising from the individual decision making is ignored and a mean view is adopted. This maps the problem to a problem of population growth in a medium of limited resources (here the total number of susceptible and suicidal individuals). Various growth patterns are highlighted depending on the contagion probability and the initial percentage of suicidal individuals. In a second approach, variability is incorporated by means of the technique of Monte Carlo simulation well suited to treat populations of limited size where randomness is expected to play an important role. This approach has been used with success in several problems arising in chemical kinetics, biochemistry and social insect behavior (Gillespie, 1992). A number of different evolution scenarios are explored and some unexpected effects are brought out. The novelty here is to give access to situations limited in space and time like e. g. those arising in a given hospital unit over the usual hospitalization time, as opposed for those accounted for in surveys where local and short scale trends are smeared out.

2. General setting

Let X_1, X_2 be the populations of suicidal and of susceptible individuals respectively. In order to bring out the role of nonlinearity and cooperativity in a clearcut manner, it is stipulated that during the phenomenon of interest there is no major reshuffling of the organization, entailing that the total population remains essentially constant:

$$X_1 + X_2 = N = \text{constant} \tag{1}$$

In addition to the above two types of individuals, a third type may also be present. In what follows its role is viewed as that of a buffer, in the sense that while it does not participate directly in the dynamics, it may play a role in determining the values of some of the parameters present.

A first instance (hereafter referred as case I) explored in the sequel is that of contagion arising though direct, physical encounters of type 1 and type 2 individuals, hereafter denoted as S_1 and S_2 which are schematized as follows:

$$\begin{aligned} S_1 + S_1 &\longrightarrow S_1 + S_1 \\ S_2 + S_2 &\longrightarrow S_2 + S_2 \\ S_1 + S_2 &\xrightarrow{p_1} 2S_1 \\ S_1 + S_2 &\xrightarrow{p_2} 2S_2 \end{aligned} \tag{2}$$

The first two steps correspond to the obvious idea that encounters of individuals of the same kind do not give rise to a mental transition. On the contrary, the last two steps account for imitation and thus cooperativity: upon encountering a susceptible individual, a suicidal one can either switch to the susceptible state with a certain probability p_2 or induce a suicidal trend to the susceptible partner with another probability p_1. The corresponding probabilities p_1 and p_2 are expected to fulfill the inequality $p_1>p_2$. Although there seems to be no direct statistical evidence in support of this, we argue that in the absence of medical treatment such a property reflects the well-established tendency of susceptible and suicidal individuals to evolve "uphill" in the search of increasingly dramatic experiences rather than to evacuate stress and evolve to the opposite way toward normality. We refer to Pommereau (2001) for the definition of susceptible individuals. In fact adolescents with mental dysfunction express affective immaturity, sensibility to frustrations, massive dependence to genitors, depressivity of the mood without depressive episode and tendency to acting out. These susceptible adolescents refer to the most deviant repairs including suicidality. We emphasize that p_1, p_2 are intrinsic parameters associated to individual 1-2 encounters, independent of the respective sizes of the populations X_1 and X_2. Depending on the latter, the overall process of contagion will of course become accentuated, as seen below. It should also be noticed that in writing scheme (2), we tacitly assumed that individuals of the type 1 and 2 can only exist in a single state. In a more refined analysis one could account for further differentiation within a single subpopulation, like e.g. different degrees of susceptibility in individuals of type 2. Other refinements would be to account for memory effects and for changes in the parameters N, p_1, p_2 arising for instance from medical care, environmental stimuli or population renewal. Such extensions are likely to be important on a long time scale. They are not carried out here, as our main purpose is to identify the role of nonlinearity and cooperativity in the outbreak of suicidal attempts, a phenomenon expected to be initiated in the short to intermediate time regime.

A second instance of interest (hereafter referred as case II) pertains to contagion through long range interactions. To account for this possibility, we imagine that individuals constitute the nodes of a network and the interactions between any two individuals give rise to a connection between the corresponding nodes. In the previously presented case I, only nearest neighbor nodes are connected (e.g. 1-2, 2-3 etc.). In the other extreme each node is connected to any other node (e.g. 1-2, 1-3, 1-4…, 2-3, 2-4,…, etc…). This corresponds to the longest possible range that interactions can achieve. Intermediate cases may also be envisaged. We emphasize that the model as defined above is in many respects generic. It should thus apply suitably adapted to other types of behavioral contagion beyond the suicidal one that constitutes the main focus of the present work.

We are now in the position to formulate the evolution of the subpopulations X_1 and X_2 in a quantitative manner. Two complementary points of view are adopted for this purpose, as specified below. The results to be reported depend crucially on the values of the contagion probabilities p_1 and p_2. These quantities or, more to the point, their difference p_1-p_2 determine the time scale over which the suicidal trend will spread. In view of the scarcity of relevant data, different values will be considered and the sensitivity of the results on the choices will be assessed. Another important parameter, responsible for the sharpness of contagion and for the importance of stochastic effects, is the total number N of the individuals in the group and the initial numbers $X_1(0)$ of suicidal ones. In the following a sensitivity analysis with respect to these parameters will be carried out and some robust trends will be identified. The following possibilities will be considered.

1. All individuals N-$X_1(0)$ other than the suicidal ones are likely to be affected by the contagion. This can be the case in a hospital unit or in an institution where non-suicidal patients are already subjected to psychiatric disorders.
2. Among the X_2=N-$X_1(0)$ individuals only a fraction $\gamma X_2(0)$ (γ much smaller than 1) are likely to be affected, the remaining ones being immune to any psychiatric disorders. This can correspond to a school class or to hospital unit in which the adolescent patients are treated for a completely different kind of disease.

3. Population dynamic approach: An averaged view

In this view, encompassing case I as well as case II above, it is assumed that individuals 1 and 2 are well mixed and interact at random. The strength of the interactions is proportional to the corresponding fractions $\Theta_1=X_1/N$, $\Theta_2=X_2/N$, and only encounters between 1 and 2 lead to changes in the populations of either 1 or 2. This leads us to a rate law of the form

Rate of change of 1 over a time interval
=p_1 x (frequency of 1-2 encounters) - p_2 x (frequency of 1-2 encounters)

Taking the limit of the shortest time interval over which interactions become effective one obtains the quantitative expression

$$d\,\Theta_1/\,dt = (\,p_1-p_2\,)\,\Theta_1\,\Theta_2$$

or, with eq. (1)

$$d\,\Theta_1\,/dt= p\,\Theta_1\,(1-\Theta_1) \tag{3}$$

where we set

$$p=p_1-p_2 \tag{4}$$

This equation is formally identical to the logistic equation (Pielou, 1969). It can be integrated exactly, yielding

$$\Theta_1(t)=\frac{\Theta_1(0)}{[1-\Theta_1(0)]\;e^{-pt}+\Theta_1(0)} \tag{5}$$

which is seen to depend solely on p and on the initial fraction $\Theta_1(0)$.
The two quantitatively different evolutions predicted by this equation are depicted in Fig. 1 and 2 corresponding respectively to $\Theta_1(0)$ being greater or smaller than 1/2. As can be seen, in the first case one witnesses a smooth evolution toward a contagion of the entire population, bound to occur on the time scale of

$$T_{cont} \sim 1/p \tag{6}$$

In the second case one observes on the contrary a first period of quiescence during which individuals 1 seem to have no contagion effect, followed by an explosive growth and eventual saturation. The explosion time, corresponding to the inflexion point of the Θ_1 versus t the curve of Fig. 2, can be evaluate explicitly and is given by

$$t^* = \frac{1}{p}\ln[\frac{1-\Theta_1(0)}{\Theta_1(0)}] \tag{7}$$

For Θ$\Theta_1(0)$ much smaller than unity it is therefore much longer than the contagion time associated to the case of Fig. 1. In practice, saturation and explosion may never be achieved if the corresponding times are longer than the hospitalization period. Nevertheless, the above results may provide valuable indications on the trends that may be in elaboration within the populations in interaction. They will also serve as reference for the Monte Carlo approach presented below.

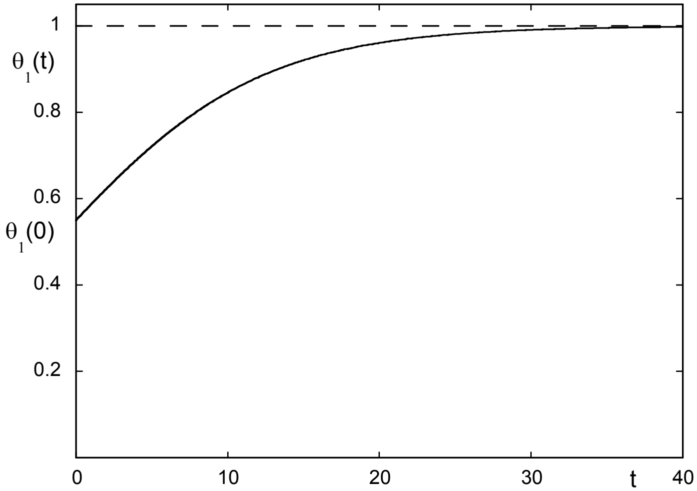


Fig. 1. Time evolution of the fraction of individuals of type 1 as deduced from eq. (5) under the condition $\Theta_1(0)>1/2$. Parameter values p=0.15, $\Theta_1(0)=0.55$.

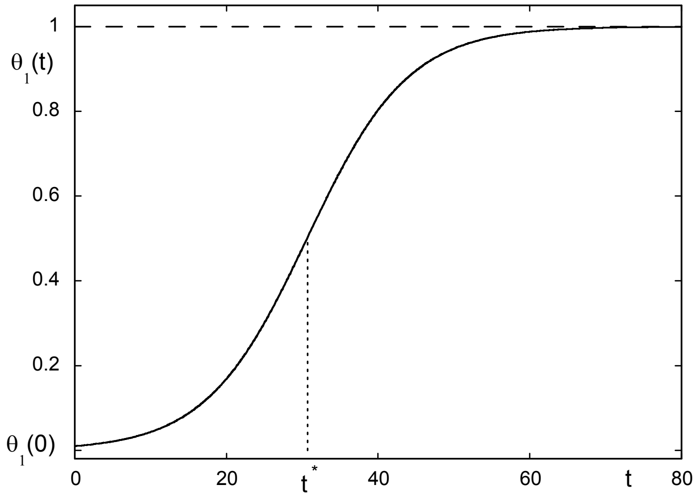


Fig. 2. As in Fig. 1 but with $\Theta_1(0)=0.01$.

4. Monte Carlo simulation

When dealing with complex realities one is often led to recognize that a modeling approach may be limited by the lack of detailed knowledge of the laws governing the system at hand and of the values of the parameters involved in the description. A central point of the present work is that to cope with this limitation it is important to set up a complementary approach aiming at a direct simulation of the underlying process, rather that at the solution of the evolution laws suggested by a certain model. The Monte Carlo simulation approach described below provides an efficient way to achieve this goal. It also allows one to incorporate in a natural way the role of individual variability expected to be of the utmost importance, since the quantities featured are now fluctuating in both space and time rather than being fully deterministic. Two types of studies have been conducted. In both cases, the population sizes have deliberately been taken to be small to emulate real world situations as they arise in a single hospital unit or in a school class. As it will turn out stochastic effects will then play a very important role. Still, the averaged description serves as a useful reference for apprehending the specific role of stochasticity in the overall process.

Case I

The physical space (school class, recreation area, hospital unit, space of common patient activities, ...) is modeled as a regular square planar lattice. Each individual performs a random walk between an initial position and its first neighbors. When two individuals are led to occupy through this process the same lattice site processes (2) are locally switched on. The various steps are weighted by the corresponding probabilities and the particular transition to be performed at a given time is decided by a random number generator (amounting essentially to throwing dice) compatible with these probabilities. After this particular step is performed the populations X_1, X_2 are updated and the process is restarted. The simulation, which records the numbers of X_1 and X_2 at different parts of space, is stopped at a number of steps beyond which the process becomes stationary in the sense of reducing to fluctuations around a constant (time-independent) plateau. In addition to a single realization of the simulation (referred as “stochastic trajectory”) averages over realization giving access to mean values, variances etc are also performed.

The following instances are considered.

i. An institution or a big hospital unit with N=30, $X_1(0)=6$ suicidal individuals and $X_2(0)=24$ individuals presenting other kinds of psychiatric disorders. The contagion probabilities are set $p_1=0.25$, $p_2=0.1$ and the individuals are initially taken to be distributed randomly.

ii. As before, but with N=20, $X_1(0)=4$ in order to test the role of population size.

iii. A school class or a mixed hospital unit with N=30, $X_1(0)=2$ suicidal individuals. It is supposed that of the $N-X_1(0)=28$ individuals 4 are susceptible of being affected and the remaining 24 ones constitute the environment within which the process will take place. Accordingly, the contagion probabilities are set to lower values $p_1=0.1$, $p_2=0.05$ since the encounters are expected to be more scarce.

iv. N=8 individuals of which $X_1(0)=4$ are suicidal and $N-X_1(0)=4$ subject to other types of disorder, functioning as a “clan” independent of its environment. This is accounted for by resetting p_1, p_2 to the values of 0.25 and 0.1 respectively.

v. As in iv. but now the two subpopulations are initially segregated (say in different hospital rooms) and meet only in common activities.

Figures 3a,b depict the time dependence of the population density X_1/N of X_1 averaged over many realizations of the process and of the associated variance $<\delta X_1^2>=<X_1^2> - <X_1>^2$. Figure 4 provides a reformulation of the results of Fig. 3 when all cases (i) to (v) are normalized to the same mean population. Figs 5 and 6a,b provide a more refined view of the role of inherent variability by showing respectively a single stochastic trajectory under the conditions of case (iii) and the probability histograms associated with (i) and (iii).

Case II

The physical space (e.g. Internet, a newsletter etc…) is here lumped into a single cell within which each individual may interact with any number of other ones with probabilities determined as before. Again, stochastic trajectories recording the individual transitions as well as averaged quantities over all trajectories are deduced. The context is now that of a small number of heavily affected individuals communicating via Internet, newsletter or any other kind of multimedia means with a small number of susceptible partners not attained so far by the disease. Fig. 7 summarizes the results for N=6, $X_1(0)=3$ using the same values for parameters p_1 and p_2 as before.

5. Discussion

Building on evidence supporting the existence of suicidal contagion, we proposed and developed a predictive model of how suicidal trends propagate in an adolescent population. The principal feature underlying the model is the cooperative character of the contagion process (last two steps in (2)). The model predictions depend entirely on two kinetic parameters, the contagion probabilities p_1 and p_2 for susceptible and for suicidal individuals to switch to the suicidal and susceptible state respectively; and on two population like parameters, the total number N of individuals that may undergo a transition in their mental state and the number $X_1(0)$ of suicidal individuals initially present.

A first result of interest has been that contagion is not always a smooth process but may rather take an explosive form, depending on the values of $X_1(0)/N$ and $p=p_1-p_2$. In this latter case there exists a well-defined time t^* of switching toward a collective suicidal state (Figs 2, 3a and 4a). This provides a quantitative basis for the phenomenon of outbreak referred in the Introduction as well as a strong support of the idea of contagion as a generic mechanism of adolescent suicidal trends. Subsequently, the population attains a mean saturation level on which is superimposed a random signal reflecting individual variability. This level may actually never be attained since on a long time scale the refinements to the original model discussed in section 2 will begin to play an increasingly crucial role.

A second series of results pertains to the role of stochasticity. The following comments are in order on inspecting the key Figure 3.

- In all cases the mean value $<X_1>$ is increasing in time, in qualitative agreement with Figs 1 and 2.
- The evolution is initially slower for segregated sub-populations (case (v)). What is happening here is that few among 1 and 2 types first meet in a limited space which constitutes a front of some sort, from which the trend can subsequently propagate.
- In cases (i), (ii), (iv) and (v), a saturation level in which the entire population of susceptible individuals switches to the suicidal state is eventually reached. The time scale for this to happen may be long with respect to the hospitalization or school period times. Still, the explosive growth for short times should be emphasized, confirming the prediction made in eq. (7) and Fig 2.

- The saturation level reached in case (iii) is significantly less than 100% in the same time scale as (i), (ii), (iv) and (v). This at first sight unexpected emergence of a state of undecidability is robust with respect to changes in the values of p_1 and p_2. It arises primarily from individual variability, here exacerbated by the smallness of the size of the overall population compared to $X_1(0)$. There are long periods of hesitation and in some realizations of the process the trend is inverted and the entire population reaches the more favorable state.

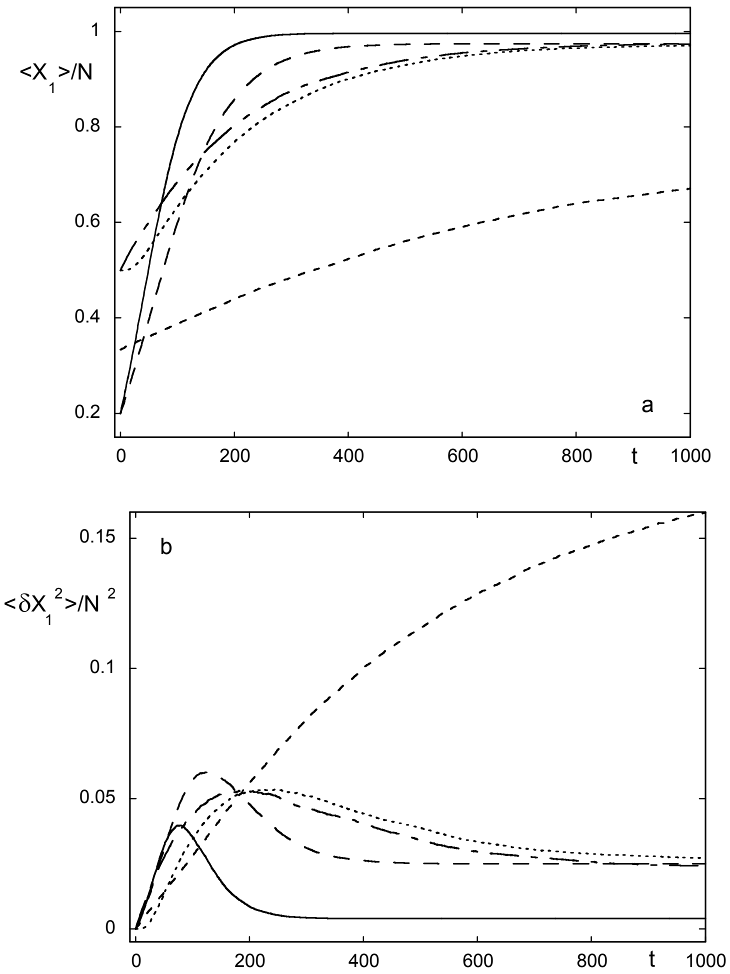


Fig. 3. (a): Time dependence of the mean density of individuals of type 1 as deduced from the Monte Carlo simulation; the full, dashed, heavy dotted, dashed-dotted and light dotted lines refer to cases (i) to (v), respectively. (b) : Time dependence of the variance under the conditions of Fig 3a. The physical space considered is a square planar lattice of size 10X10 space units, the number of statistical realizations is 10,000 and the initial positions of the populations are random in space.

These trends are further illustrated in Fig. 3b where the variance$<\delta X_1^2>=<X_1^2> - <X_1>^2$ is represented. In all case but (iii) $<\delta X_1^2>$ is seen to reach a low final value, but prior to this it goes though a well - marked maximum grossly at a time corresponding to the inflexion point of the curves in Fig. 3a. As for case (iii), $<\delta X_1^2>$ steadily increases and reaches a final value orders of magnitude larger than for (i), (ii), (iv) and (v) which is comparable to the mean value itself. This is in agreement with and provides an explanation of the statement in Jones and Jones on the behavior of variance.

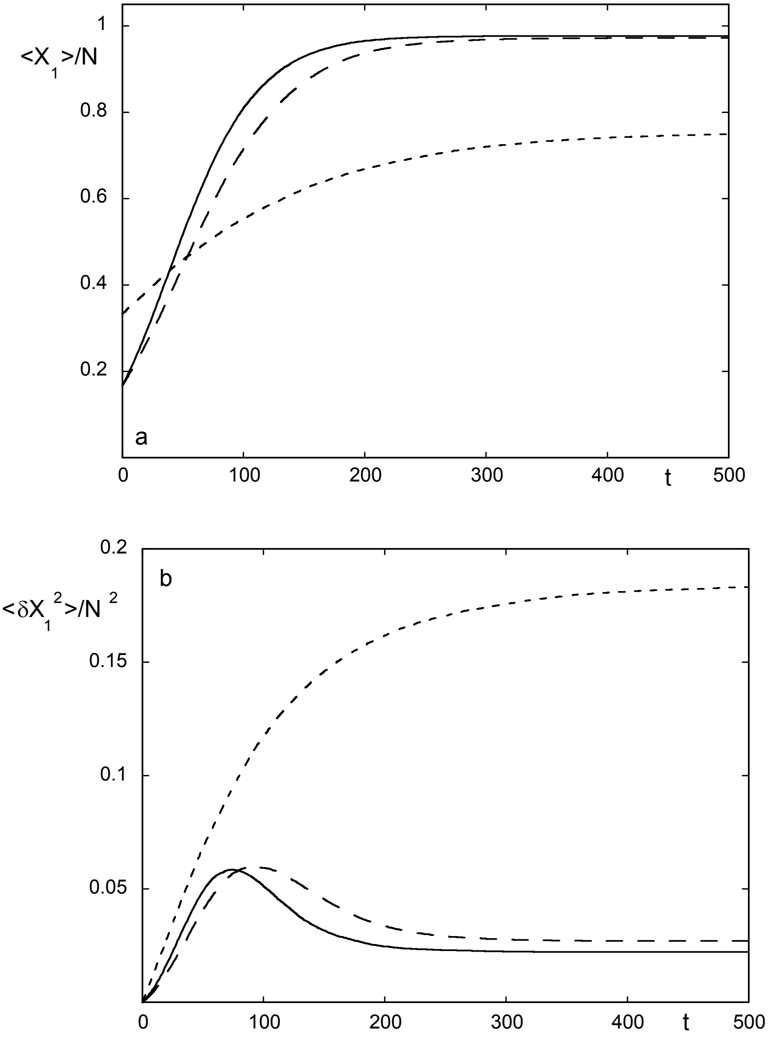


Fig. 4. (a): As in Fig. 3 but under conditions of identical overall population densities. Full, dashed and dotted lines refer to cases (i), (ii) and (iii), respectively. Initial positions and number of realizations as in Fig. 3.

Interestingly, when all cases above are normalized to the same mean population density, cases (i), (ii), (iv), and (v) are essentially reduced to a “universal" behavior both for the mean and the variance while case (iii) still constitutes a different class (Fig. 4a, 4b). This suggests that the model of eq. (3) is rather adequate for intermediate to long times as long as N is sufficiently large (which in practice could be reached already for the rather modest value of N=8), but even in these cases it may prove inadequate for short times and especially for times around the maximum of the variance.

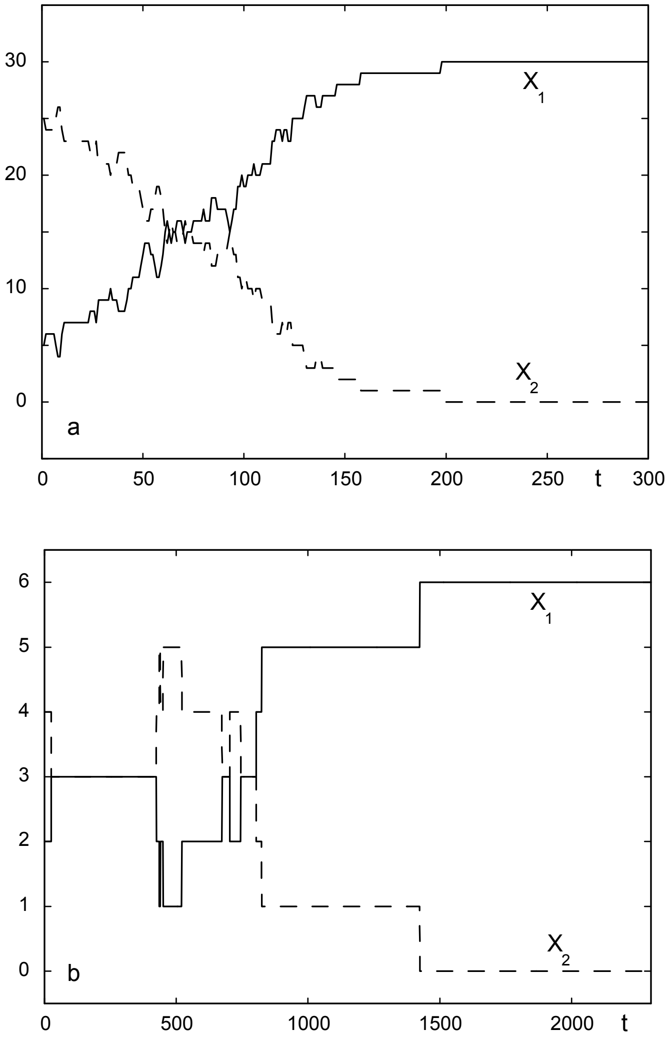


Fig. 5. (a): Quasi-deterministic behavior modulated by small scale variability under the conditions of case (i). (b): Situation of undecidability induced by the individual variability in a small size population (case (iii)).

At the level of a single stochastic realization of the process (the analog of the type of evolution observed in practice) variability and undecidability are reflected by the fact that while in case (i) the switching of the population to state 1 occurs quite early in time (Fig.5a), it needs a much longer induction time under the conditions of case (iii) (Fig. 5b). We next comment on Figs 6a,b which depict the probability histograms associated with (i) and (iii) respectively. In 6a, drawn after 80 time units (the time at which the variance reaches its maximum in Fig. 3b) the histogram is clearly unimodal. It is peaked at a value corresponding

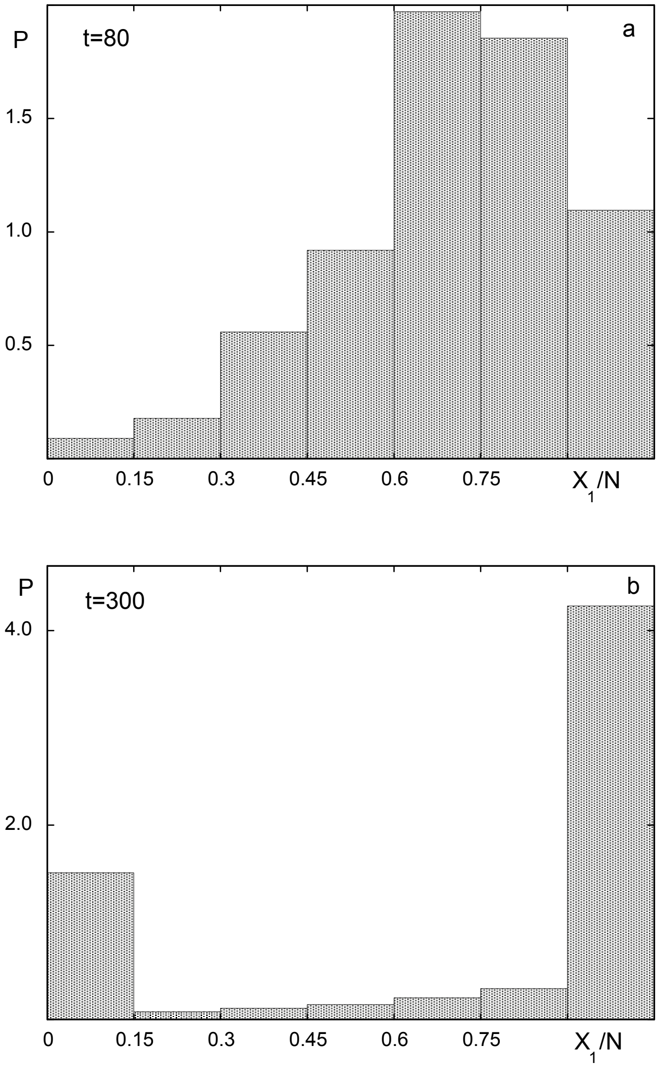


Fig. 6. Probability histograms associated with cases (i), Fig. 6a and (iii), Fig. 6b with an initial population density 0.3. Initial positions as in Fig. 3 and number of realizations is 20,000.

to the instantaneous X_1/N as deduced from Fig. 3a. For longer times the maximum slides to the right and eventually tends to 1. The structure is radically different for Fig. 6b drawn after 300 time units (the time for the value of the variance to exceed that of cases (i), (ii), (iv) and (v)) which displays a bimodal structure. As can be seen, the two peaks are located at low (close to 0) and high (close to 1) density of X_1, reflecting the possibility of switching from individuals of type 1 to type 2 with a non-negligible probability. Clearly, this type of structure is quite different from the binomial distribution usually featured when interpreting results of surveys (Jones & Jones, 1994). This reflects the cooperative character of the contagion dynamics, an idea that has been central throughout this chapter.

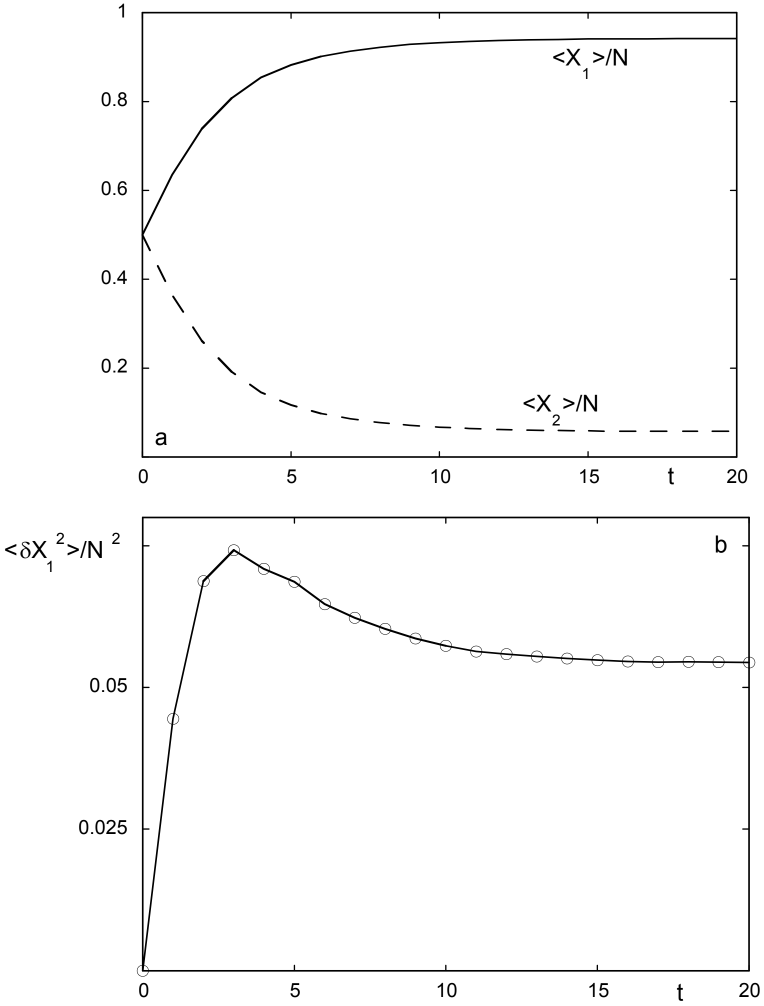


Fig. 7. Time dependence of the mean density of individuals of type 1 and 2 (7a) and of the variance of individuals of type 1 (7b) in the presence of long range interactions. Number of realizations as in Fig. 3.

The results discussed so far pertain to Case I. Regarding now the new features concerning Case II, summarized in Fig. 7, their most striking difference with Figs 3 and 4 is that the process is now accelerated dramatically, such that saturation level is reached within an observable time scale. Owing the small numbers involved this level is less than 100% in a way analogous to case (iii) above. The variance remains substantial at saturation (Fig. 7b) and goes through a maximum.

6. An augmented model

The results in the preceding sections depend crucially on the validity of the conservation condition of the total population of suicidal and susceptible individuals (eq. (1)). Although this may be a reasonable assumption over short to intermediate time scales it is bound to fail in the long run, as the system becomes open to different kinds of interactions with its environment. In this section we develop an augmented version of the model of eqs (2) accounting for key processes expected to be present in real-world situations. Specifically, we allow for the following additional steps.

- The influx of susceptible individuals S_2 from an external population A of size much larger than S_2:

$$A \xrightarrow{a} S_2 \tag{8a}$$

- The possibility that suicidal individuals may be removed from the population S_1 (recovery or on the contrary isolation):

$$S_1 \xrightarrow{k_1} S_1^* \tag{8b}$$

- The possibility that susceptible individuals may likewise be removed from the space of coexistence with S_1, spontaneously or deliberately:

$$S_2 \xrightarrow{k_2} S_2^* \tag{8c}$$

The rate equations associated to this augmented model read

$$\begin{aligned}\frac{d\Theta_1}{dt} &= p\Theta_1\Theta_2 - k_1\Theta_1 \\ \frac{d\Theta_2}{dt} &= a - p\Theta_1\Theta_2 - k_2\Theta_2\end{aligned} \tag{9}$$

Choosing as before p>0, we notice that in the limit a=0, $k_1=k_2=0$ the total population $\Theta_1+\Theta_2$ is conserved and one recovers for Θ_1 the logistic equation (3). Here we are interested in the new effects arising (a), from the opening of the susceptible population towards the influx a of freshly arriving individuals; and (b), from the process by which both suicidal and susceptible individuals tend to leave the system though the above mentioned mechanisms of medical treatment, recovery or spatial constraints.

Contrary to eq. (3), eqs (9) do not admit an explicit analytic solution. We therefore proceed by identifying first the stationary states in which the variables Θ_1 and Θ_2 no longer evolve in time. Setting d Θ_1/dt= d Θ_2/dt=0 in eqs. (9), one finds:

- A semi trivial solution

$$\Theta_1 = 0, \qquad \Theta_2 = \frac{a}{k_2} \tag{10a}$$

- A fully non-trivial solution

$$\Theta_1 = \frac{1}{k_1}(a - \frac{k_1 k_2}{p}), \qquad \Theta_2 = \frac{k_1}{p} \tag{10b}$$

To determine the conditions under which the system will eventually settle in (10a) or (10b) we perturb slightly each of these states and seek for conditions on the parameters under which the perturbations are amplified or on the contrary damped. In the first case the state - which will be qualified as unstable- will not be sustainable under real-world conditions, where perturbations of different origins are inevitable. In the second case the state –which will be qualified as stable- will represent the asymptotic regime towards which the system will evolve after a transient period whose duration depends on the values of the parameters. A standard linear stability analysis (Nicolis, 1995)) leads to the conclusion that there is a well-defined transition separating these two situations, occurring at a value of the influx parameter a given by

$$a_c = \frac{k_1 k_2}{p} \tag{11}$$

For $a<a_c$ state (10a) is the unique, stable steady state solution of eqs. (9) since state (10b) is physically unacceptable ($\Theta_1<0$). For a> a_c state (10a) still exists but is unstable, and the system evolves spontaneously towards state (10b) which becomes physically admissible as Θ_1 is now positive. Notice that in the limit a=0, $k_1=k_2=0$, p>0 the semi-trivial state is always unstable and the non-trivial one is always stable. This corresponds, in fact, to the situation depicted in Figs 1 and 2 pertaining to the model of eq. (3).

Figures 8a,b summarize the time evolution of the fractions of Θ_1 and Θ_2 prior to the steady state, under the condition $a>a_c$ (state (10b) is stable). We start with a sizable pool of susceptible individuals in which a small fraction of suicidal ones has been introduced. The evolution of Θ_1 follows first a course quite similar to that of Fig 2, but once near the plateau the situation changes radically: owing to the increasing effect of suicidal contagion, the pool of susceptibles tends to be depleted and this in turn induces a sharp decrease of suicidal incidents. The result is the appearance of a marked overshoot in the population of Θ_1 and a concomitant undershoot in Θ_2. Subsequently both Θ_1 and Θ_2 experience a slight undershoot and overshoot respectively, before settling to their long terms values. We have here a second manifestation of suicidal outbreak beyond the one identified for the model of eq. (3), where outbreak was associated with the occurrence of an inflexion point of the function $\Theta_1(t)$ prior to the attainment of the plateau (eq. (7)).

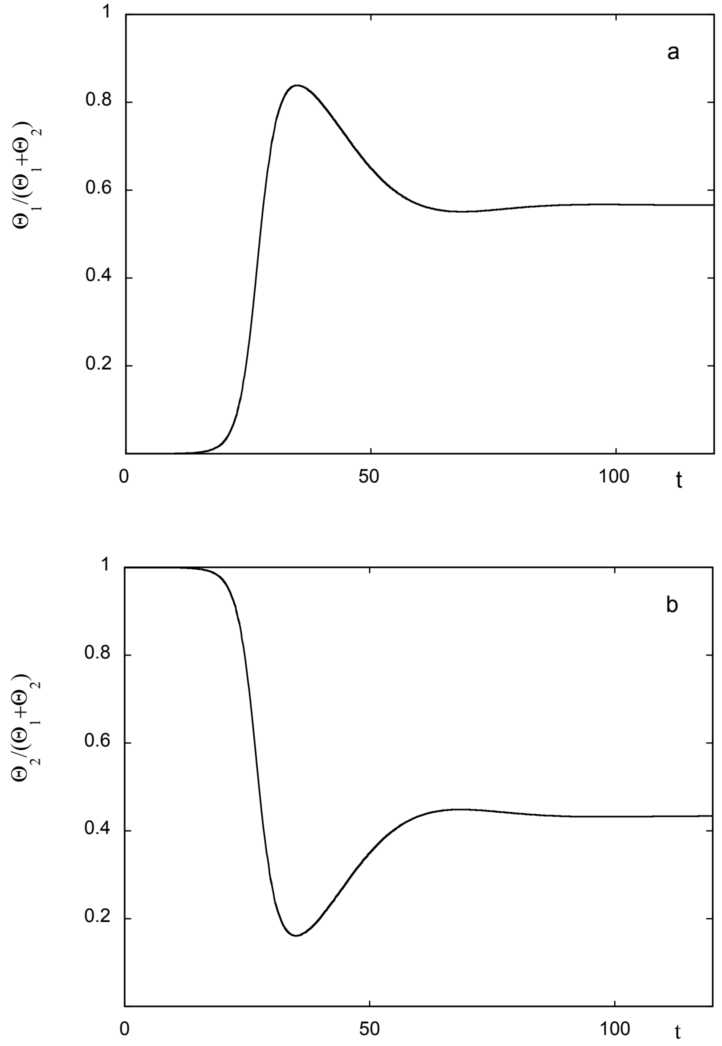


Fig. 8. Transient evolutions of the fractions of Θ_1 (a) and Θ_2 (b) obtained by solving numerically eqs. (9). Parameter values a=1, k_1=0.12, k_2=0.01, k=0.02 and initial conditions equal to 0.001 and 0.999, respectively.

Following the logic of the Monte Carlo analysis previously carried out for the scheme of eqs (2), we now inquire on the effect of variability in the results derived so far in this section. Rather than perform a full scale Monte Carlo study, we resort to a more phenomenological approach in which variability is accounted for by adding to the right hand sides of both eqs (9) uncorrelated random noises sampled from a Gaussian distribution. Fig 9 depicts the response of Θ_1 to a variability source of this kind. Keeping parameters values as in Fig. 8 we see that variability tends to depress the extent of suicidal outbreak, presumably by desynchronizing the action of the suicidal individuals that would otherwise have manifested itself in a concerted fashion.

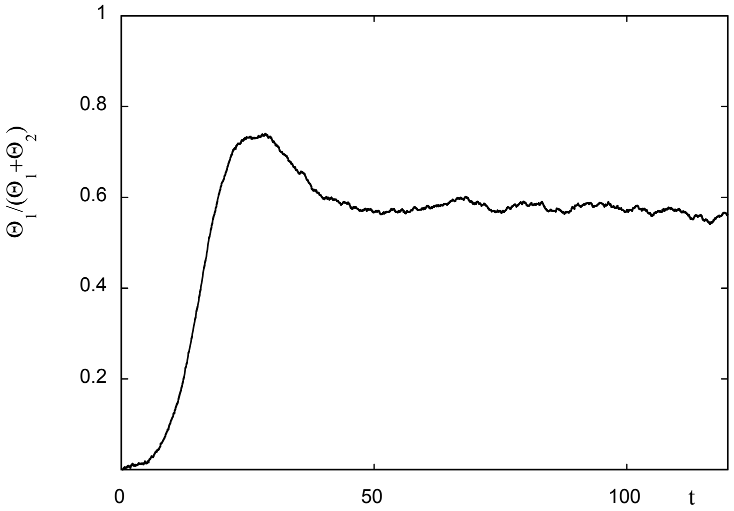


Fig. 9. Effect of variability in the form of uncorrelated Gaussian noise sources of variance equal to 10^{-2} added to eqs (9), on the evolution of the fraction of Θ_1. Parameter values and initial conditions as in Fig. 8.

7. Conclusions and perspectives

We believe that the ideas put forward in this work have a methodological interest that may be further enhanced by e.g. refining the model to account for several internal states or for memory effects. In addition to this fundamental aspect we suggest that our results as they stand can be the starting point for two kinds of applications. Firstly, the reassessment of some of the results available from surveys. In particular the bimodal character of the probability in Fig.6b, reflecting the cooperativity and the smallness of the population size, suggests that the process does not always follow the trend of a purely random event as reflected by a binomial probability distribution. Secondly, the elaboration of prevention strategies. In particular, one may use the switching time t* (eq. (7)) and inflexion point in Figs 2 and 8a) as alert level beyond which the process may get out of control. It may happen as mentioned in Sec. 3 that under the conditions actually prevailing in a given environment this time is much too long compared to the time scale imposed by the local conditions. If so one should switch to a second indicator of an imminent catastrophic evolution, which in our view is provided by the standard deviation $(<\delta X_1^2>)^{1/2}$ or more significantly the ratio

$(<\delta X_1^2>)^{1/2}/<X_1>$. As seen in Sec. 5 this quantity, easily monitored, tends to be enhanced in the vicinity of a collective transition encompassing the populations of interest.

For all the situations analyzed in Sec. 4 with the exception of (iii), the propagation of suicide is explosive and inevitable. The evolution of the propagation of suicide in case (v) is slower because of the limited cooperativity between individuals who have few contacts between them. It would be worthwhile to analyze in the future from this perspective contagion trends in other behavioral disorders typical of adolescence such as running away and addictions.

Another potential application pertains to prevention of situation (iii) in connection with the nature of the class group. There is much discussion about the possibility to create classes with mixed difficult adolescents, that is teenagers with conduct and affective disorders inhibiting the faculty to learn and to succeed in school. In fact the adolescents suffering of conduct disorder have often difficulties in mentalization of their essential depressive symptoms. Even if they do not have the problem of suicidal symptoms in first place, they commit repeatedly a lot of accidents, such as motor vehicle fatalities or even delinquent acts, equivalent to suicidal act. Regrouping this kind of adolescents may be, in our view and according to our results, an error as it will tend to induce further accidents. We see actually that the mixing of susceptible individuals in a "healthy" class group limits the risk of suicidal contagion.

Finally, there is according to our results an interactive "Werther" effect in the form of cyber suicide. In 1774 Johann Wolfgang von Goethe published his by now famous novel "Die Leiden des jungen Werther", in which his hero a young artist, takes his own life after a series of failed attempts to gain the love of beautiful Lotte. The novel had an immediate and an immense impact: men of society used to dress like Werther and as many as 2000 readers seem to have imitated the way he acted and died. As a result of this catastrophic situation, Goethe's novel was banned for a long time in many European countries. More than 200 years later, it appears that the availability of easy communication channels through the mass media and in particular through the advent of the Internet, an increasingly important mode of information and communication among adolescents and young adults is at the origin of a comeback of an interactive "Werther effect". Many studies have addressed the question of an observer copying suicidal behavior that he has seen modeled in the media. Case reports about cyber-suicide have been published, whereby indirect exposure to suicide through media or Internet accounts contributes to subsequent suicide.

Suicide information is easily accessible over the web, as are special chat rooms for discussions with like-minded people. Chat rooms are typical of adolescents and young adults, a group at the highest risk for imitative suicidal behavior (Davidson et al, 1989). In fact mass clusters are media related phenomena. They are regrouped more in time than in space, and are purportedly in response of actual or fictional suicide.

Our results (case II, cf. Fig. 7) provide insights on the mechanisms underlying this collective behavior. They also suggest certain ways of control of the phenomenon and of its follow ups. Health group sites and qualified treatment for suicidal youths should be better promoted. Psychiatrists, parents and teachers should take more interest in their patient's/children's Internet consumption and discuss with them. Question on media and Internet should be part of the anamnesis. The legal options to prevent cyber suicide should be discussed from a national and international perspective because of the dramatic contagion and the criminal abuse of the Internet communities (Becker, Schmidt, 2004). This is crucial especially in view of our results on the dramatically fast pace of the process.

In summary, the major clinical insights afforded by our models are: the elaboration of guidelines for slowing down the propagation of suicide; the identification of possible "alert" indicators; and controlling Internet consumption. The main limitations of the models in their present form are that memory effects are not incorporated and that an individual is taken to be in either of only two mental states.

All in all we believe that in addition to and as complement of the all-important insights afforded by the statistical analysis of surveys, a "first principles" approach of the kind suggested in this chapter may contribute to the unveiling of some of the multiple facets of the dramatic episodes surrounding adolescent suicidal trends.

8. References

Baume, P. Cantor, C.H., Rolfe, A. (1997). Cyber suicide: the role of interactive suicide , notes on the Internet. *Crisis*, 18: 2, 73-79.

Becker, K., Schmidt, M.H. (2004). Internet chat rooms and suicide. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43: 3, 246.

Brent, D.A., Perper, J., Moritz, G., Allman, C., Liotus, L., Schweers, J., Roth, C., Balach, L., Cannobbio, R. (1993). Bereavement or depression? The impact of the loss of a friend to suicide. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32: 6, 1189-1197.

Brent, D.A., Perper, J., Moritz, G., Allman, C., Schweers, J., Roth, C., Balach, L., Cannobbio, R., Liotus, L. (1993). Psychiatric sequelae to the loss of an adolescent peer to suicide. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32: 3, 509-517.

Bridge, J.A., Day, N.L., Day, R., Richardson, G. A., Birmaher, B., Brent, D.A. (2003). Major depressive disorder in adolescents exposed to a friend's suicide. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42: 11, 1294-1300.

Davidson, L.E., Rosenberg, M.L., Mercy, J.A., Franklin J., Simmons, J.T. (1989). An epidemiologic study of risk factor in two teenage suicide clusters. *JAMA*, 17: 262, 2687-2692.

Gillespie, D. T. (1992). *Markov Processes*. New York: Academic Press.

Gould, M.S. (2001). Suicide and the media. *Annals of the New York Academy of Sciences,* 932, 200-224.

Joiner T.E., Jr. (1999). The clustering and contagion of suicide. *Current directions in psychological science,* 8,3, 89-92.

Jones, M.B. and Jones, D.R. (1994). Testing for behavioural contagion in a case-control design. *Journal of psychiatric research*, 28, 35-55.

McKenzie, N., Landau, S., Kapur, N., Meehan, J., Robinson, J., Bickley, H., Parsons, R., Appleby, L. (2005). Clustering of suicides among people with mental illness. *The British Journal of Psychiatry,* 187: 476-480.

Mosekilde, E. (1996). *Topics in Nonlinear Dynamics*. Singapore: World Scientific.

Nicolis G. (1995*). Introduction to Nonlinear Science*. Cambridge: Cambridge University Press.

Pielou, E. K. (1969). *An Introduction to Mathematical Ecology*. NewYork: Wiley-Interscience.

Pommereau, X. (2001). *L'Adolescent Suicidaire*. Paris: Dunod.

Stolley, P.D., Tamar Lasky (1995). *Investigating Disease Patterns : The Science of Epidemiology*, Freeman, New York.

Wheeler L. (1970). *Interpersonal Influence*. Boston : Allyn&Bacon.

12

The Effect of Spatially Inhomogeneous Electromagnetic Field and Local Inductive Hyperthermia on Nonlinear Dynamics of the Growth for Transplanted Animal Tumors

Valerii Orel and Andriy Romanov
Medical Physics and Bioengineering Laboratory
National Cancer Institute
Ukraine

1. Introduction

Cancer is often characterized as a chaotic, poorly regulated growth. Cancer can be viewed as a complex adaptive system. Complex adaptive systems can be described mathematically by nonlinear (chaos) theory including asymmetry, fractal structure and autocorrelation factor (Cramer, 1993). Atypical shape of tumor cells and chaotic structures of blood flow is one from characteristic of cancer process. Atypical change of cell shape in conglomerates of tumor cells and structure of blood flow is accompanied by increase of deterministic chaos (Baish & Jain, 2000; Orel & Dzyatkovskaya, 2000). Complex natural phenomena such as cancer are dynamical systems whose state changes by perturbation. The concept of deterministic chaos is hierarchical for host in contemporary ideas about role of chaos for potential application in oncology (Sedivy & Mader, 1997; Blazsek, 1992). The authors introduced concepts related to chaos theory, such as attractors, fractals and the Lotka-Volterra equations, as potentially useful approaches to allow for the analysis of carcinogenic biological processes as related to selection and competition. In certain situations, these equations give chaotic, non-linear, and nonpredictable results. Given what is known about the enormous complexity of the carcinogenic process, use of models such as these may be perfectly justified, and might provide the theoretical framework that is so desperately needed in this age of data overload to make real progress in the understanding of human carcinogenesis (Garte, 2003).

Entropy is a measure of disorder. The thermodynamic entropy of a cancerous cell is different from that of a normal cell due to the more disordered structure of the cancerous cell. The reversal of entropy flow in tumour tissues may halt tumour development due to reversed signal transmission in the tumour-host entity. This thermodynamic approach may help in the design of cancer therapy (Molnar et al., 2009).

Transplanted animal tumors which can only be experimentally induced by transplanting living tumor cells significant influence on complex adaptive systems include developing of tumor formation for experimental animals. During recent years there has been increasing public concern on potential cancer risks from radiofrequency radiation emissions (Hardell &

Sage, 2008). Inhomogeneous pulsing electromagnetic fields (EF) stimulation of biological tissue was associated with the increase in the number of cells and/or with the enhancement of the cellular differentiation (Diniz et al., 2002). Inhomogeneous (asymmetric) and sinusoidal EF can cause different changes in protein synthesis of cells. It should be noted, that pulsed asymmetric EF and heat shock produced different patterns of polypeptide synthesis (Goodman & Henderson, 1988). Inhomogeneous pulsing EF caused significant reductions in osteoclast formation of tumor necrosis factors, interleukins (Chang et al., 2004) and in osteoblast-like cell of proliferation and gene expression (De Mattei et al., 2005). These observations provide evidence that in vitro inhomogeneous EF affects the mechanisms involved in cell proliferation and differentiation.

Magnetic resonance images demonstrate that malignant tumor can be inhomogeneous media for spatially inhomogeneous EF (Fig. 1). Cancer patient exhibited higher values within the spread parameter *S* range than healthy individual (Fig.2). Each wavefront will be continued independently by an arbitrary inhomogeneous structure of tumor. Propagation of inhomogeneous radio waves in tumor is accompanied by nonlinear effects with greater changes in direction and energy of electromagnetic field than in normal tissues (Kattapuram et al., 1999).

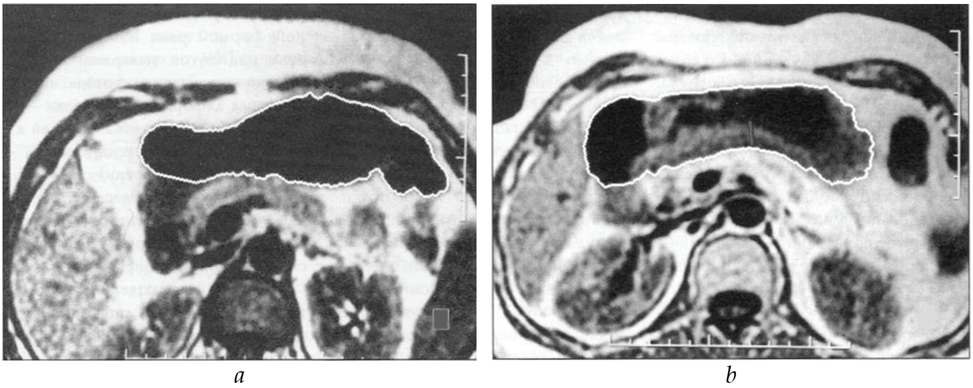


Fig. 1. T1-weighted MR images of the stomach: *a* - healthy individual; *b* - cancer

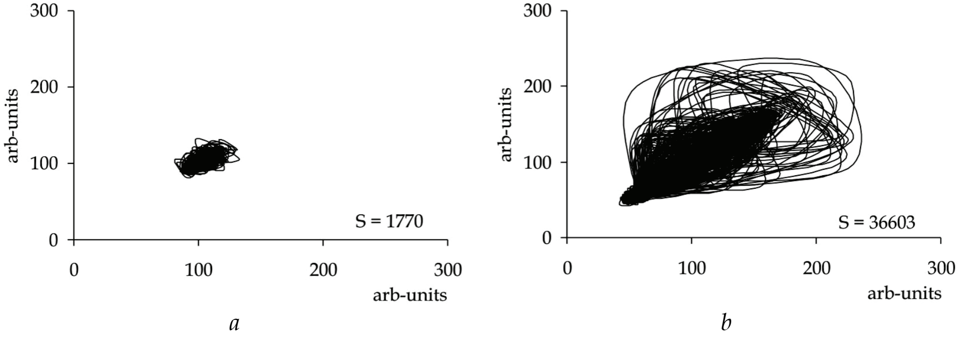


Fig. 2. Phase map of T1-weighted MR images of the stomach: *a* - healthy individual; *b* - cancer

The complete wave field at a tumor will be then obtained as an integral superposition of all wavefront arriving in some neighbourhood of the object. Inhomogeneous electromagnetic wave can be written from Maxwell's equations in the form of an inhomogeneous electromagnetic wave equation (or often "nonhomogeneous electromagnetic wave equation") (Purcell, 1985). Relationships between transplanted animal tumors and external inhomogeneous EF that initiated in them local hyperthermia are important for understand of the principles nonlinear dynamics in cancer process and multimodal approach (and typically nonlinearly) for him treatment (Furusawa & Kaneko, 2000).

Doxorubicin (DOXO) is an anthracycline quinone antineoplastic antibiotic that has been shown to have a wide spectrum of clinical activity against a variety of solid tumors. The mechanisms of DOXO-induced cytotoxicity have been extensively studied and have been shown to include free radical formation and absorption of DOXO into the double helix of DNA resulting in topoisomerase II-mediated DNA damage . DOXO also causes depolarization of the membrane lipid bilayer in different cancer cell lines (Reszka et al., 2001).

Current forms of DOXO are higly toxic to the patient and can cause systematic comlications, most notably cardiotoxicity. Systemic toxicity can seriously decrease the effectness of the drug since a lower dose must be administrated to avoid toxicity. Another approach to avoid toxicity include targeted delivery, however, it is often difficult to ensure that the chemotherapy targets only the cancer tissue and the agent is localising in the target tissue. Therefore in several studies DOXO was combined with electromagnetic hyperthermia with an aim at enhancing antitumor efficacy of the drug (Shen et al., 2008). However, the cytotoxicity of this antitumor agent is increased by elevated temperatures as shown in vitro and in vivo (Marmor, 1979; Chen et al., 2004). Nonetheless, studies of DOXO and electromagnetic hyperthermia are still controversial and often show no synergism or synergism only at the doses that cannot be tolerated by subjects (Gaber, 2002). Positive clinical results of combined treatment with DOXO and electromagnetic hyperthermia are still unsatisfactory. Widespread clinical application of electromagnetic hyperthermia in the patients is limited because temperatures in the range of 41–50°C produce heat shock proteins and initiate drug resistance (thermoresistant) in tumor cells (Roemer, 1999).

Drug resistance is the single most important cause of cancer treatment failure and carries a massive burden to patients, healthcare providers, drug developers and society. It is estimated that multi drug resistance plays a major role in up to 50% of cancer cases. Today, most drug therapies involve multiple agents, as it is almost universally the case that single drugs (or single-target drugs) will encounter resistance. Drug resistance presents some of the greatest challenges to the treatment and eradication of cancer. There are many studies and reports on drug resistance in cancer cells. P-glycoprotein, the expression product of the MDR-1 gene, is strongly associated with both *de novo* and acquired resistant. The protein function as a transmembrane drug efflux pump, transporting cytostatic agents. Glutation and it is dependent enzymes may be involved in resistance to drug by proving cellular protection against free radicals damage. Resistance to drug occurs when the damaged DNA undergoes excision repair. It is likely that many mechanisms of DOXO resistance exist and that such mechanisms are cell specific. Thus, problems related to the development multidrug resistance have led researchers to investigate alternative forms of administrating DOXO for treatment of cancer.

One of complex approach may be in use of inhomogeneous pulsing EF for treatment of drug resistance tumor (Miyagi et al., 2000). Pulsing EF used for stimulation of antiresistant nonthermal effect in mouse osteosarcoma cell line (Hirata et al., 2001). It is known that exposure to the pulsing EF causes depolarization of cell membranes and modifies drug resistance of tumor cells (Pasquinelli et al., 1993; Ruiz-Gómez et al., 2002).

One of the branches in electromagnetic hyperthermia known as inductive hyperthermia (IH) is based on the use of magnetic and electric components of EF in the radiofrequency spectrum for the localization and the concentration of heat during anticancer neoadjuvant therapy or activation of susceptor material implanted in the tumor. The equivalent power density (power density of the plane wave having the same field intensity) for magnetic field is greater than that for the electric field by a factor of ten (Martino, 1962; Moseley, 1988).

During IH of tumor the process of irradiating realize by near-field. In the near-field the maxima and minima of electric and magnetic fields do not occur at the same points along the direction of propagation as they do in the case of the far-field. In this region, the electromagnetic field structure may be highly spatially inhomogeneous and typically, there may be substantial variations from the plane wave impedance i.e., in some regions, almost pure electric fields may exist and, in other regions, almost pure magnetic fields (Jordan & Balmain, 1968). The magnetic component of EF causes heating in tumor tissues through induced eddy currents. Incorporation of antitumor agents into the tumor cells is increased by eddy current stimulation, which is induced by pulsing magnetic fields. Therefore, the cell cycle shifts from the non-proliferative to proliferative phase that leads to increased antitumor activity of the drug (Ivkov et al., 2005; Jin et al., 1998; Orel et al.,2005).

It is well known that EF can influence the chemical reactions to raise their activation energies above threshold levels of thermal noise (Weaver et al.,1999). Nonthermal effects can reduce existing disadvantages on all of the classical thermal treatment (Blank & Soo, 2001; Longo & Ricci, 2007).

In paper (Boddie et al., 1987), it was suggested to produce an inhomogeneous EF pattern with eddy current orthogonal to the magnetic force lines during regionally-focused hyperthermia of a tumor. Really, it is possible to suppose that increased inhomogeneity of EF will activate a non-equilibrium thermodynamical process in a tumor and increase antitumor activity of DOXO. Separately nonthermal and hyperthermal effects (41–50°C) of amplitude-frequency modulation for initiation EF inhomogeneity during treatment of animal tumor is generally used. However, the influence of spatial inhomogeneity of EF and local IH in the range physiological hyperthermia (37–40°C) on nonlinear dynamics of animal tumor growth hasn't been well enough studied yet.

This paper examines the effects of spatially inhomogeneous EF, local IH in the range physiological hyperthermia on nonlinear dynamics of the growth for transplanted animal tumors and entropic action during treatment by DOXO of DOXO-resistant Guerin's carcinoma.

2. Materials and methods

2.1 Experimental animals

In the study, 180 male rats weighing 170 ± 20 g bred in the vivarium of National Cancer Institute and 20 C57BL/6 male mice weighing 19 ± 1 g bred in the vivarium of Bohomolets Institute for Physiology Research, NAS of Ukraine (Kyiv, Ukraine) were used.

2.2 Tumor transplantation

The transplantation of Guerin carcinoma, Lewis lung carcinoma, sarcoma 45, Walker 256 carcinosarcoma and Pliss lymphosarcoma were performed according to the established procedure. All animal procedures were carried out according to the rules of the regional ethic committee. Animals were housed in 2 groups: group 1 – control (no treatment); group 2 – irradiation by elliptic applicator with straight profile (ASP) (40 MHz).

DOXO-resistant Guerin's carcinoma was acquired according to (Solyanik et al., 1999). Thirty sequential subcutaneous transplantations of Guerin carcinoma cells ($3{\cdot}10^6$ per animal) received from DOXO-treated rats. The transplantation of DOXO-resistant Guerin's carcinoma was performed subcutaneously by standard method into the right hind leg. Animals were housed in four groups: 1 – control (no treatment); 2 – DOXO-administration; 3 – DOXO-administration + electromagnetic irradiation (EI) by ASP; 4 – DOXO-administration + EI by elliptic applicator with the circular arc in profile (AAP). Each group contained ten animals.

2.3 Electromagnetic irradiation

First prototype of the device for medical treatment called "Magnetotherm" (Radmir, Ukraine) was used (Nikolov et al., 2008). The frequency of EI was 40 MHz with an initial power of 100 W. The animal tumors irradiated locally (Fig. 3) by inductive coaxial applicators that had differed by the geometry and spatial inhomogeneity of EF.

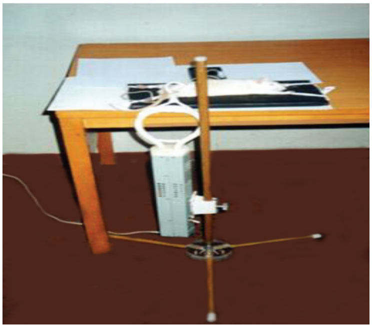


Fig. 3. Electromagnetic irradiation of animal tumors

ASP was an ellipse on a horizontal plane with the semi-axes 1.5×2.5 cm and straight profile (Fig. 4*a*). AAP profile was an arc of the circle with the radius 2.3 cm (Fig. 4*b*) (Ares et al., 1996).

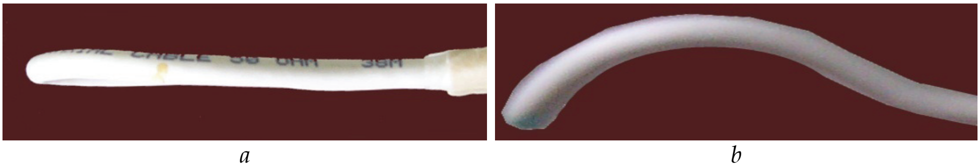


Fig. 4. Appearance of inductive applicator: *a* – ASP; *b* – AAP

EF distribution was computed according to (Mittra, 1973) (Fig. 5). Spatial inhomogeneity of EF was estimated by asymmetry parameter of electric a_E and magnetic a_H field strength distribution according to (Korn & Korn, 1968). Animal tumor was positioned in the center of applicator loop at the distance 0.3 cm from tumor surface. Specific adsorption rates (SAR) of EI were calculated according to (Mittra, 1973). Similar design was used in helical field stellarator for the plasma to increase entropy of EF (Weller et al., 2001).

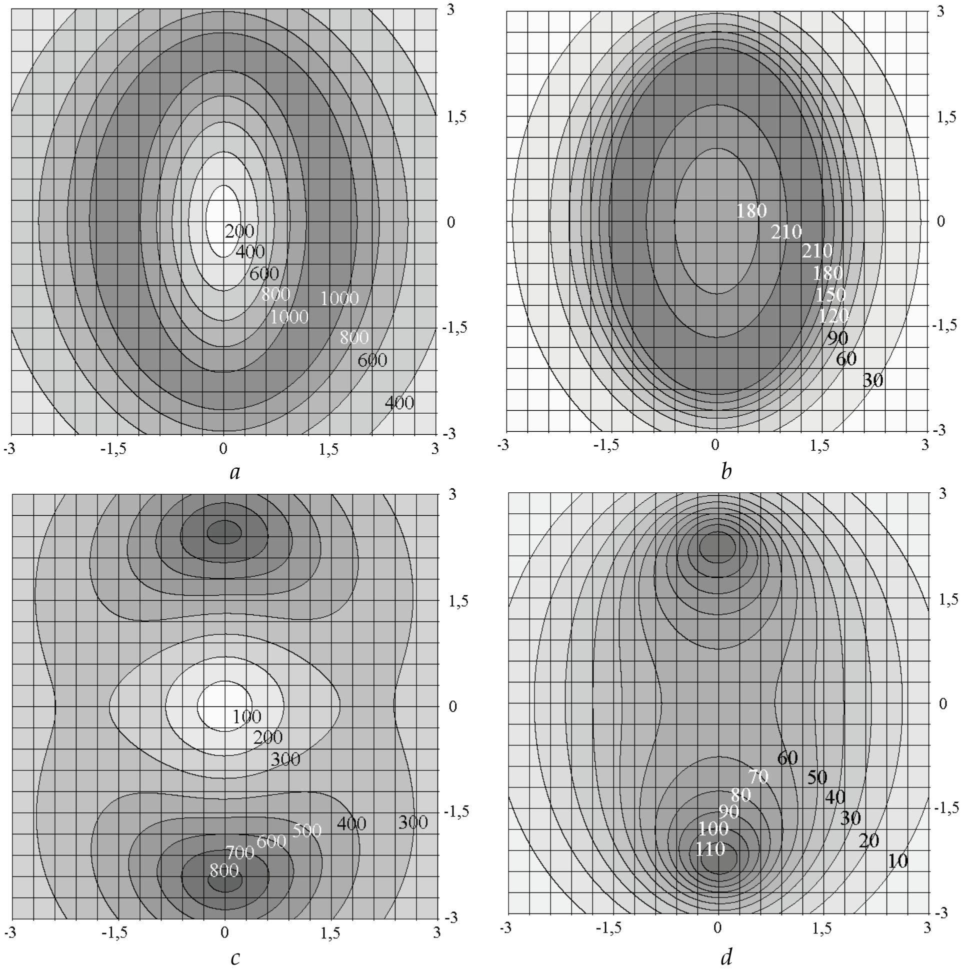


Fig. 5. The isolines of the electromagnetic field: *a* – ASP, electrical component with a_E = – 0.03 a.u.; *b* – ASP, magnetic component with a_H = 0.16 a.u., SAR = 8.8 W/kg; *c* – AAP, electrical component with a_E = 0.89 a.u.; *d* – AAP, magnetic component with a_H = 0.48 a.u., SAR = 1.6 W/kg. Distance to the plane of applicator was 0.5 cm; the values on isolines indicated the tension of the electrical field in V/m and the magnetic field in A/m; the distance in cm is indicated on the axis of abscissas and ordinates

The change of thermal pattern on surface of phantom from fatty tissue of the pig after irradiated by EF shown in Fig. 6. The structure of heat formation on the surface of phantoms depends on the degree of electrmagnetic field nonuniformity and it is similar to computed.

2.4 Treatment of animals with doxorubicin-resistant Guerin's carcinoma

Experimental animals were treated by DOXO (Pharmacia & Upjohn) in the dose 1.5 mg/kg. The treatment was performed five times by DOXO and EI from 10 to 18 days after tumor transplantation every other two days. Tumor volume before treatment was 0.43 ± 0.05 cm^3.

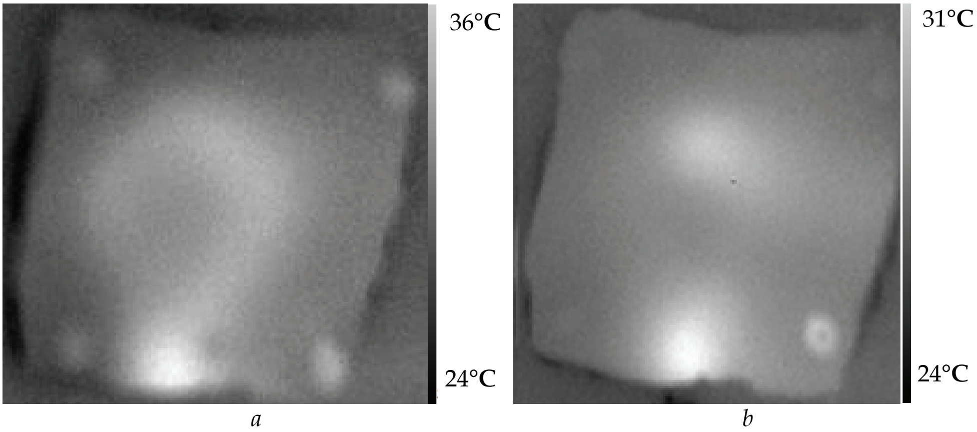


Fig. 6. The change of thermal pattern on surface of phantom from fatty tissue of the pig after irradiated by: *a* –ASP; *b* – AAP

2.5 Temperature studies

The temperature was measured in the tumor centre of DOXO-resistant Guerin's carcinoma by the fiber-optic thermometer TM-4 (Radmir, Ukraine). The kinetics of typical temperature changes for animal tumor under EI is represented in Fig. 7.

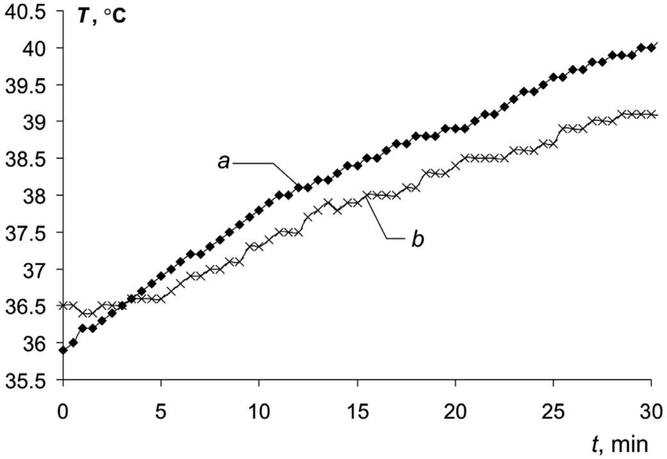


Fig. 7. The temperature changes in the centre of DOXO-resistant Guerin's carcinoma during EI by ASP (*a*) and AAP (*b*)

The temperature was reached up to 39.1°C after 15 min and 40°C after 30 min EI by ASP, as for AAP that was 37.9 and 38.4°C, accordingly. The time between two measurements was 4 hours. It is necessary to notice, that tumor temperature was slightly increased after EI by ASP in comparison with AAP. The kinetics of temperature growth in the tumor was quasilinear. The fluctuations of experimental values evaluated by standard error of temperature in linear regression model. The standard error was 0.15°C for ASP and 0.1°C for AAP.

Preliminary research showed that 15 and 30 minutes of local EI on conventional Guerin carcinoma initiated practically identical strengthening of DOXO antineoplastic activity. Therefore, with aim of milder hyperthermic non-equilibrium effects at physiological temperatures the irradiation was being performed during 15 minutes at once after treatment by DOXO.

The animals were immobilized on the special panel to indicate the heat generation pattern of EF. The thermography was conducted by remote thermograph (B.E. Loshkarev Institute of semiconductors of NAS of Ukraine). The inhomogeneity structure of digital thermograms was estimated by the Shannon entropy (*S*) equation meant for a statistical measure of the disorder (non-equilibrium of thermodynamical process) of a system (Korn & Korn, 1968).

2.6 The analysis of nonlinear kinetics of tumor volume

Nonlinear kinetics of tumor volume was evaluated by growth factor φ according to autocatalytic equation:

$$\frac{dx}{dt} = \varphi(x + x_0)(1 - x)\,, \tag{1}$$

where $x = \frac{\Phi - \Phi_0}{\Phi_\infty - \Phi_0}$ is relative tumor growth by time *t*; $x_0 = \frac{\Phi_0}{\Phi_\infty - \Phi_0}$ is relative tumor volume at the moment of time $t = 0$; Φ_0 and Φ_∞ is initial and limiting tumor volume accordingly; Φ is tumor volume at the moment of time *t* (Emanuel, 1977).

The solution of equation (1) is

$$\Phi = \Phi_0 + \Phi_0 \cdot \frac{e^{\varphi \frac{\Phi_\infty}{\Phi_\infty - \Phi_0} \cdot t} - 1}{1 + \frac{\Phi_0}{\Phi_\infty - \Phi_0} \cdot e^{\varphi \frac{\Phi_\infty}{\Phi_\infty - \Phi_0} \cdot t}}\,. \tag{2}$$

The effect of EF and local IH on nonlinear dynamics of the growth of animal tumors was evaluated with the braking ratio:

$$\kappa = \frac{\varphi_c}{\varphi_{EI}}\,, \tag{3}$$

where φ_c – is growth factor for control group of animals, φ_{EI} – is growth factor for group after EI.

2.7 The heterogeneity of tumor structure in ultrasound image

Ultrasonic studies were done before and right after EI by ultrasonic apparatus ATL HDI 3000 (Fillips, USA) with the use of 6 MHz transducer. During ultrasonic studies the transducer was stationary fixed relative to animal tumor.

The heterogeneity of ultrasound image *G* in tumor tissues for studies of tumor vessels was evaluated with spatial autocorrelation statistics *r* by Moran (Bailey & Gatrell, 1995; Orel et al., 2007a):

$$G = 1 - r, \tag{4}$$

$$r = \frac{n\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i-\overline{x}\right)\left(x_j-\overline{x}\right)}{\left(\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2\right)\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} w_{ij}} \tag{5}$$

where n is the number of pixels in selected region of interest in ultrasound image, x_i is the intensity of ith pixel, $\overline{x}$ is the mean intensity of whole region of interest, and w_i is a distance-based weight which is the inverse distance between pixels i and j ($1/d_{ij}$).

2.8 Statistical and correlation analysis

Statistical processing of numerical results was carried out using Statistica 6.0 (© StatSoft, Inc. 1984–2001) computer program with parametric Student's *t*-test. Correlation analysis was performed with the MATLAB 7.0 (©1984–2004 The MathWorks, Inc.) software.

3. Results

3.1 Changes in nonlinear dynamics of the growth for animal tumors under the influence of spatially inhomogeneous electromagnetic field and local inductive hyperthermia

As it is shown in table 1 the growth kinetics of animal tumors had very different nonlinear responses under the influence of spatially inhomogeneous electromagnetic fields (a_E = – 0.03 a.u.; a_H = 0.16 a.u.) and local IH initiated by ASP. The strongest inhibition effect under the influence of EI was in Pliss lymphosarcoma and sarcoma 45. The growth stimulation of animal tumors after EI was recorded in Walker 256 carcinosarcoma. Animal tumors for Lewis lung carcinoma grew nonsignificantly but average number of metastases on a mouse in the lungs was increased on 86%. Nonlinear dynamics of tumors' growth was much differed for each single animal in all investigated groups.

EI of Gueren carcinoma by AAP with inhomogeneous electromagnetic fields (a_E = 0.89 a.u.; a_H = 0.48 a.u.) statistically not significant changed nonlinear dynamics of malignant growth in comparison with control group of animal without treatment.

Tumor	Parameters		
	φ_c, day^{-1}	φ_{EI}, day^{-1}	κ
Guerin carcinoma	0.45 ± 0.01	0.46 ± 0.05	0.99
Lewis lung carcinoma	0.39 ± 0.02	0.36 ± 0.01	1.07
Sarcoma 45	0.60 ± 0.03	0.45 ± 0.01*	1.31
Walker 256 carcinosarcoma	0.60 ± 0.01	0.66 ± 0.01*	0.91
Pliss lymphosarcoma	0.42 ± 0.02	0.32 ± 0.01*	1.32

* Statistically significant difference from control group

Table 1. The growth kinetics of animal tumors

The ultrasonic studies were used for interpretation of peculiarities in tumor blood flow during EI. Guerin carcinoma only was researched because there were problems in visualization of ultrasound images on the monitor for other experimental tumors. Fig. 8 shows the sonogram of Guerin carcinoma on the 10th day after tumor transplantation before and after EI. The sonograms show that tumor heterogeneity parameter G for Guerin carcinoma was higher in 2.9 times after EI than without irradiation. This is in accordance with well known medical observations that EI and mild hyperthermia in tumor is characterized by intensive tumor blood flow (Song et al., 2005).

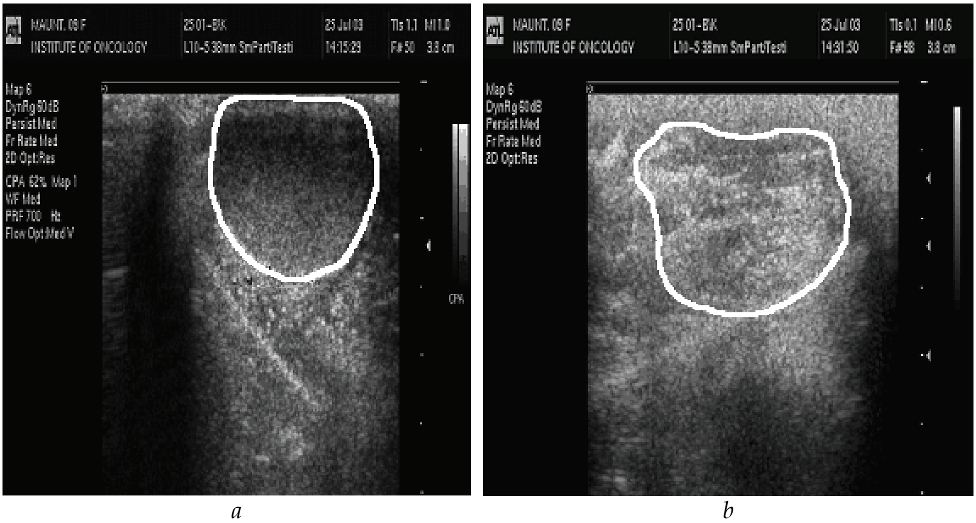


Fig. 8. The sonogram of Guerin carcinoma and tumor heterogeneity parameter G:
a – without EI (G = 0.24); *b* – after 15 min EI (G = 0.69)

According to the presented data, one may suppose that recorded effects of inhibition or stimulation growth for animal tumors after electromagnetic stimulation may be caused by peculiarity of vascular damages in different experimental tumors.

3.2 The effect of spatially inhomogeneous electromagnetic field, local inductive hyperthermia and doxorubicin on nonlinear dynamics of tumor growth for animals with doxorubicin-resistant Guerin's carcinoma

As it is shown in Fig. 9, nonlinear dynamics of the growth for tumor volumes on 10 and 12th day after tumor transplantation was identical. Since 14th day after transplantation tumor volumes for animals from 4 groups were statistically significant decreased in comparison with the animals of 1, 2 and 3 groups on 88%, 79% and 82% ($p < 0.05$) accordingly in average. The growth kinetics of animal tumors is shown in table 2. The growth kinetics for 3 group had minimal response under the influence of DOXO and EI by ASP generated EF with a_E = –0.03 a.u.; a_H = 0.16 a.u. At the same time the complete resorption were observed on 20th day after tumor transplantation for 40% animals from 4 group (DOXO + EI by AAP, a_E = 0.89 a.u. and a_H = 0.48 a.u.). The recurrent tumor growth hadn't been detected for 4 months after the treatment. Obtained results were testified by the study repeated in 4 months.

Our research showed that antitumor effect of DOXO was not depended on the rotation of applicator on horizontal plane relative to tumor. Antitumor effect of DOXO didn't changed significantly under EF after mechanochemical activation of drug before treatment.

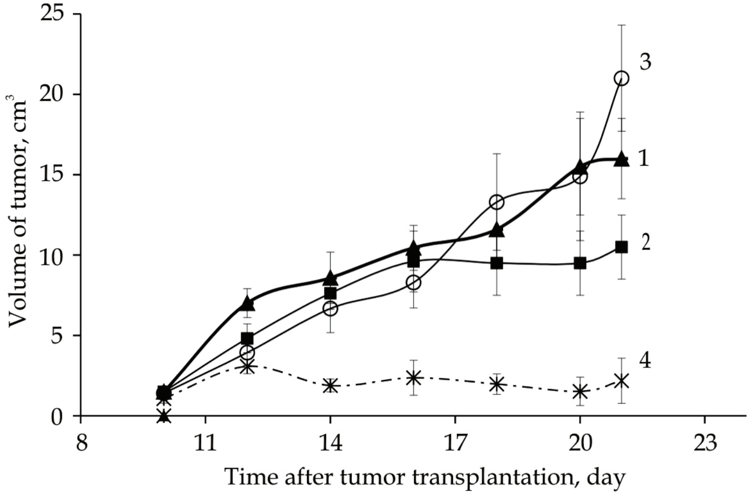


Fig. 9. EI and DOXO-induced changes in nonlinear dynamics of the growth for DOXO-resistant Guerin's carcinoma: 1 – without DOXO and EI (control); 2 – DOXO; 3 – DOXO + EI by ASP; 4 – DOXO + EI by AAP

N	Treatment	Parameters	
		φ, day^{-1}	κ
1	Without DOXO and EI (control)	0.46 ± 0.01	
2	DOXO	0.42 ± 0.01	1.08
3	DOXO + EI by ASP	0.47 ± 0.02	0.97
4	DOXO + EI by AAP	0.32 ± 0.02*	1.43

* Statistically significant difference from control group

Table 2. The growth kinetics of animal tumors

3.3 Thermography

Thermal patterns of tumor's surface and the panel after EI are presented in Fig. 10. Maximal inhomogeneity of tumor surface and indicative panel that estimated by entropy was

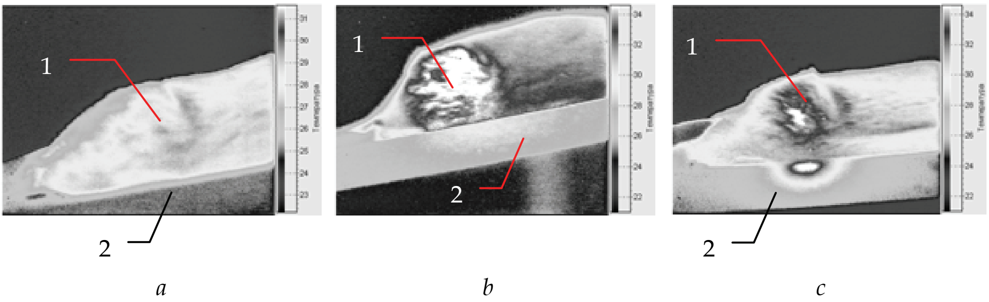


Fig. 10. Change of thermal pattern on tumor surface after transplantation on 15 day (1) and indicative panel (2) after EI; *a* – without EI (control); *b* – EI by ASP; *c* – EI by AAP

obtained for AAP with increased spatial inhomogeneity of EF (Fig. 11). It testifies, that the use of EF with increased spatial inhomogeneity influenced on nonuniform temperature distribution on the surface of animal tumor.

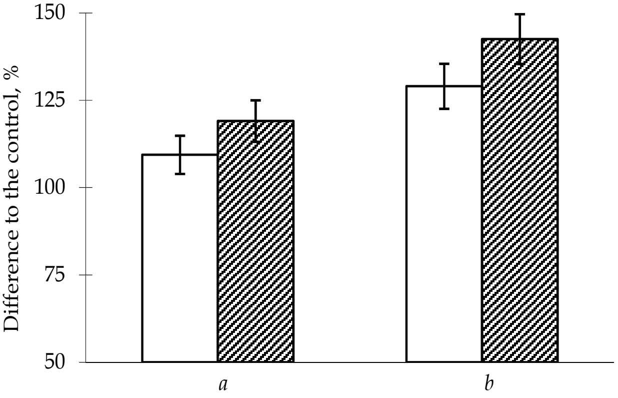


Fig. 11. The inhomogeneity (entropy) of thermal pattern on tumor surface after transplantation on 15 day (*a*) and indicative panel (*b*) after EI: □ – by ASP; ▨ – by AAP. On an axis there is a difference to the control (without EI)

3.4 Ultrasonic studies

Typical tumor sonograms on the 15th day after the tumor transplantation and 15 minutes of EI are shown in Fig. 12. The computer nonlinear analysis of composite B-mode and steered color Doppler acoustic image demonstrated that heterogeneity *G* was decreased by 30% after EI with increased spatial inhomogeneity by AAP. It testifies, that the use of EF with increased spatial inhomogeneity influenced on the vessel dilation in malignant tissues. This is in accordance with aforementioned observations that EI and moderate hyperthermia in a tumor is characterized by the typical change of a tumor's blood flow and increased oxygenation of tumor cells (Song et al., 2005).

4. Discussion

4.1 The influence of spatially inhomogeneous electromagnetic field and inductive hyperthermia on nonlinear aspects of malignant growth

Our study demonstrated that spatially inhomogeneous electromagnetic fields with asymmetry parameters $a_E = -$ 0.03 a.u. and $a_H = 0.16$ a.u. and local IH in the range physiological hyperthermia cause influence on nonlinear dynamic of the growth of transplanted animal tumor (Orel et al., 2007b). The cancer processes are an example of non-equilibrium, non-linear process. It is predictable locally in the very short-term, but not in the medium- and long-term, as typical of systems exhibiting deterministic chaos (Rubin, 1984). The effects of spatially inhomogeneous EF and local IH in the range physiological hyperthermia warrant increased to create chaos for animal with cancer process. It effects of inducing extremely large and very rapid surges of stochastic endogenous signals in tumor

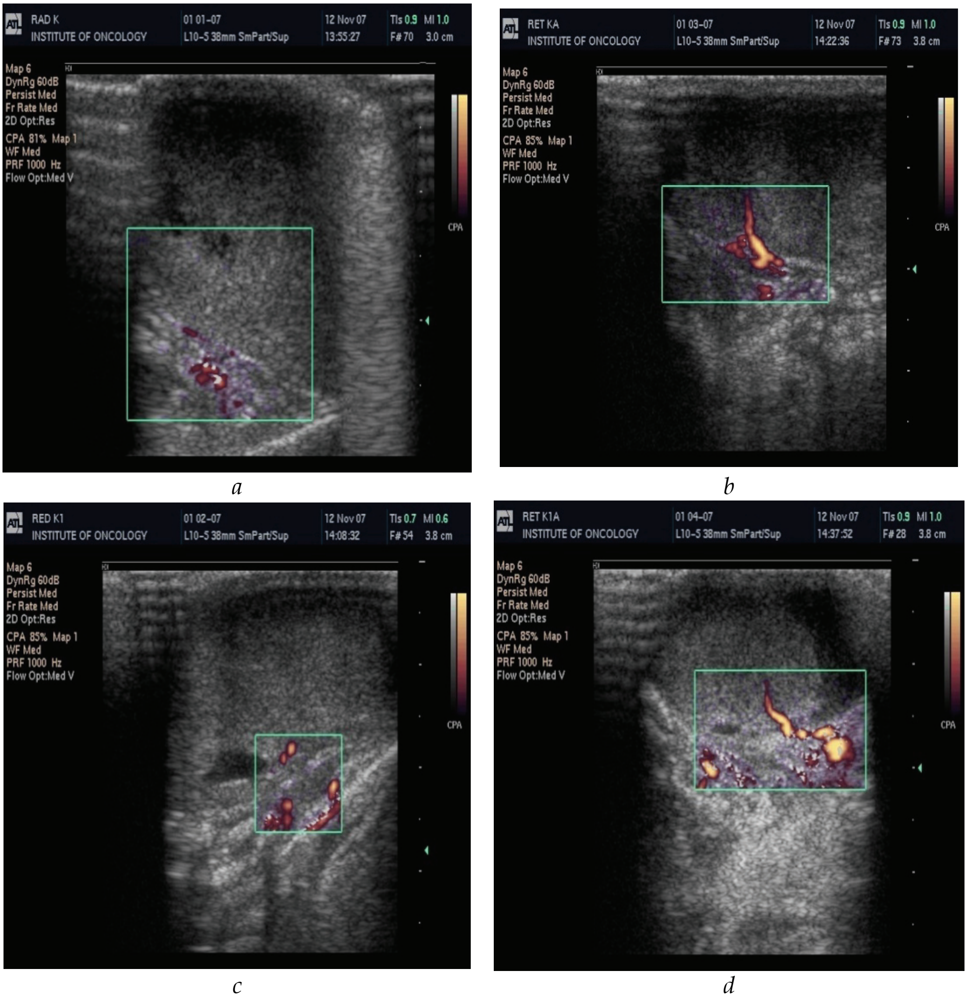


Fig. 12. The change of heterogeneity (*G*) in composite B-mode and steered color Doppler acoustic image of tumor: *a* – without EI (control), *G* = 0.55; *b* – EI by ASP, *G* = 0.56; *c* – without EI (control), *G* = 0.60; *d* – EI by applicator with AAP, *G* = 0.42

cells. They tend to be quasi (almost but not quite)-periodic, the periodicities are a complex of many periods, and they can swing between different quasi-periodic states. But they are not at all random (Waliszewski et al., 1998; Marino et al., 2000,2009).

Living systems are organized such that they manifest operational features ascribed to hierarchical and heterarchical structures from quantum to organism levels (Dirks, 2008). In mainstream biology that would enable us to understand how EF below the "thermal threshold" could have any effects. That, despite the fact that consistent changes in gene expression and DNA breakages – considered to the 'most solid' evidence – have now been obtained. Some biological effects are indeed associated with EF so weak that the energies in those fields are below the energy of random thermal fluctuations. Molecular signaling in eukaryotic cells is accomplished by complex and redundant pathways converging on key molecules that are allosterically controlled by a limited number of signaling proteins. p53-signaling pathway is an example of a complicated sequence of signals produced in response to DNA damage. This pattern of signaling may arise from chance occurrences at the origin of life and the necessities imposed on a nanomolar system (Yarosh, 2001; Schneider et al., 2004). Signals from tumor cells look like stochastic processes although their latent mechanism is deterministic. These are the 'butterfly' effects: the molecule of DNA could affect the metabolism in organism (in common with a proverbial butterfly flapping its wings in the Amazon rainforest could affect the weather in London) (Carrubba et al., 2007; Carrubba et al., 2008).

Thereby inhomogeneous EF influence on genetic instability gives rise to the diversity of cancer process. Evidently above mentioned can incarnate of foundation for interpretation different in nonlinear dynamics for transplanted animal tumors.

According to the presented data, one may suppose that recorded effects of inhibition or stimulation growth for animal tumors after spatially inhomogeneous electromagnetic stimulation may be caused by peculiarity of vascular damages in different experimental tumors. These results are important for clinical application of medical technologies because they testify against the use of electromagnetic hyperthermia as a basis for the monotherapy of malignant human tumors and the necessity to facilitate local EI during anticancer neoadjuvant therapy with the use of drugs or magnetic nanoparticles. In general, the application of local electromagnetic hyperthermia in clinical oncology is effective when combined with chemotherapy or radiochemotherapy as shown in (Falk & Issels, 2001).

4.2 An increase of doxorubicin antitumor effect by entopictic action of spatially inhomogenous electromagnetic and heat fields

The spatially inhomogeneous field is definitely changed by the geometric and mass/structure variance of the tumor itself. The effect of spatially inhomogeneous EF during EI on transformation of radio waves and thermal descriptions in malignant tumors was investigated. It is shown that structure of heat formation in the range physiological hyperthermia on tumor surface depends on the degree of inhomogeneity of EF. In our next experiments revealed entropic action in antitumor effect for DOXO of inhomogenous electric ($a_E = 0.89$ a.u.), magnetic fields ($a_H = 0.48$ a.u.) and temperature in the range physiological hyperthermia during EI.

This action we visualized for other antitumor drug too. The highest antitumor and antimetastatic activity was caused by the combined action of cisplatin and irradiation by spatially inhomogeneous EF and local IH of animals with resistant to cisplatin substrain of Lewis lung carcinoma too (Orel et al., 2009).

The heterogeneous structure of blood vessels in malignant tissue specified by greater specific area of interaction with antitumor drug in comparison with normal tissue. Chaotic signals of inhomogeneous EF can be applied to increase creativity of artificial intelligence, in fluid dynamics of blood to induce turbulence to increase therapeutic effects for antitumor drug, in biochemical processes to drive reactions toward otherwise improbable biochemical compounds, or to raise bond energies above threshold levels without destructive heat. It can be applied to the breaking up of separative attitudes among metastasized cancer cells and aiding in the recovery from cancer (Orel et al., 2004).

What is physicochemical property of spatially inhomogeneous electric, magnetic and temperature fields which influenced on nonlinear dynamics of biological process in the tumor and initiated action as increased antitumor effect for DOXO?

The heterogeneity for tumor structure usually is more variable than for normal tissues. Therefore, we studied influence of EF on transformation of electric, magnetic and thermal fields in heterogeneous (rubber foam + 0.9% NaCl solution) and homogeneous (0.9% NaCl solution) phantoms.

Preliminary research showed that transformation of EF and thermal patterns in phantoms was investigated during EI by spatially inhomogeneous EF (Orel et al., 2008). The change of electric (ΔE) and magnetic (ΔH) component under the influence of phantoms was calculated as follows:

$$\Delta E = E - E_{0,}, \tag{6}$$

$$\Delta H = H - H_0, \tag{7}$$

where E and H is electric and magnetic field intensity under phantom, E_0 and H_0 is electric and magnetic field intensity in the air, respectively.

It is shown in Fig. 13 that the structure of heat formation on the surface of phantoms depends on the degree of EF nonuniformity and it is similar to computed in Fig. 5 EF distribution. Relative increase of magnetic field strength $\Delta H/H_0$ in phantoms after EI by AAP was in 3.5 times greater than by ASP on the average (Table 3). Relative increase of temperature $\Delta T/T_0$ in phantoms was smaller in 5.4 times after EI by AAP compared to ASP on the average. In rubber foam phantom the ratio $\Delta T/T_0$ increased in 8.6 times after EI by AAP compared to 0.9% NaCl solution phantom. It testifies stronger transformation of spatially inhomogeneous EF for heterogenous structure of rubber foam phantom than for homogeneous structure of 0.9% NaCl solution phantom. The transformation of inhomogeneous EF to thermal patterns for phantoms was similarly to an effect for animal tumors (see chapter 3.3).

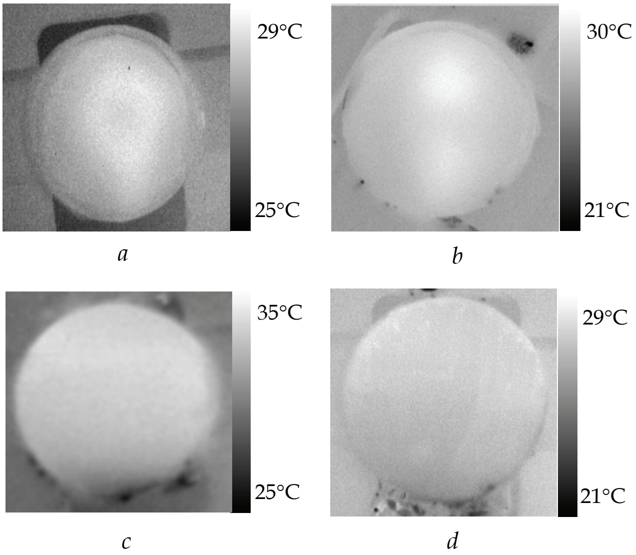


Fig. 13. The change of thermal pattern on phantom surface after electromagnetic irradiation by ASP of foam rubber + 0.9% NaCl solution (*a*), AAP of foam rubber + 0.9% NaCl solution (*b*), ASP of 0.9% NaCl solution (*c*), AAP of 0.9% NaCl solution (*d*)

Phantom	Applicator	$\Delta E/E_0$, %	$\Delta H/H_0$, %	$\Delta T/T_0$, %
NaCl 0.9% solution	ASP	47 ± 3	8.0 ± 1.0	0.20 ± 0.02
NaCl 0.9% solution	AAP	19 ± 3*	20.0 ± 3.1*	0.10 ± 0.01
Foam rubber	ASP	49 ± 6	7.0 ± 0.5	6.2 ± 1.0
Foam rubber	AAP	28 ± 4*	31.0 ± 3.5*	0.7± 0.2*

* $p < 0.05$ compared to similar parameter of ASP

Table 3. The ratios $\Delta E/E_0$, $\Delta H/H_0$ and $\Delta T/T_0$ for phantoms

We studied the transformation of EF and thermal patterns in physiological phantoms – muscular, fatty, liver tissues and packed red blood cells too. The result was similarly to physical phantoms.

Analyzing the above-mentioned phantom researches, it is possible to mark the problem in our discussion. Is an increase of antitumor effect for drug during treatment under the action of spatially inhomogeneous EF and nonuniform temperature field with temperature peak 37.9°C accompanied by the tendency of biological system to move toward randomness or disorder that increased thermodynamical entropy in the tumor? As contrasted with our experiments in classic electromagnetic hyperthermia the uniform heat with discrete peaks temperature more 41°C is basic for cancer therapy (Franckena et al., 2009) that is not enough for essential change of the thermodynamic entropy in the tumor.

To answer on this question we studied the growth dynamics for Guerin carcinoma during treatment by DOXO under influence of inhomogeneous EF and accessory uniform and nonuniform heat in tumor activated by external water heating. Experimental animals were treated by DOXO (Pharmacia & Upjohn) in the dose 1.5 mg/kg. The treatment was performed four times by DOXO, EI and external uniform and nonuniform heating by the rubber hot-water bottles from 9 to 15 days after tumor transplantation every other two days. The growth kinetics of Guerin carcinoma was varied for different groups (Table 4). Spatially inhomogeneous EF and nonuniform heat field in the range of physiological hyperthermia was maximally increased antitumor effect of DOXO for transplanted Guerin carcinoma. But temperature in the tumor for this case had a lesser value.

We can suppose that increase of antitumor effect by inhomogeneous EF for drug during treatment of the tumor accompanied by the change of thermodynamical entropy.

Treatment	Temperature in the centre of tumor, °C	Parameters	
		φ, day^{-1}	κ
Control (without DOXO, EI and accessory heat)	36.5	0.54 ± 0.06	1.00
DOXO	36.5	0.42 ± 0.02*	1.28
DOXO + accessory uniform heat + EI by AAP	41.5	0.38 ± 0.01*	1.43
DOXO + accessory uniform heat	40	0.37 ± 0.01*	1.45
DOXO + accessory nonuniform heat	38	0.36 ± 0.01*	1.50
DOXO + EI by AAP	37.9	0.35 ± 0.01*	1.53

* Statistically significant difference from control group

Table 4. The growth kinetics of Guerin carcinoma during 15 days after tumor transplantation

It is well known that EF can initiate electro- and magnetocaloric effects. The electro- and magnetocaloric effects are electro- and magneto-thermodynamic phenomenons in which a reversible change in temperature of a suitable material is caused by exposing the material to a changing EF. It was accompanied by changes in transfers from electromagnetic to thermodynamic entropy and enthalpy (Nikiforov, 2007; Crosignani & Tedeschi, 1976). Therefore, we can symbolically included high-frequencies electromagnetic IH in separate class of electro- and magnetocaloric effects.

Described above physicochemical interaction between spatially inhomogeneous electric, magnetic and temperature fields in the phantoms was probably similar to physicochemical interaction in the tumor. They could influence on nonlinear dynamics of biological process. We suppose, that it was interconnection between nonlinear conversion effects of spatial inhomogeneous electric, magnetic fields (a_E = 0.89 a.u.; a_H = 0.48 a.u.) and initiated spatial inhomogeneous temperature field in the heterogeneity tumor structure during propagation of radio waves through malignant tissues. Entropy action is expressed in increase of antitumor effect for DOXO. Alongside located normal tissue toxicity effect was minimal through low level their heterogeneity.

In future we will be able to develop of novel and effective strategies for prevention and treating cancers on the basis of understanding of nonlinear dynamics of adaptive systems associated with tumorigenesis aspects during signaling interaction between cancer cells and the host for complex treatment of patients by whole-body irradiation with local varying spatial inhomogeneous EF.

4.3 Nonlinear model of growth dynamics for transplanted animal tumor during irradiation by spatially inhomogeneous electromagnetic field and inductive hyperthermia

Spatially inhomogeneous EF and initiated it heat manage the formation and disintegration of dissipative structures lying in the basis of self-organization processes in organism at physiological hyperthermia. We applied Waddington's epigenetic landscape model which is a metaphor for how gene regulation modulates development to interpret the changes in thermodynamical parameters (entropy, enthalpy etc.) during nonlinear tumor growth of transplanted animal tumors (Goldberg et al., 2007). The traditional mechanist, pathway-centered explanation assumes that a specific, "instructive signal" i.e., a messenger molecule or external signal of that interacts with its cognate cell surface receptor, tells the cells which particular genes to active in order to establish a new cell phenotype. Essentially, cell distortion triggered the cell to "select" between different preexisting attractor states (Sole, R. et al., 2006). A certain chemical reaction is performed at different temperatures and the reaction rate is determined. The reaction rate (k) for a reactant or product in a particular reaction is intuitively defined as how fast a reaction takes place according to the Eyring–Polanyi equation:

$$k = \frac{k_B T}{h} e^{\frac{\Delta S}{R}} e^{-\frac{\Delta H}{RT}} , \tag{8}$$

where: k_B is Boltzmann's constant, h is Planck's constant, T is absolute temperature, ΔS is entropy of activation, ΔH is enthalpy of activation, R is gas constant (Polanyi, 1987).

The interaction effect of spatially inhomogeneous EF with heterogenous structure of animal tumors just as described above for the phantoms initiated spatially inhomogeneous thermal field gradient in malignant tissues in the range physiological hyperthermia. It was accompanied by stochastic changes in transfers from electromagnetic to thermodynamic entropy ΔS and enthalpy ΔH of activation and, respectively, stochastic changes of the reaction rate that influence on nonlinear (chaotic) aspects in malignant growth (random effect of increase or decrease) for transplanted animal tumors (see chapter 3.1). Spatially inhomogeneous EF with increased asymmetry parameters during treatment of animal tumors by DOXO (Table. 4) accompanied by the change of entropy of activation (ΔS), the reaction rate k (eq.8) and initiate enzyme catalysis topoisomerase II-mediated DNA damage and free radical formation, absorbing them into double helix of DNA and resulting damage of tumor cells. In this case the number of free radicals increased, in our opinion, as a result of the effect of spin conversion in radical electron pair.

Let us consider kinetic model of tumor growth under the action of DOXO and nonuniform heat field in the range of physiological hyperthermia initiated by spatially heterogeneous EF. Let tumor cells multiplied with the growth factor λ, and DNA of some part of cells loses their ability for replication under the action of DOXO and nonuniform heat field. The appropriate equation can be written as

$$\frac{dx}{dt} = \lambda x - v \, . \tag{9}$$

where x is the number of tumor cells in unit volume with capable of replication DNA, v is the rate of appearing of tumor cells with damaged DNA, which is unable to replicate.

Doxorubicin is known to interact with DNA by intercalation and inhibits the progression of the enzyme topoisomerase II, which unwinds DNA for transcription. Doxorubicin stabilizes the topoisomerase II complex after it has broken the DNA chain for replication, preventing the DNA double helix from being resealed and thereby stopping the process of replication. Schematically this reaction can be written down as:

$$DOXO + [TOP{+}DNA] \rightarrow DNA^*, \tag{10}$$

where [*TOP*+*DNA*] is topoisomerase II complex, *DNA** is damaged DNA.

Let $y = C_{DOXO}$ is the concentration of DOXO, $y(0) = y_0$ – beginning maximal concentration of DOXO, $y \geq 0$; $u = C_{TOP}$ is the concentration of topoisomerase II, $u > 0$. For the open system the concentration of *DOXO* and *TOP* in the reaction (10) is described taking into account diffusion:

$$\begin{cases} \dfrac{\partial y}{\partial t} = -r + D_y \dfrac{\partial^2 y}{\partial l^2}, & (11) \\ \dfrac{\partial u}{\partial t} = -r + D_u \dfrac{\partial^2 u}{\partial l^2}, & (12) \end{cases}$$

where r is reaction rate, D_y and D_u is effective diffusion rate, l is spatial coordinate.

In accordance with kinetic law of mass action during steady quasistationary regime in the system the rate r of reaction (10) is expressed as

$$r = kyu, \tag{13}$$

where k is the constant of reaction rate (Ederer & Gilles, 2007).

The concentration u of topoisomerase II is related with the number x of tumor cells in unit volume:

$$u = ax, \tag{14}$$

where a is a coefficient.

The rate v of appearing of tumor cells with damaged DNA determined by the cells with topoisomerase II reacted in (10):

$$v = -\frac{1}{a}\frac{du}{dt}. \tag{15}$$

Putting in (15) the expression for $\frac{du}{dt}$ from (12) and taking (14) into account, we will get

$$v = \frac{r}{a} - D_x \frac{\partial^2 x}{\partial l^2}. \tag{16}$$

Thus, equations (9) and (11) it is possible to write down as a system:

$$\begin{cases} \dfrac{dx}{dt} = \lambda x - \dfrac{r}{a} + D_x \dfrac{\partial^2 x}{\partial l^2}, \\ \dfrac{dy}{dt} = -r + D_y \dfrac{\partial^2 y}{\partial l^2}. \end{cases} \tag{17}$$

The constant of reaction rate k depends on temperature T according to Arrhenius equation:

$$k = Ae^{-\frac{E}{RT}}. \tag{18}$$

Taking (13) and (18) into account the system (17) will look like:

$$\begin{cases} \dfrac{dx}{dt} = \lambda x - Ae^{-\frac{E}{RT}} xy + D_x \dfrac{\partial^2 x}{\partial l^2}, \\ \dfrac{dy}{dt} = -aAe^{-\frac{E}{RT}} xy + D_y \dfrac{\partial^2 y}{\partial l^2}, \end{cases} \tag{19}$$

with initial condition $y(0) = y_0$ and edge conditions $x > 0$ and $y > 0$.

The system of equations (11) describes the nonuniform thermal effect of the spatially inhomogeneous EF on the growth kinetics of the number of tumor cells under the action of DOXO.

According to the presented data, one may suppose that recorded effects of growth inhibition for DOXO-resistant Guerin's carcinoma after treatment by DOXO and local EI by EF with increased spatial inhomogeneity (a_E = 0.89 a.u.; a_H = 0.48 a.u.) may be connected with the initiation of membrane depolarization due to two steps. Firstly – ionic cyclotron resonance and next – paramagnetic resonance (Liboff AR, 1985; Blanchard & Blackman 1994; Bezrukov & Vodyanoy, 1997), which initiated the antitumor activity of DOXO. Its biochemical mechanisms may be the alteration of the tumor microenvironment via changes in the pH gradient between the extracellular environment and the cell cytoplasm (De Milito & Fais, 2005) and probably EF influency on free radical metabolism of human body (Jin et al., 1998). Thus, we can assert that spatially inhomogeneous EF and local IH initiated in tumor of the reactions with multiple physicochemical properties.

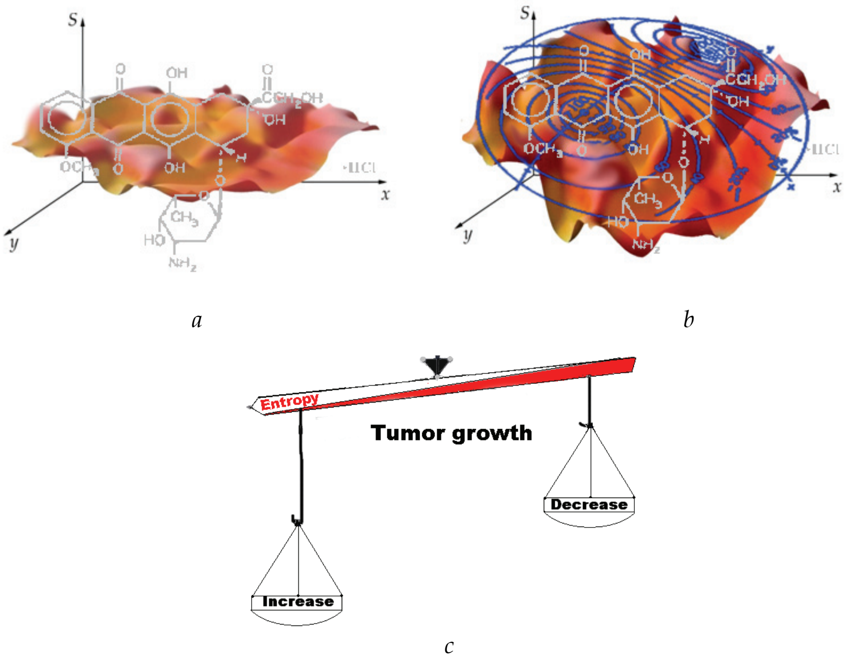


Fig. 14. Spatial distribution of entropy of activation in the tumor during treatment by Doxorubicin hydrochloride $C_{27}H_{29}NO_{11}{\cdot}HCl$ and spatial inhomogeneity electromagnetic field with increased asymmetry parameters: *a* – Doxorubicin hydrochloride; *b* – Doxorubicin hydrochloride under the action of spatially inhomogeneous EF and IH; *c*- entropy of activation and tumor growth

Our preclinical and early clinical data suggest that combining superficial and intracellular agents can synergize and leverage single-agent activity. The aforementioned effect of influence of spatially inhomogeneous EF and local IH at physiological temperatures on increase of antitumor activity for drug used in clinical practice during chemotherapy of cancer patients (Nikolov et al., 2008).

5. Conclusion

1. EI by spatially inhomogeneous EF and local IH in the range physiological hyperthermia of transplanted animal tumors manifests many of nonlinear (chaotic) aspects in malignant growth.
2. An increase of spatially inhomogeneous EF and local IH in the range physiological hyperthermia increased antitumor effect of DOXO for transplanted DOXO-resistant Guerin's carcinoma and accompanied by the change of thermodynamical entropy.
3. Understanding the chaotic theory for cancer and its interplay may enable similar strategies to be employed in the treatment of cancer by spatially inhomogeneous EF and local IH in the range physiological hyperthermia.

6. Acknowledgements

The authors would like to thank Ph.D. Dunaevsky V.I. for thermography measurements. We wish to give special thanks to Ph.D. Dzyatkovskya N.N., Dr.-Ing. Nikolov N.A., Dr.-Ing. Mel'nik Yu.I. and Dzyatkovskya I.I. for experimental studies and joint discussions.

7. References

Ares, F., Rengarajan, S, Lence, J., Trastoy, A. & Moreno, E. (1996). Synthesis of antenna patterns of circular arc arrays. *Electronics Letters*, Vol. 32, No. 20, – P. 1845–1846.

Bailey, T. & Gatrell, A. (1995). *Interactive Spatial Data Analysism*, Wiley, New York.

Baish, J. & Jain, R. (2000). Fractals and Cancer. *Cancer Res*., Vol. 60., – P. 3683–3688.

Bezrukov, S. & Vodyanoy, I. (1997). Signal transduction across alamethicin ion channels in the presence of noise. *Biophys. J.*, Vol. 73, – P. 2456–2464.

Blanchard, J. & Blackman, C. (1994). Clarification and application of an ion paramagnetic resonance model for magnetic field interactions with biological resonance systems. *Bioelectromagnetics*, Vol. 15, – P. 217–238.

Blank, M. & Soo, L. (2001). Electromagnetic acceleration of electron transfer reactions. *J. Cell. Biochem.*, Vol. 81, – P. 278–283.

Blazsek, I. Innate chaos: I. (1992). The origin and genesis of complex morphologies and homeotic regulation. *Biomed. Pharmacother.*, Vol. 46, No. 5–7, – P. 219–235.

Boddie, A.; Frazer, J. & Yamanashi, W. (1987). RF electromagnetic field generation apparatus for regionally-focused hyperthermia, United States Patent N4674481 on 23. 06.1987.

Carrubba, S.; Frilot, C; Chesson, A. & Marino, A. (2007). Evidence of a nonlinear human magnetic sense. *Neuroscience*, Vol. 144, No. 1, – P. 356–367.

Carrubba, S.; Frilot, C.; Chesson, A.; Webber, C.; Zbilut, J. & Marino, A. (2008). Magnetosensory evoked potentials: consistent nonlinear phenomena. *Neurosci. Res.*, Vol. 60, No. 1, – P. 95–105.

Chang, K., Chang, W., Yu, Y, Shih, C. (2004). Pulsed electromagnetic field stimulation of bone marrow cells derived from ovariectomized rats affects osteoclast formation and local factor production. *Bioelectromagnetics*, Vol. 25, No. 2, – P. 134–141.

Chen, Q.; Tong, M. & Dewhirst F. (2004). Targeting tumor microvessels using doxorubicin encapsulated in a novel thermosensitive liposome. *Mol. Cancer Ther.*, Vol. 3, – P. 1311–1317.

Cramer, F. (1993). *Chaos and order. The complex structure of living systems*, VCH Verlagsgesellschaft, Weinheim.

Crosignani, B. & Tedeschi, A. (1976). Variation of the entropy of an electromagnetic field due to scattering. *Lettere Al Nuovo Cimento*, Vol. 17, No. 4, – P. 141–143.

De Mattei, M., Gagliano, N., Moscheni, C., Dellavia, C., Calastrini, C., Pellati, A., Gioia, M., Caruso, A. & Stabellini, G. (2005) Changes in polyamines, *c-myc* and *c-fos* gene expression in osteoblast-like cells exposed to pulsed electromagnetic fields. *Bioelectromagnetics*, Vol. 26, No. 3, – P. 207–214.

De Milito, A. & Fais, S. (2005). Proton pump inhibitors may reduce tumour resistance. *Expert Opinion on Pharmacotherapy*, Vol. 6, – P. 1049–1054.

Diniz, P., Shomura, K., Soejima, K. & Ito, G. (2002). Effects of pulsed electromagnetic field (PEMF) stimulation on bone tissue like formation are dependent on the maturation stages of the osteoblasts. Bioelectromagnetics.V. 23, Issue 5, Pages 398 – 405.

Dirks, P. (2008).Brain tumor stem cells: bringing order to the chaos of brain cancer. *J Clin Oncol*, Vol. 26, No. 17, – P. 2916–2924.

Ederer, M. & Gilles E. (2007). Thermodynamically feasible kinetic models of reaction networks. *Biophysical Journal*, Vol. 92, No. 6, – P. 1846–1857.

Emanuel, N. (1977). *Kinetics of experimental tumor processes*, Nauka, Moscow (in Russian).

Falk, M. & Issels, R. (2001). Hyperthermia in oncology. *Int J Hyperthermia,* Vol. 17, – P. 1–18

Franckena, M.; Fatehi, D.; de Bruijne, M.; Canters, R.; van Norden, Y.; Mens, J.; van Rhoon, G. & van der Zee, J. (2009). Hyperthermia dose-effect relationship in 420 patients with cervical cancer treated with combined radiotherapy and hyperthermia. *Eur J Cancer*, Vol. 45, No. 11, – P. 1969–1978.

Furusawa, C. & Kaneko, K. (2000). Origin of complexity in multicellular organisms. *Phys. Rev. Lett.*, Vol. 84, No. 26, Pt. 1, – P. 6130–6133.

Gaber, M. (2002). Modulation of doxorubicin resistance in multidrug-resistance cells by targeted liposomes combined with hyperthermia. *J Biochem Mol Biol Biophys*, Vol. 6, – P. 309–314.

Garte, S. (2003). Cancer epidemiology.Theory in carcinogenesis and epidemiology. *Journal of Epidemiology and Community Health,* Vol. 57, – P. 85.

Goldberg, A.; Allis, C. & Bernstein, E. (2007). Epigenetics: A landscape takes shape. *Cell*, Vol. 128, – P. 635–638.

Goodman, R. & Henderson, A. (1988). Exposure of salivary gland cells to low-frequency electromagnetic fields alters polypeptide synthesis. *Proc. Natl. Acad. Sci. USA*, Vol. 11, – P. 3928–3932.

Hardell, L. & Sage, C. (2008). Biological effects from electromagnetic field exposure and public exposure standards. *Biomed Pharmacother,* Vol. 62, No. 2, – P. 104–109

Hirata, M.; Kusuzaki, K.; Takeshita, H.; Hashiguchi, S.; Hirasawa, Y. & Ashihara, T. (2001). Drug resistance modification using pulsing electromagnetic field stimulation for multidrug resistant mouse osteosarcoma cell line. *Anticancer Res.*, Vol. 21, – P. 317–320.

Ivkov, R., De Nardo, S., Daum, W., Foreman, A., Goldstein, R., Nemkov, V. & De Nardo G. (2005). Application of high amplitude alternating magnetic fields for heat induction of nanoparticles localized in cancer. *Clinical Cancer Research,* Vol. 11, – P. 7093–7103.

Jin, Y., Wang, H., Cheng, Y. & Gu, H. (1998). Effects of static magnetic fields on free radical metabolism of human body. *Wei Sheng Yan Jiu*, Vol. 27, No. 2, –P. 97–99.

Jordan, E. & Balmain, K. (1968). *Electromagnetic waves and radiating system*, Prentice Hall, New Jersey.

Kattapuram, S., Rosol, M., Rosenthal, D., Palmer, W. & Mankin, H. (1999). Magnetic resonance imaging features of allografts. *Skeletal Radiol,* Vol. 28, No. 7, – P. 383–389.

Korn, G. & Korn, T. (1968). *Mathematical handbook for scientists and engineers*, MacGraw-Hill Book Company, New York.

Liboff, A. (1985). Cyclotron resonance in membrane transport. In: *Interaction between electromagnetic field and cells*, Chiabrera, A.; Nicolini, C. & Schwan, H. (Eds.), – P. 281–296, Plenum, New York.

Longo, I. & Ricci A. (2007). Chemical activation using an open-end coaxial applicator. *J. Microw. Power Electromagn. Energy*, Vol. 41, – P. 4–19.

Marino, A.; Wolcott, M.; Chervenak, R.; Heuil, F.; Nilsen, E. & Frilot, C. (2000). Nonlinear response of the immune system to power-frequency magnetic fields. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, Vol. 279, No. 3, – P. 761–768.

Marino, A.; Carrubba, S.; Frilot, C. & Chesson, A. (2009). Evidence that transduction of electromagnetic field is mediated by a force receptor. *Neurosci. Lett.*, Vol. 452, No. 2, – P. 119–123.

Martino, F. Alternative inductothermia in cancer. (1962). A confirmation with therapeutic applications of Warburg's theory. *Cancro,* Vol. 15, – P. 358–385.

Marmor, J. (1979). Interactions of hyperthermia and chemotherapy in animals. *Cancer Research*, Vol. 39, – P. 2269–2276.

Mittra, R. (1973). *International Series of Monographs in Electrical Engineering. Computer Techniques for Electromagnetics*, Pergamon Press, Oxford & New York.

Miyagi, N., Sato, K., Rong, Y., Yamamura, S., Katagiri, H., Kobayashi, K. & Iwata, H. (2000). Effects of PEMF on a murine osteosarcoma cell line: drug-resistant (P-glycoprotein-positive) and non-resistant cells. *Bioelectromagnetics,* Vol. 21, No. 2, – P. 112–121.

Molnar, J.; Thornton, B.; Thornton-Benko, E.; Amaral, L.; Schelz, Z. & Novak, M. (2009). Thermodynamics and Electro-Biologic Prospects for Therapies to Intervene in Cancer Progression. *Current Cancer Therapy Reviews*, Vol. 5, No. 3, – P. 158–169.

Moseley, H. (1988). *Non-ionizing radiation. Medical physics handbooks*, Adam Hilger, Bristol & Philadelphia.

Nikiforov, V. (2007). Magnetic induction hyperthermia. *Russian Physics Journal*, Vol. 50, No. 9, – P.913–924.

Nikolov, N.; Orel, V.; Smolanka, I.; Dzyatkovskaya, N.; Romanov, A.; Mel'nik, Y.; Klimanov M. & Chernish, V. (2008). Apparatus for Short-Wave Inductothermy "Magnetotherm", *Proceedings of NBC 2008*, pp. 294–298, Katushev, A.; Dekhtyar, Yu. & Spigulis, J. (Eds), Springer-Verlag, Berlin, Heidelberg.

Orel, V. & Dzyatkovskaya, N. (2000). Mechanoemission of blood and oncogenesis. In: *Biophotonics and coherent systems*, Beloussov, L.; Popp, F. & van Wijk, R. (Eds.), –P. 347–363, Moscow University Press.

Orel, V.; Grinevich, Y.; Dzyatkovskaya, N.; Danko, M.; Romanov, A.; Mel'nik, Y. & Martynenko, S. (2004). Spatial & Mechanoemission Chaos of Mechanically Deformed Tumor Cells. *Journal of Mechanics in Medicine & Biology*, Vol. 4, – P. 31–45.

Orel, V.; Kudryavets, Y.; Bezdenezhnih, N.; Danko, M.; Khronovskaya, N.; Romanov, A.; Dzyatkovskaya, N. & Burlaka, A. (2005). Mechanochemically activated doxorubicin nanoparticles in combination with 40MHz frequency irradiation on A-549 lung carcinoma cells. *Drug Delivery*, Vol. 12, – P. 171–178.

Orel, V.; Kozarenko, T.; Galachin, K.; Romanov, A. & Morozoff, A. (2007a). Nonlinear Analysis of Digital Images and Doppler Measurements for Trophoblastic Tumor. *Nonlinear Dynamics, Psyhology and Life Science*, Vol. 11, –P. 309–331.

Orel, V.; Dzyatkovskaya, N.; Romanov, A. & Kozarenko, T. (2007b). The effect of electromagnetic field and local inductive hyperthermia on nonlinear dynamics of the growth of transplanted animal tumors. *Experimental Oncology*, Vol. 29, No. 2, – P. 156–158.

Orel, V.; Nikolov, N.; Dzyatkovskaya, N.; Romanov, A.; Melnik, Y.; Dunaevsky V. & Dzyatkovskaya, I. (2008). Influence of change of spatial nonuniformity of the electromagnetic field on transformation of radio-waves and thermal characteristics of phantoms and Lewis lung carcinoma. *Physics of the Alive*, Vol. 16, No. 2, – P. 92–98 (In Ukrainian).

Orel, V.; Dzyatkovskaya, I.; Nikolov, N.; Romanov, A.; Dzyatkovskaya, N.; Kulik, G.; Todor, I.; Chranovskaya, N. & Skachkova, O. (2009). The influence of spatially nonuniform electromagnetic field on antitumor activity of cisplatin during treatment of resistant substrain of Lewis lung carcinoma. *Ukrainian Radiology Journal*, Vol. 17, –P. 72–77 (in Ukrainian).

Pasquinelli, P., Petrini, M., Mattii, L., Galimberti, S., Saviozzi, M. & Malvaldi G. (1993). Biological effects of PEMF (pulsing electromagnetic field): an attempt to modify cell resistance to anticancer agents. *J Environ Pathol Toxicol Oncol, Vol.* 12, – P. 193–197.

Polanyi, J. (1987). Some concepts in reaction dynamics. *Science*, Vol. 236, – P. 680–690.

Purcell, E. (1985). *Electricity and Magnetism,* McGraw-Hill, New York.

Reszka, K., Mc Cormick, M. & Britigan, B. (2001). Peroxidase- and nitrite-dependent metabolism of the anthracycline anticancer agents daunorubicin and doxorubicin. *Biochemistry,* Vol. 40, – P. 15349–15361.

Roemer, R. (1999). Engineering aspects of hyperthermia therapy. *Annual Review of Biomedical Engineering,* Vol. 1, – P. 347–376.

Rubin, H. (1984). Cancer as a dynamic developmental. *Cancer Res.*, Vol. 45, – P. 2935–2942.

Ruiz-Gómez M., de la Peña, L., Prieto-Barcia, M., Pastor, J., Gil, L. & Martínez-Morillo, M. (2002). Influence of 1 and 25 Hz, 1.5 mT magnetic fields on antitumor drug potency in a human adenocarcinoma cell line. *Bioelectromagnetics*, Vol. 23, No. 8, –P. 578 – 525.

Sedivy, R. & Mader, M. (1997). Fractals, chaos, and cancer: do they coincide? *Cancer Invest.*, Vol. 15, – P. 601–607.

Schneider, B. & Kulesz-Martin, M. (2004). Destructive cycles: the role of genomic instability and adaptation in carcinogenesis. *Carcinogenesis*, Vol. 25, No. 11, – P. 2033–2044.

Shen, J.; Zhang, W.; Wu, J. & Zhu, Y. (2008). The synergistic reversal effect of multidrug resistance by quercetin and hyperthermia in doxorubicin-resistant human myelogenous leukemia cells. *Int J Hyperthermia*, Vol. 24, No. 2, – P. 151–159.

Sole, R.; Garsia, I. & J.Costa. (2006). Spatial Dynamics in Cancer. In: *Complex Systems Science in Biomedicine Series: Topics in Biomedical Engineering,* Deisboeck, T. & Kresh, J.(Ed.), – P. 557–572, International Book Series, Springer US.

Solyanik, G.; Todor, I.; Kulik, G. & Chekhun, V. (1999). Selective mechanism of the emergence of Guerin`s carcinoma resistance to doxorubicin. *Experimental oncology*, Vol. 21, – P. 264–268.

Song, C.; Park, H.; Lee, C. & Griffin, R. (2005). Implications of increased tumor blood flow and oxygenation caused by mild temperature hyperthermia in tumor treatment. *Int. J. Hyperthermia*, Vol. 21, – P. 761–767.

Waliszewski, P.; Molski, M. & Konarski, J. (1998). On the holistic approach in cellular and cancer biology: nonlinearity, complexity, and quasi-determinism of the dynamic cellular network. *J. Surg. Oncol.*, Vol. 68, No. 2, – P. 70–78.

Weaver, J.; Vaughan, T. & Martin, G. (1999). Biological effects due to weak electric and magnetic fields: the temperature variation threshold. *Biophys J.*, Vol. 76, No. 6, – P. 3026–3030.

Weller, A.; Anton, M.; Geiger, J.; Hirsch, M.; Jaenicke, R.; Werner, A. & Nührenberg, C. (2001). Survey of magnetohydrodynamic instabilities in the advanced stellarator. *Phys. Plasmas*, Vol. 8, – P. 931.

Yarosh, D. Why is DNA damage signaling so complicated? (2001). Chaos and molecular signaling. *Environmental and Molecular Mutagenesis*, Vol. 38, No. 2–3, – P. 132–134.

Advanced Computational Approaches for Predicting Tourist Arrivals: the Case of Charter Air-Travel

Eleni I. Vlahogianni, Ph.D. and Matthew G. Karlaftis, Ph.D.
Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, 5, Iroon Polytechniou Str., Zografou Campus, Athens 157 73, Greece

1. Introduction

Tourism is one of the major industries profiting various sectors of the economy, such as the transportation, accommodation, entertainment and so on. According to the World Tourism Organization (2008), international tourism grew at around 5% during the first four months of the year 2008. Fastest growth was observed in the Middle East, North-East and South Asia, and Central and South America. Even though, uncertainty over the global economic situation is affecting consumer confidence and could hurt tourism demand, for 2008 as a whole, UNWTO maintains a cautiously positive forecast. Moreover, international trends show that tourists are becoming more discerning in their choice of destinations and, therefore, becoming less predictable and more spontaneous in terms of their consumption patterns (Burger et al. 2001).

Air transportation is probably the most important mode for international travel and leisure. A typical characteristic of air tourism in Europe is the extensive use of non-scheduled/charter flights and the existence of low-cost carriers in the leisure travel market, that account for 8% of passengers and 3% or revenues in the aviation industry (Dresner 2006). Non-scheduled demand is typical in Mediterranean countries where connections are essentially touristic and characterized by non-scheduled services.

In this type of air travel, the ability to accurately predict tourist arrivals is of importance in the successful management and operation of the airport facilities, as well as the adjacent transportation network. Yet, the literature has little to offer in modeling demand stemming from non-scheduled flights, as such series exhibit seasonality, intense variability and inherent unpredictability.

This paper develops and tests advanced computational approaches in order to predict non-scheduled/charter international tourist demand. The computational challenges that may arise in such a problem are twofold: first, to treat seasonal and stochastic characteristics of non-scheduled tourist demand, and, second, to develop models that consider past tourist demand characteristics. This paper focuses on developing ARFIMA models that consider both non-stationarity and long-term memory effects in the auto-regressive process and temporal neural networks with advance genetically optimized characteristics that treat both nonlinearity and non-stationarity.

2. Motivators and prediction of non-scheduled air-travel demand

A major motivator for the emergence and growth of non-scheduled air travel has been the low-cost carriers (LCC) and their prevalence in global aviation. From the period after 9/11 period that caused a decreasing trend in the airline travel demand, global aviation and travel demand, particularly in Europe and the Mediterranean Region LCCs offered an attractive alternative for price-sensitive clients during the tight economic times. Whereas traditional airlines have concentrated on large cities and major airports, low-cost airlines have turned to under-utilized airports at some distance from the main population centers embracing a business model much different in its customer base, air network, and provision of services by focusing on the more cost-sensitive leisure travel and working in a way that traditional airlines cannot (Barrett 2000).

LCC market providing point-to-point (rather than hub-based) service owes its growth not only to low-cost service, but also to the ability to focus on customer segments not emphasized by larger carriers; European low-cost leaders Ryanair and EasyJet, for instance, focus on providing air services for travelers seeking to visit friends and relatives. By focusing on these groups, LCC have demonstrated an ability to grow the overall passenger market, particularly on routes with strong tourist appeal (Dennis 2004).

Literature emphasizes the role of LCC in the development of multiple airport systems and the emergence of secondary airports (Bonnefoy & Hansman 2004). LCC appeal to secondary airport is in that they provide reduced congestion and lower cost, while still providing access to key population centers.

The shift to secondary airports, along with the reduced gap between charter flights and "no-frills" / budget flights have significant impact on the volatility of traffic for the entire airport system; literature indicates that periods of high volatility and uncertainty in demand exist during the developmental phases of secondary airports that can last up to 20 year after the opening of such facilities (de Neufville, 1995).

Regarding leisure airline traffic, the ability to provide custom-made services to tourists has been shown to be critical. Tourists increasingly expect to experience a personalized and close to their life-style service (Graham 2006). A characteristic example of charter airports is Greece where approximately 80% of the total tourist arrivals every year are accommodated by air transportation. The importance of non-scheduled international arrivals is depicted in Figure 1 that depicts annual evolution of total arrivals for 1989 and 2006 period, along with the evolution of non-scheduled and scheduled international arrivals. As can be observed, for the period after 2001, nearly 70% of air-travel arrivals concern international flights and 62% of the international arrivals are accommodated by non-scheduled flights.

From a methodological standpoint, although the prediction of tourist demand has been extensively treated (a review of approaches can be found in Law et al. 2007, Song & Li 2008) little has been done towards the prediction of non-scheduled arrivals. Summarizing the methodologies implemented to date for to tourist demand prediction, both econometrics and other computational methods have been applied and compared. Law et al. (2007) state that, comparing classical econometric prediction techniques that are highly exploited but with marginal improvement to modeling touristic demand, the incorporation of data mining techniques has led to some "ground breaking outcomes".

Moreover, several papers on tourism forecasting problems report neural networks as having better performance than classical statistical techniques, such as ARIMA models, exponential smoothing and so on (Law and Au 1997, Law 2000, Burger et al. 2001, Kim et al. 2003, Cho 2003). These studies compare advanced computational approaches that have enhanced capabilities in modeling nonlinear characteristics (for example neural networks) with simple linear and stationary approaches such as the ARIMA models. Quite recently, hybrid ARIMA and simple static neural networks, as well as mixtures of static neural network models have also been found to perform better that classical time-series approaches (Aslanargun et al. 2007).

Regarding modeling of non-scheduled demand, previous work has applied regression models to predict charter international arrivals to major Greek airports and has highlighted that although there is uncertainty and variability in their evolution, historical data can be used to provide good predictions (Karlaftis and Papastavrou 1998). However, no previous work has been conducted in the direction of predicting non-scheduled international arrivals in secondary airports with intense seasonal characteristics.

3. Computational approaches

3.1 Fractionally integrated autoregressive moving average processes

Commonly applied AR(I)MA models are able to describe processes that are covariance stationary *I(0)* or non-stationary through differencing *I(1)*. It has been observed that the erroneous consideration of having a unit root leads to models with inflated estimates of the moving average component (Box-Steffensmeier and Smith, 1998). In order to account for long memory processes Fractional integration is introduced to autoregressive processes to account for the processes that are neither I(0) or I(1) in the form of the differentiation operator (Baillie 1999):

$$(1-L)^d = \left\{1 - dL - d(d-1)\frac{L^2}{2!} - d(d-1)\frac{L^3}{3!} - \ldots\right\} \tag{1}$$

In the conditional mean, the fractionally integrated autoregressive moving average process of orders *p* and *q* – ARFIMA(*p*,*d_m*,*q*) introduced by Granger and Joyeux (1980) and Hosking (1981) is represented by the following equation:

$$(1-L)^{d_m}\Psi(L)(y_t - \mu) = \Theta(L)\varepsilon_t \tag{2}$$

$$\varepsilon_t = \sigma_t z_t \quad z_t \sim N(0,1) \tag{3}$$

where μ is the unconditional mean of y_t, $\Psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \ldots - \psi_p L^p$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$ are the AR and MA polynomials having all roots outside the unit cycle, while innovations ε_t are i.i.d distributed with σ_t^2 being the conditional variance and a positive, time-varying, and measurable function with respect to the information set, which is available at time *t*-1 (Baillie et al. 2002). The differentiation parameter (d_m) is associated with the following statistical properties of a (time) series (Hosking 1981, Odaki 1993):

- For every region where $d_m < 1/2$, then y_t is stationary,
- When $-1 < d_m < -1/2$, the series exhibits invertibility,

- When $-\frac{1}{2} \le d_m < 0$, the stationary process y_t is antipersistent,[1]
- When $d_m = 0$, the stationary process y_t has short memory and is mean reverting,
- When $0 < d_m \le \frac{1}{2}$, y_t is fractionally integrated and exhibits long memory,
- When $\frac{1}{2} < d_m < 1$, the process y_t is mean-reverting, but the stationarity property cannot be verified and,
- When $d_m = 1$, y_t is a unit root process.

Fractionally integrated processes are significant in dealing with two issues: first, data is being modeled more precisely, as the knife-edged restriction of an *I(0)* or *I(1)* process is avoided and both long term persistence and, second, short-term correlation structure of a series can be modeled (Hosking 1981).

3.2 Temporal genetically optimized neural networks

Temporal Neural Networks can be considered as an extension of the static Multi-layer Perceptrons (MLP) that has been extensively applied to touristic demand prediction. They differ from the commonly used MLPs in that they incorporate memory mechanisms in their structure that can be limited to the input layer or extend to the entire network. The memory acts as a time-series reconstruction module with the aim to embed the scalar series $S(t)$ to a vector $\mathbf{S}(t) = \{S(t-\tau),...,S(t-(m-1)\tau)\}$ in an *m*-dimensional vector space known as Phase Space, where *τ* is the time delay of and *m* is the dimension.

We implement a neural network called time-lagged neural networks (TLNN) with a complex Gamma memory mechanism in the input layer and the hidden layer (de Vries and Principe 1992). Moreover, in order to develop a fully non-stationary model we set the network to predict under the iterative consideration: Given the time-series of a variable a single step ahead model is constructed to produce a prediction $\hat{S}(t)$ at time *t* that is then fed backwards to the network and is used as new input data in order to produce the next step $\hat{S}(t+1)$ prediction at *t*+1:

$$\hat{S}(t+1) = \left\{\hat{S}(t), S(t), S(t-1)...\right\} \tag{4}$$

The training of TLNN under iterative consideration feeds back the prediction at time *t*+1 and utilizes it as an input for the generation of next prediction step *t*+2. The training in the specific iterative neural network model is conducted via the temporal back-propagation algorithm known as Back-propagation to time (BPTT) (Webros 1990); all weights are duplicated spatially for an arbitrary number of time steps *τ*; as such, each node that sends activation to the next has *τ* number of copies as well. For a training cycle *n*, the weight update is given by the following equation (Haykin 1999):

$$\mathbf{w}_{ji}(n+1) = \mathbf{w}_{ji}(n) + \eta\delta_j(n)\mathbf{x}_i(n) \tag{5}$$

where, $\mathbf{w}_{ji}(n+1)$ and $\mathbf{w}_{ji}(n)$ are the weights of the *i*-th synapse of the neuron *j* at training cycle *n*+1 and *n* respectively, *η* is the learning rate, **x***i*(*n*) (i=1,2,…n) is the input vector and $\delta_j(n)$ is given by:

[1] Anti-persistence is a property of an ACF that exhibits slow decay, but the original series are not characterized by the long memory property; rather, the autocorrelations (in the ACF) alternate in signs.

$$\delta_j(n) = \begin{cases} e_j(n)\phi'(\upsilon_j(n)),\ j \text{ neuron in the output layer} \\ \phi'(\upsilon_j(n))\sum_{r\in \mathrm{A}} \mathbf{\Delta}_r^{\mathrm{T}}(n)\mathbf{w}_{rj},\ j \text{ neuron in the hidden layer} \end{cases} \qquad (6)$$

where, $e_j(n)$ is the network's error, ϕ is the nonlinear activation function. Moreover, if A is a set of all neurons whose inputs are fed by the j neuron in the hidden layer is a forward manner, then $\upsilon_j(n) = \sum_{i=1}^{m} \mathbf{w}_{ji}\mathbf{x}_i(n) + b_j$ is the induced local field of neuron r that belongs to the A and $\mathbf{\Delta}_r^{\mathrm{T}}(n) = [\delta_r(n), \delta_r(n+1), ..., \delta_r(n+m)]^T$ is the local gradient vector.

The iterative neural network approach introduced provides a fully non-stationary and nonlinear environment for treating time series problems. However, regardless of being static or dynamic, neural networks suffer from the lack of an automatic manner to self-optimization mainly with respect to their structure (number of hidden units) and learning parameters. Recently, genetic algorithms have gained significant interest as they can be integrated to the neural network training to search for the optimal architecture without outside interference, thus eliminating the tedious trial and error work of manually finding an optimal network. Genetic algorithms are based on three distinct operations: selection, cross-over and mutation (Mitchell 1998); these operations run sequentially in order for a fitness criterion (in the specific case the minimization of the cross-validation error) to converge. Details for the specific optimization approach can be found in Vlahogianni et al. (2005).

4. Case study: greek island airports

We focus on the influence of Non-Scheduled International (NSI) arrivals to the secondary airports and a prediction of their temporal evolution. Three case studies from Greek island secondary airports are evaluated: Heraklion (Crete), Kefalonia and Rhodes. All three cases exhibit significant demand during the peak summer period; however, these case studies differ in the overall demand levels, as well as their seasonal arrival characteristics. As can be observed in Figure 2, where the evolution of arrivals (passengers per year) and flights per year and per airport for the period of 1999-2006 is depicted, Kerkyra is characterized by low volumes, whereas Heraklion and Rhodes exhibit high demand during the year. The difference is in the volume of the NSI arrivals; as can be seen in Figure 3, where monthly arrival variation is depicted for all airports tested, Kerkyra and Rhodes have significantly more acute monthly variation, reaching extremely low NSI demand during the off-peak periods.

The analysis to follow will, first, focus on revealing long-term memory features in the manner NSI arrivals evolve in time and, second, search for similarities or differences in the dynamics of NSI arrivals across the airports selected with different demand distributions. Third, advanced neural network predictors will be developed that will apply the iterative approach in order to learn to approximate the dynamics of NSI arrivals; models will be developed for all the three airports and compared to each other.

4.1 Fractional dynamics in NSI arrivals

Several ARFIMA models were fitted to the available time –series in order to test whether there exist fractional dynamics in the evolution of non-scheduled international arrivals. The models are fitted to both three study airport, as well as to the pooled data, as well as data from the peak (months from May to September). Moreover, in the same datasets *I(1)* ARIMA processes are also fitted in order to compare the estimated autoregressive and moving average parameters from ARFIMA and ARIMA models. The choice of the best fitting model is done via Akaike's ($-2\frac{LogL}{n}+2\frac{k}{n}$, where logL is the log likelihood value, n is the number of observations and k the number of estimated parameters) and Schwartz's ($-2\frac{LogL}{n}+2\frac{\log k}{n}$) criteria. Furthermore, the Jarque-Bera test (JB test) goodness-of-fit test measuring the of departure from normality, $Q^2(i)$ statistics that indicate the possible existence of serial correlation in the standardized residuals, as well as the LM ARCH statistics that test the null hypothesis of no ARCH effect in the series tested are also presented; the above test will provide information of the quality of the ARFIMA models developed.

Results for the best fitted ARFIMA models are shown in Tables 1 to 3(parameter estimates depicted in the tables are significantly different from zero at the 1% significance level). Interestingly, for all case studies the fractional dynamics are similar. NSI arrivals in all airports tested are found to be best described by a fractionally integrated ARMA process with p=1 (autoregressive term) and q=1 (moving average term). Parameter d is found to vary between 0.24 and 0.46 indicating that NSI arrivals regardless of study period (peak or off-peak), as well as of the airport tested, exhibit long-term memory (for more details on the memory properties see Washington et al. 2003). We observe that the ARFIMA modeling results exhibit an apparent "inability" to approximate the monthly variability of NSI arrivals, particularly at low demand levels (off-peak months) (Figure 4).

4.2 Iterative predictions of NSI arrivals using temporal neural networks

For iterative predictions, the specifications of the TLNN are shown in Table 4. As can be observed, the depth of the Gamma memory of the TLNN (parameters τ and m) are genetically optimized during the learning, along with the number of hidden units h in the hidden layer and the learning rate η and momentum μ of back-propagation. The available data is separated into three subsets in order to test the training (cross-validation) and then test the performance of the network (testing). Moreover, the genetic algorithm optimization specifications are also depicted on Table 4; a roulette selection method is applied in order to select the parents according to their fitness. Moreover, the probabilities of cross-over and mutation are to be equal to 0.9 and 0.09 respectively, following literature that indicates that crossover should usually be selected at high values and mutation should approximate the inverse of the number of chromosomes (population) and be much lower than the crossover probability to avoid permutation (Gen and Cheng, 2000).

Results concerning the optimization of the look-back time window, or else the depth of the memory of the iterative temporal neural networks, are shown in Table 5. Interestingly, the required data to produce accurate predictions – as determined by the genetic optimization of the parameters τ and m during learning – differ between Heraklion airport and the rest of the cases examined. The recurrence of the dynamics in the Heraklion case is every 6 months, whereas NSI arrivals of Kerkyra and Rhodes are affected by data from up to 4 months in the past.

Results of the predictions (test set) using TLNN are seen in Table 6; predictions for the same period using ARFIMA (averaged for the three airports) are also illustrated. As can be observed, the TLNN has, overall, better accuracy that is evident both in the high and low demand periods in all three airport cases examined. The averaged behavior of the ARFIMA and TLNN models developed with respect to the actual and predicted NSI arrivals is graphically represented in Figure 5. Interestingly, the accuracy of predictions seems to decrease significantly in the case of low demand time periods, such as months between November and March, where touristic arrivals to Greek islands are, in general, significantly lower than the ones during summer months. The decreased accuracy in the case of Kerkyra indicates the existence of significant stochasticity in the manner in which arrivals evolve in low demand and off-peak period cases.

5. Discussion and conclusions

A large portion of tourist demand is conducted by air. Several air links can have intrinsic characteristics concerning the touristic demand evolution with strong non-stationary and seasonal characteristics. In this paper we implemented recent data mining techniques to model tourist demand and developed two advanced models of time-series prediction: a fractionally integrated autoregressive moving average model (ARFIMA) and a temporal genetically optimized iterative neural networks. These models originate from different methodological backgrounds and aim to evaluate different statistical properties of tourist demand (such as the existence of long-term memory, the parameters of memory depth for predictions and so on). To evaluate the proposed methodologies, three cases studies were examined that encompass three secondary airport located in the Greek Islands which exhibit different yearly and monthly demand distributions.

In terms of prediction accuracy, the advanced form of temporal neural networks implemented seems to outperform the ARFIMA model. This applies to both high and low tourist demand periods. In terms of the knowledge acquired by the modeling, both approaches revealed very interesting results; the fractional dynamics observed in both the pooled data and the peak demand period, show that the tourist arrivals are not always stationary or best described as most frequently - assumed - by ARIMA models. The fractionally integrated processes fitted to the available data showed that all case studies examined have similar fractional dynamics and exhibit long term memory. This finding has significant implications to the process of modeling NSI arrivals, as it suggests the persistence of the effect of several socio-economic issues to the evolution of NSI arrivals.

Moreover, the iterative approach to predicting NSI arrivals showed significant improvement to the prediction accuracy. The advanced genetic optimization implemented with regards to the look-back time window of the TLNN shows that the past could be utilized to predict the evolution of tourist demand. Nevertheless, the differences in the memory depth of the three TLNN models developed to approximate the dynamics of NSI arrivals in the three airports indicates the stochasticity of the temporal evolution of NSI arrivals during periods of low volume that significantly affect the accuracy of predictions.

Finally, lack of prediction accuracy during transitional conditions reveals that, as expected, the demand evolution can have multiple causal dimensions that need to be considered in an effective methodological framework that could integrate both the temporal and causal/relational characteristics of other possible influential variables in the prediction process. Our ongoing work focuses on extending the present methodological framework to iterative neural network prediction that incorporates other socio-economic data to develop influential relationships and evaluate whether they can improve predictability during periods of stochasticity in tourist demand.

6. References

Aslanargun, A., Mammadov, M., Yazici, B. and Yolacan, S. (2007), Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting, *Journal of Statistical Computation and Simulation*, 77(1), 29-53.

Baillie, R.T., Han, Y.W. and Kwon, T.-G. (2002) Further long memory properties of inflationary shocks, South. Econ. J. 68 496–510.

Barrett, S. D. (2000). Airport competition in the deregulated European aviation market. *Journal of Air Transport Management*, 6(1), 13 - 27.

Bonnefoy, P., and Hansman, R. (2004). Emergence and Impact of Secondary Airports in the United States. *Proceedings of the 4th AIAA ATIO Forum in Chicago, Illinois*. Retrieved August 7, 2007.

Burger, C.J.S.C., Dohnal M., Kathrada M., Law R. (2001), A practitioners guide to time-series methods for tourism demand forecasting: a case study of Durban, South Africa, *Tourism Management*, 22, 403-409.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting, Tourism Management 24, 323–330.

de Neufville, R. (1995). Management of multi-airport systems: A development strategy. *Journal of Air Transport Management, 2* (2), 99-110.

Dennis, N. (2004). Can the European low cost airline boom continue? Implications for regional airports: *The 44th European Congress of Regional Science Association*. Porto, Portugal. Retrieved online on January 4, 2007.

Dresner, M. (2006). Leisure versus business passengers: Similarities, differences, and implications, *Journal of Air Transport Management*, 12, 28–32.

Graham, A. (2006). Have the major forces driving leisure airline traffic changed?, *Journal of Air Transport Management*, 12(1), 14-20.

Granger, C.W.J. and Joyeux, R., (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15– 29.

Haykin, S. (1999), *Neural Networks: A comprehensive foundation*, Prentice Hall Upper Saddle River, NJ.

Hosking, J.R.M. (1981) Fractional differencing, *Biometrika*, 68, 165–176.

Karlaftis, M. G. and Papastavrou, J. D. (1998), Demand characteristics for charter air-travel, *International Journal of Transportation Economics*, XXV(1), 19-35.

Kim, J., Wei, S. and Ruys, H. (2003), Segmenting the market of West Australian senior tourists using an artificial neural network, *Tourism Management*, 24, 25–34.

Law, R. (2000), Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting, *Tourism Management*, 21, 331–340.

Law, R. and Au, N. (1999), A neural networks model to forecast Japanese demand for travel to Hong Kong, *Tourism Management*, 20(1), 89–97.

Law, R., Mok, H. and Goh, C. (2007), Data Mining in Tourism Demand Analysis: A Retrospective Analysis, Advanced Data Mining and Applications, Book Series Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 508 – 515.

M. Gen and R.W. Cheng, Genetic Algorithms and Engineering Optimization, John Wiley & Sons, New York (2000).

Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT Press, ISBN: 0262631857.

Song, H. and Li, G. (2008), Tourism demand modeling and forecasting—A review of recent research, *Tourism Management,* 29, 203–220.

Vlahogianni, E. I. Karlaftis M. G. and Golias, J. C. (2005), Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach, *Transportation Research C,* 13(3), 211-234.

Washington, S.P. Karlaftis M.G. and Mannering, F.L. (2003), *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC Press, London.

Webros, P. J. (1990), Backpropagation Through time: What it does and How to do it, *IEEE Proceedings*, 78(10), 1550-1567.

World Tourism Organization – UNWTO (2008). "Firm Tourism Demand - Advanced Results", World Tourism Barometer, June, accessed at: http://www.unwto.org/media/news/en/press_det.php?id=2532.

		Pooled	Peak Period
		p=1,q=1	p=1,q=1
Degree of differentiation	d_m	0.24	0.46
AR polynomial coefficients	ψ_1	0.66	0.02
	ψ_2	-	-
MA polynomial coefficients	θ_1	0.52	0.55
	θ_2	-	-
	θ_3		-
	θ_4		-
	θ_5		-
Log-likelihood		-2622.36	-1079.26
JB Test Null: normality		2.02	1.42
$Q^2(7)$ Null: serial independence		136.25**	66.18**
LM ARCH (1) Null: no ARCH effect		1.41	1.32

* rejection at 5% significance level

** rejection at 1% significance level

Table 1. Estimation Results for the ARFIMA(p,d_m,q) models for the Heraklion airport.

		Pooled	Peak Period
		p=1,q=1	p=1,q=1
Degree of differentiation	d_m	0.15	0.31
AR polynomial coefficients	ψ_1	0.66	0.05
	ψ_2	-	-
MA polynomial coefficients	θ_1	0.35	0.48
	θ_2	-	-
	θ_3		-
	θ_4		-
	θ_5		-
Log-likelihood		-2588.14	-1002.80
JB Test Null: normality		4.43	1.24
$Q^2(7)$ Null: serial independence		145.25**	64.54**
LM ARCH (1) Null: no ARCH effect		1.65	0.03

* rejection at 5% significance level

** rejection at 1% significance level

Table 2. Estimation Results for the ARFIMA(p,d_m,q) models for the Kerkyra airport.

		Pooled	Peak Period
		p=1,*q*=1	*p*=1,*q*=1
Degree of differentiation	d_m	0.34	0.37
AR polynomial coefficients	ψ_1	0.67	0.05
	ψ_2	-	-
MA polynomial coefficients	θ_1	0.43	0.58
	θ_2	-	-
	θ_3		-
	θ_4		-
	θ_5		-
Log-likelihood		-2689.31	-1017.48
JB Test Null: normality		3.48	2.48
$Q^2(7)$ Null: serial independence		122.52**	75.67**
LM ARCH (2) Null: no ARCH effect		0.82	0.10

* rejection at 5% significance level

** rejection at 1% significance level

Table 3. Estimation Results for the ARFIMA(p,d_m,q) models for Rhodes airport.

		Specifications
	DATA	TR–CV–TE *: 60%-20%-20%
	Structure	Input layer: Gamma memory (genetically optimized memory depth) 1 hidden layer (genetically optimized number of hidden units *h*)
	Learning	Back-propagation
Genetic algorithm optimization	Chromosome	$h \in [5,15]$, $\gamma \in [0.01 - 0.1]$, $\mu \in [0.5 - 0.9]$, $\tau \in [1,5]$, $m \in [1,12]$ **
	Fitness function	Mean square error (cross-validation set)
	Selection	Roulette
	Cross-over	Two point (*p*=0.9)
	Mutation	Probability *p*=0.09

* *Training - Cross-validation - Testing*

** *h: neurons in hidden layer, γ: learning rate, μ: momentum, τ: time delay, m:dimension*

Table 4. Data and neural network specifications for iterative short-term prediction.

	Pooled NSI Arrivals	
	τ	m
Heraklion	1	6
Kerkyra	1	4
Rhodes	1	4

Table 5. Estimates of the depth of the Gamma memory (parameters τ and m) of the genetically-optimized TLNNs for the three cases.

	Pooled Data	Peak Demand Period
GA-TLNN*		
Heraklion	17%	2.8
Kerkyra	26%	3.4
Rhodes	18%	3.2
ARFIMA *(average over cases tested)*	37%	8.2

**genetically optimized TLNN*

Table 6. Mean Absolute Percent Error of predictions using ARFIMA and genetically optimized TLNN.

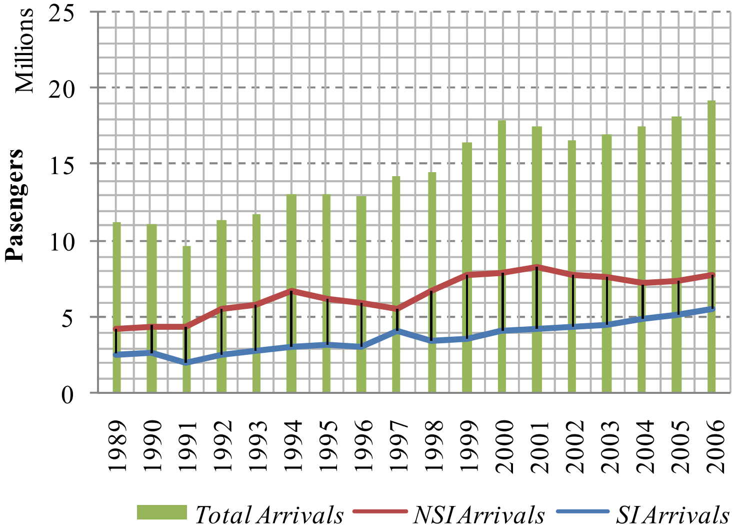
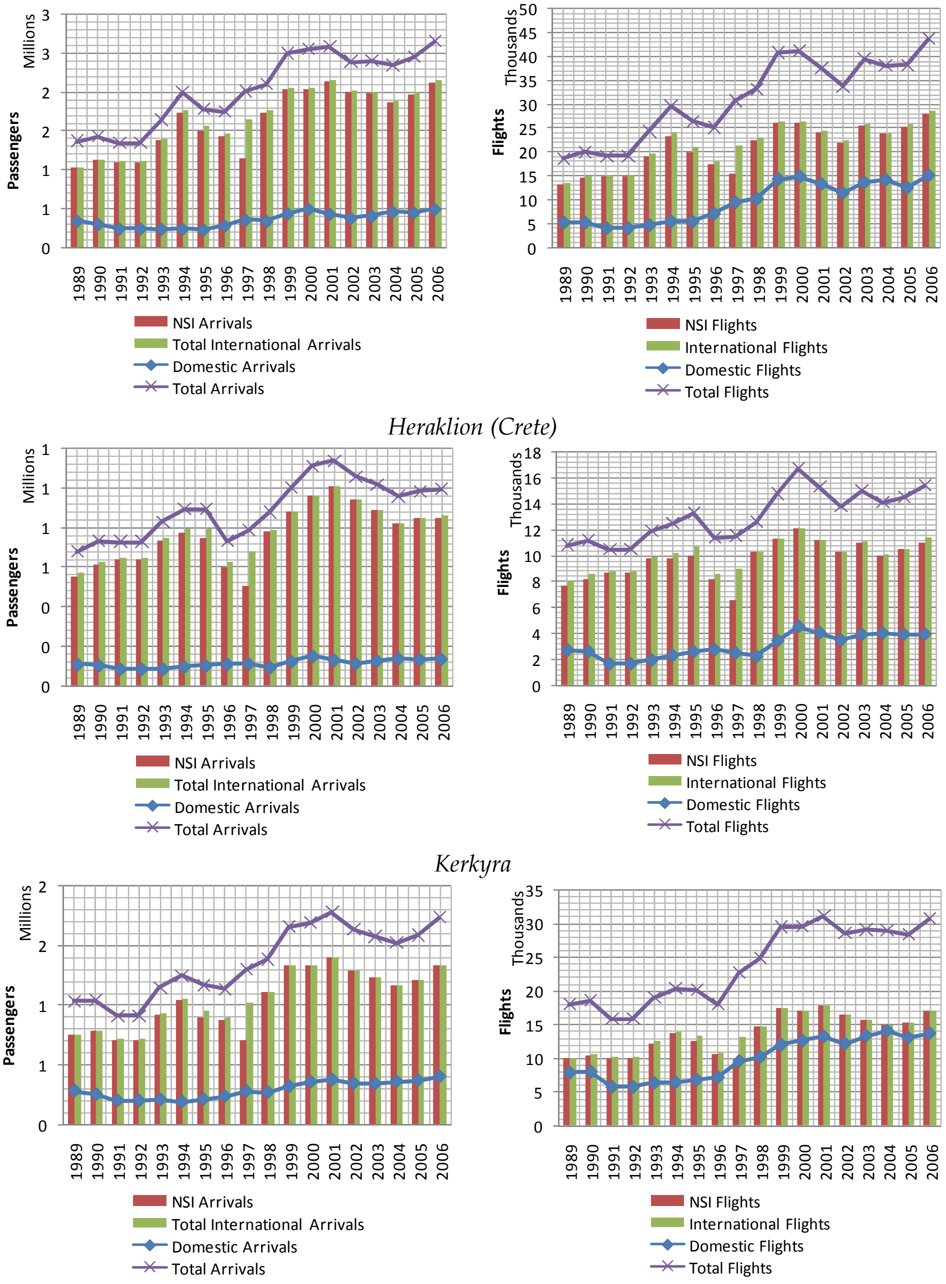


Fig. 1. Yearly evolution of the total arrivals, non-scheduled international arrivals (NSI Arrivals) and scheduled international arrivals (SI Arrivals) for the Greek airports.

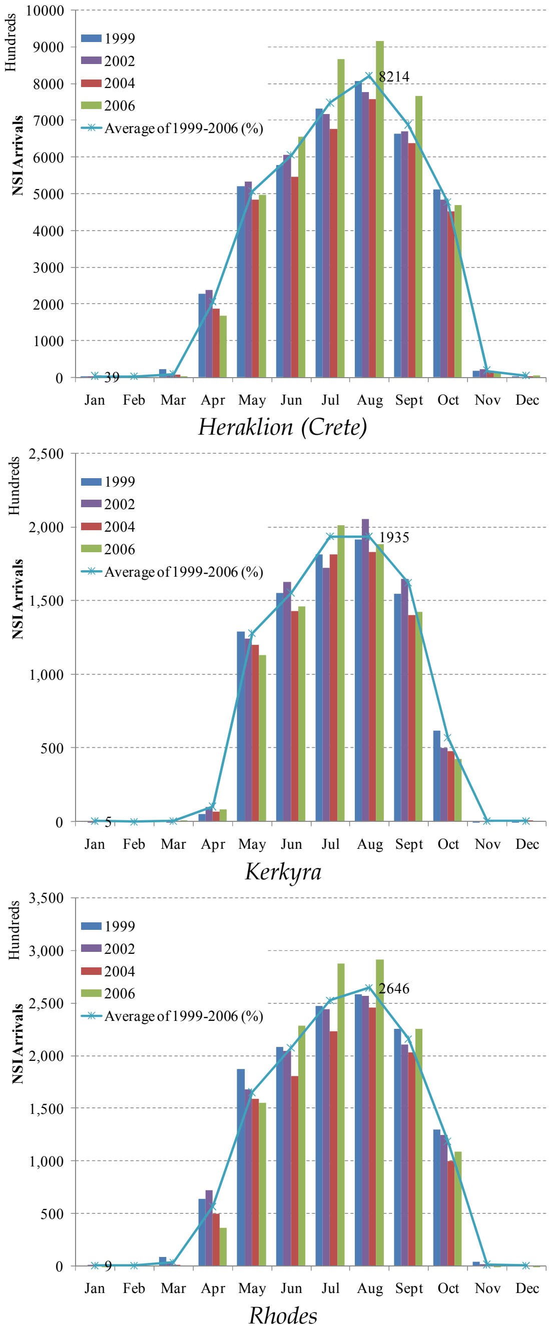


Heraklion (Crete)

Kerkyra

Rhodes

Fig. 2. Evolution of arrivals (passengers per year) and flights per year for the period of 1999-2006.



Heraklion (Crete)

Kerkyra

Rhodes

Fig. 3. Monthly variation of non-scheduled international arrivals in Rhodes for the period between 1999 and 2006.

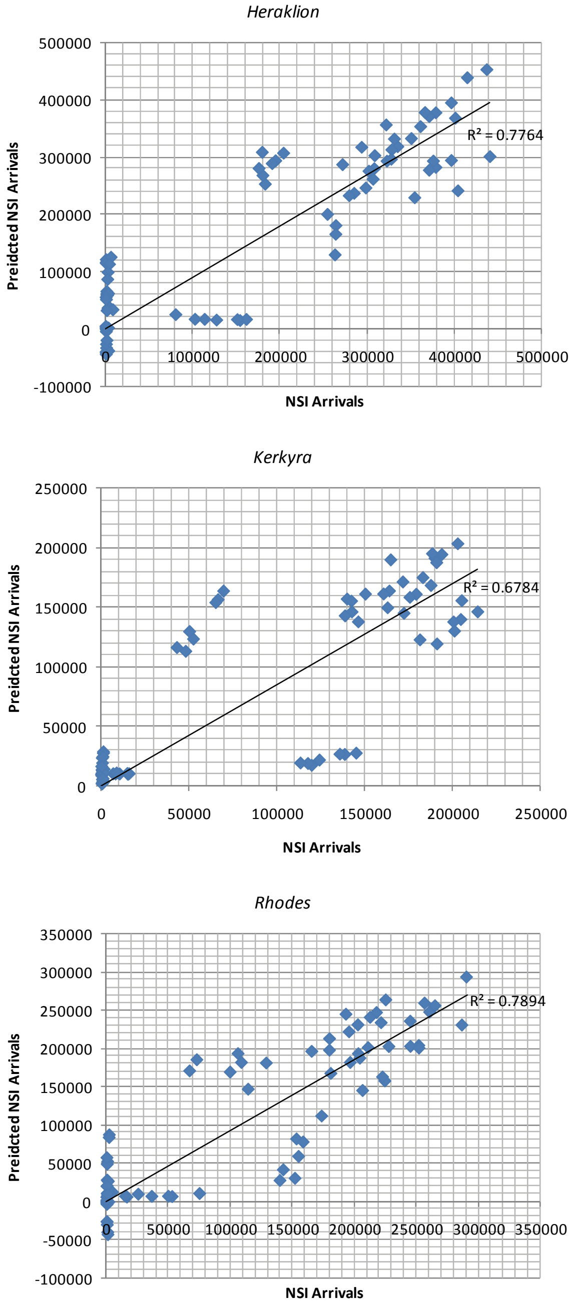


Fig. 4. Scatter plots of actual versus predicted values of NSI arrivals for the three airports.

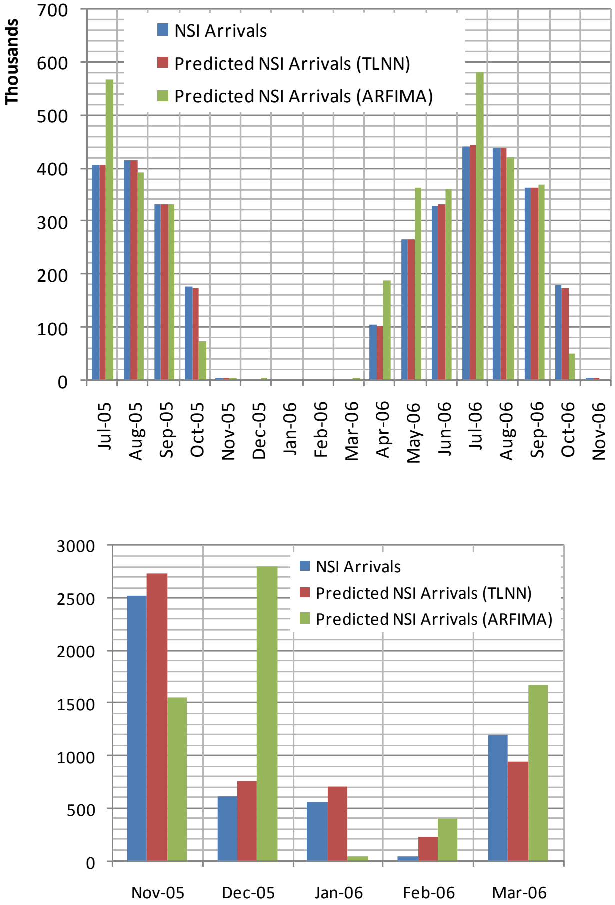


Fig. 5. Predictions using the ARFIMA and genetically optimized TLNN. Results from the three case study airports are aggregated both for ARFIMA and TLNN.

14

A Nonlinear Dynamics Approach for Urban Water Resources Demand Forecasting and Planning

Xuehua Zhang, Hongwei Zhang and Baoan Zhang
[1]*Department of Environmental Economics, Tianjin Polytechnic University*
[2]*School of Environmental Science and Technology, Tianjin University*
China

1. Introduction

Over the past decades, controversial and conflict-laden water allocation issues among competing domestic, industrial and agricultural water use as well as urban environmental flows have raised increasing concerns (Huang & Chang, 2003); Particularly, Such competition has been exacerbated by the growing population, rapidly economic growth, deteriorating quality of water resources, and shrinking water availability due to a number of natural and human-induced impacts. A sounding strategy for water resources allocation and management can help to reduce or avoid the losses which are caused by water resources scarcity. However, in the water management system, many components and their interactions are uncertain. Such uncertainties could be multiplied not only by fasting changes of socioeconomic boundary conditions but also by unpredictable extreme weather events which caused by climate change. Thus, water resources management should be able to deal with all challenges above. Therefore, an effective integrated approach is desired for urban water adaptive management.

Many methods, such as stochastic, fuzzy, and interval-parameter programming techniques, have been employed to counteract uncertainties in different fields of water management and have made great progresses in managing uncertainties in model scale. Water resource is an integral part of the socio-economic-environmental (SEE) system, which is a complex system dominated by human. In order to reach a sounding decision, it is necessary for decision-makers to obtain a better understanding of the significant factors that shape the urban and the way the water resources system reacting to certain policy. Therefore, study of sustainable water resource management should be based on general system theory that addresses dynamic interactions amongst the related social-economic, environmental, and institutional factors as well as non-linearity and multi-loop feedbacks.

System dynamics (SD) aims at solving of complex systems problems by simulating development trends of the system and identifying the interrelations of each factor in the system. This will help to explore the hidden mechanism and thus improve the performance of the whole system. Hence, after proposed by W. Forrester (Forrester, 1968), SD model has been widely used in global, national, and regional scales for sustainability assessment and system development programme (Meadows 1973; Mashayekhi, 1990; Saeed, 1994). Due to the complexity of problems in the water system, the use of dynamic simulation models in water management has a long tradition (Biswas 1976; Roberts et al., 1983; Abbott and Stanley, 1999; Ahmad & Simonovic, 2004). The development journaey of several sections of applying system dynamics as a tool for integrated water management system analysis can be traced as from focusing on water system itself, to having a strong economic examinations on feedback relationships between industry and water availability, and then to having interaction with population growth (Liu et al., 2007). The above development make SD model has the flexibility and capability to support deliberative-analytical processes effectively. Meanwhile, SD and Multi Objective Programme (MOP) integrated model as an extension of the previous SD applications has been presented and used in urban water management in recent years, which takes into account both optimization and simulation (Guo et al, 1999; Zhang & Guo, 2002). This chapter will introduce a nonlinear dynamics approach for urban water resources demand forecasting and planning based on SD-MOP integrated model.

2. Uncertainties in Urban water system

2.1 Urban water system analysis

Generally, urban water system could be divided into four subsystems, i.e., social subsystem, economic subsystem, environmental subsystem and water resources subsystem. The relationships and interactions are complicate, as Fig. 1.

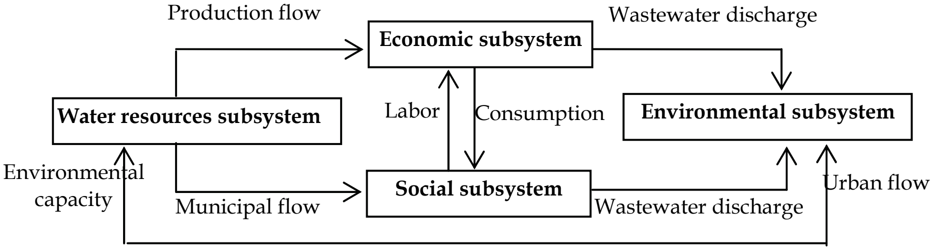


Fig. 1. Urban water management subsystems and relations

2.2 Uncertainties of urban water management system analysis

Urban water resources demand forecasting and planning are two important parts of urban water integrated management. Commonly, integrated water management should provide a framework for integrated decision-making and could be consists of system analysis, action results forecast, planning formulate and implementation, and evaluation and monitoring the goals and effects of implementation. At the system analysis stage, information collection and investigation are the basic work. A system structure is built based on a careful consideration of interactions among factors and subsystems. Long-term and short-term goals, problems, and priority focused will then are identified with both experts and stakeholders take part in. At the forecast stage, simulation model and evaluation model will be set up. Fixing on parameters and variable values of models and listing alternative solutions are the key process of the stage, based on field investigation, literature review and interviews with local stakeholders. Then according to the simulation and evaluation results of the alternatives, the selected solution can be identified and the corresponding desired actions can be determined.

Implementation and re-evaluation can't be separated completely. Management and re-evaluation is the mechanism that improves management goals and practices constantly.

Uncertainties limit the forecasting ability of and thus influence the quality of decision making. They can be categorized into four types : (1) intransience uncertainties caused by fasting changes of urban socioeconomic conditions; (2) external uncertainties caused by the stress of factors beyond the urban boundary (Liu, 2007); (3) uncertainties associated with raw data and model parameters driven from outdated or absent issues news, events, or statistic data; and (4) uncertainties arising from multiple frames (e.g. people's cognizing/perceiving technique/ability advance, world and ethical view change) (Jamieson, 1996; Pahl-Wostl, 2009). The above uncertainties are associated with all four stages, the details as Fig. 2.

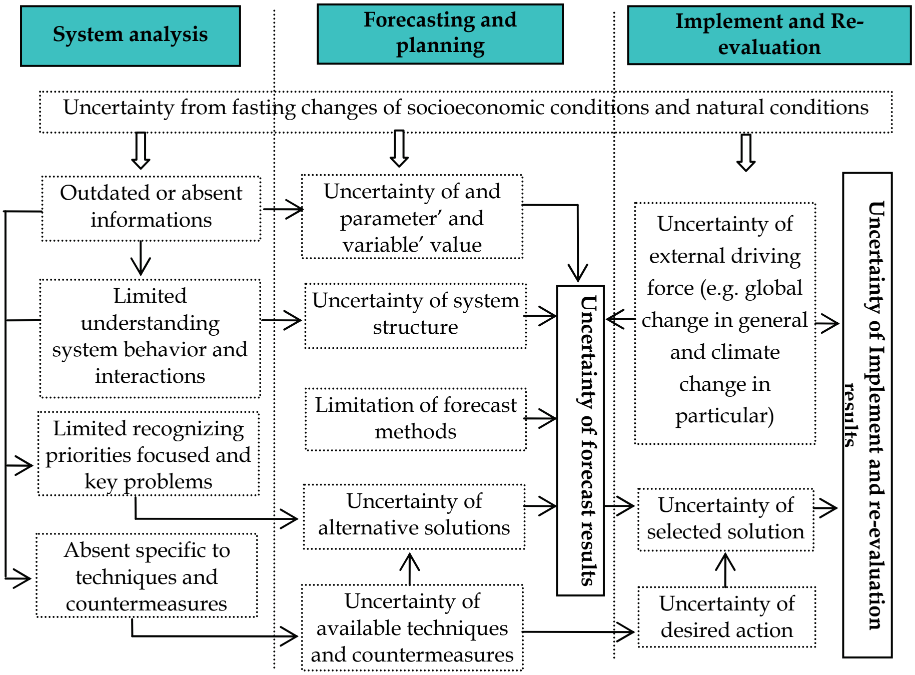

System analysis
Forecasting and planning
Implement and Re-evaluation
Uncertainty from fasting changes of socioeconomic conditions and natural conditions
Outdated or absent informations
Limited understanding system behavior and interactions
Limited recognizing priorities focused and key problems
Absent specific to techniques and countermeasures
Uncertainty of and parameter' and variable' value
Uncertainty of system structure
Limitation of forecast methods
Uncertainty of alternative solutions
Uncertainty of available techniques and countermeasures
Uncertainty of forecast results
Uncertainty of external driving force (e.g. global change in general and climate change in particular)
Uncertainty of selected solution
Uncertainty of desired action
Uncertainty of Implement and re-evaluation results

Fig. 2. The uncertainties in urban water management system

We can find that all above uncertainties are raised from the cognitive dimension (e.g. limited understanding system behavior and interactions among composing factors, uncertainty from fasting changes of socioeconomic conditions and change of natural conditions) and technical dimension (e. g. outdated or absent issues news/events/data, absent specific to techniques and countermeasures, limited of forecasting method) two aspects.

2.3 Overlook of counteracting measures to water system uncertainties

Whether we recognize it or not, socioeconomic laws and the natural laws are located in the objective world. So we can say that uncertainty is raised from the limitations of human cognition. Due to human cognitive abilities change, their understanding of the current world and their forecast of the future world will change over time. Furthermore, SEE system is a complexity system reflecting the mutual and complicated functions amongst the internal elements, which can be characterized by the complicated system structure properties far from balance status and with dissipation structures, as well as the behaviors of which the input-output response shows uncertainty that beyond people's experiential and qualitative cognition. We can be in virtue of SD model as well as interactions between modelers and stakeholders to interact the behavior uncertain from input-output response. The SD model can be run by different scenarios, and thus the optimal scenario can be selected by the analyses and discussions.

However, simulation model could be run in almost limitless scenarios according SEE complex system parameters changed in different policies. Thus it is difficult to simulate all possible scenarios constrained in time and fund. So it is difficult to ensure the optimal level of selected scenarios and its corresponding programme design. Therefore, SD-MOP integrated model (Zhang & Guo, 2002) is proposed to counteracts uncertainties with SD model applying in different scenarios simulation and analysis, and MOP model applying in optimization.

3. System dynamics model

3.1 The basic concepts of SD

The SD model takes certain steps along the time axis in the simulation process. At the end of each step, the system variables denoting the state of the system are updated to represent the consequences resulting from the previous simulation step. Initial conditions are needed for the first time step. Variables representing flows of information and initials, arising as results of system activities and producing the related consequences are named as level variables described as ▭ in the flow diagram, and rate variables described as ⧖. Auxiliary variable means the detailed steps by which information associated with current levels are transformed into rates to bring about future changes. In addition, the symbol ☁ represents the sinks or sources.

Fig. 3 is a sample flow diagram for the total population, in which the total population (TP) is a level variable; birth population (BP), death population (DP), and net migrated population (NP) are rate variables; and birth rate (BR), death rate (DR), and net migration rate (NR) are auxiliary variables.

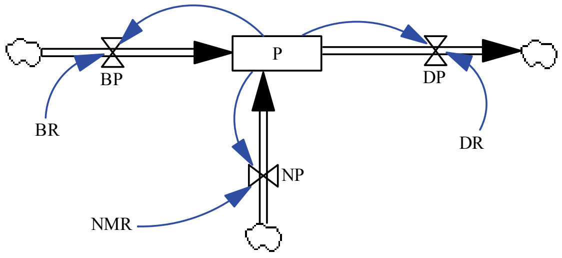


Fig. 3. SD flow chart of population subsystem

In SD level equation, three time points are denoted as J (past) , K (present), and L (future). The step from J to K is referred to as JK and that from K to L as KL. The duration period between successive points is named DT. Therefore, a level variable could be referred to as LEVEL.J, LEVEL.K, or LEVEL.L at a time point，RATE.JK and RATE.KL will function in the duration period. We can express:

LEVEL.K=LEVEL.J+DT*RATE.JK

3.2 The procedures for applying SD model to simulate target system behavior

The proedures for applying SD model to simulate target system behavior can be summarized into three steps.

(1) Construction SD model

The first step of the procedures is constructing SD model through analyses of the total system, and identifying the model validity by historical examination, and sensitivity analysis. Accordingly, parameters and relevance can be modified and confirmed.

(2) Validity examination

Validity examination examination includes direct observation, historical examination, and sensitivity analyses. Direct observation is through SD model run, if there is no obviouse unreasonable simulation results, we can to the historical examination.

Historical examination is checking the error between simulation and reality. The errors of main forecasting level variables are accepted is one of the requirements of SD model being used in reality system.

Another requirement is that the target system responds in lower degree sensitivity to most of the parameters through a series of sensitivity analyses conducted to examine the system's responses to variations of input parameters and/or their combinations. A concept of sensitivity degree is defined as follows:

$$S_Q = \left| \frac{\Delta Q_{(t)}}{Q_{(t)}} \cdot \frac{X_{(t)}}{\Delta X_{(t)}} \right| \tag{1}$$

where t is time; $Q_{(t)}$ denotes system state at time t; $X_{(t)}$ represents system parameter affecting the system state at time t; S_Q is sensitivity degree of state Q to parameter X; and $\Delta Q_{(t)}$ and $\Delta X_{(t)}$ denote increments of state Q and parameter X at time t, respectively.

For the n state variables (Q_1, Q_2,…, Q_n), the general sensitivity degree of a parameter at time t can be defined as follows:

$$S = \frac{1}{n}.\sum_{i=1}^{n} S_{Q_i} \tag{2}$$

Where n denotes a number of state variables; S_Q is sensitivity degree of state Q_i; and S is general sensitivity degree of the n states to the parameter X.

If there are some departures from the model validity requirement standards, the SD model should be adjusted until fix to the standards. Then, SD model could be used in target system behavior simulation.

3.3 SD model validity in simulating nonlinear feedback mechanism

Although SD equations are linearity, they simulating in computer can describe nonlinear characteristics produced by multi-feedback when consider temporal dynamic affection.

Figure 4 is a piece of water resources subsystem delay feedback circle- water supply capacity building flow chart, which included two simple first-order delay feedbacks.

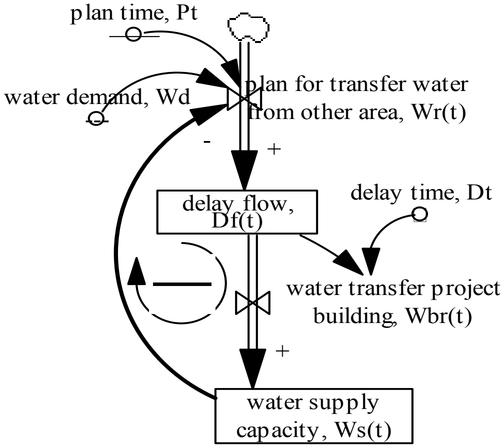

plan time, Pt
water demand, Wd
plan for transfer water from other area, Wr(t)
-
+
delay flow, Df(t)
delay time, Dt
water transfer project building, Wbr(t)
+
water supply capacity, Ws(t)

Fig. 4. Water supply capacity flow chart

Plan for transfer water from other area (*Wr (t)*) expression , in which had a first order delay, was shown as the basic divided differences formula: $Wr(t) = (Wd - Ws(t)) / Pt$.

Due to delay time to implement from confirming water transfer scheme to water supply formation, water transfer project building (*Wbr(t)*) could be expressed as a simple first order mater delay function: $Wbr(t) = Df(t) / Dt$.

As known, initialization of *Df (t)* is A m³, initialization of *Ws (t)* is B m³, *Wd* = C m³, *Pt* = a, *Dt* = b. According the above conditions can be established equations (3):

$$\begin{cases} Wr(t) = (Wd - Ws(t)) / Pt \\ Wbr(t) = Df(t) / Dt \\ Wd = \mathrm{C} \\ Pt = \mathrm{a} \\ Dt = \mathrm{b} \\ Df(t)\,|_{t=0} = \mathrm{A} \\ Ws(t)\,|_{t=0} = \mathrm{B} \end{cases} \tag{3}$$

Confluence rate was the derivative of the flow to time t. Hereby, it could be obtained the corresponding differential equations (4).

$$\begin{cases} Df^{'}(t) = \dfrac{1}{\mathrm{a}}(\mathrm{C} - Ws(t)) - \dfrac{Df(t)}{\mathrm{b}} & \text{(4-1)} \\ Df(0) = \mathrm{A} & \text{(4-2)} \\ W^{'}s(t) = \dfrac{Df(t)}{\mathrm{b}} & \text{(4-3)} \\ Ws(0) = \mathrm{B} & \text{(4-4)} \end{cases} \tag{4}$$

By equations (4), it could be derived the expression of flow, and the following equation could be obtained by on both sides of equation (4-3) of equations (4) derivation.

$$Ws''(t)+\frac{1}{2}Ws'(t)+Ws(t)=\mathrm{C} \tag{5}$$

Solve equation (5), the curve of water supply capacity, the curve of the delay flow , the curve of the plan rate, and the curve of project building could be derivate. Thus the results is according follow three conditions.

1. Condition 1

when b>$\frac{\mathrm{a}}{4}$, $\frac{1}{\mathrm{b}^2}-\frac{4}{\mathrm{ab}}<0$, then $\lambda_{1,2}=-\frac{1}{2\mathrm{b}}\pm\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1i}$

The solution of the equation (5) corresponding homogeneous equation is shown as:

$$W_{s(t)}=e^{-\frac{1}{2\mathrm{b}}t}(C_1\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-t}+C_2\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-t}) \tag{6}$$

Seeking the special solution of equation (5):

$$W^*_{s(t)}=C \tag{7}$$

According to equation (6) and (7), we can obtain the general solution of equation (5), which is shown as equation (8).

$$W_{s(t)}=e^{-\frac{1}{2\mathrm{b}}t}(C_1\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-t}+C_2\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-t}+C \tag{8}$$

$W_{s(0)}$ =B will be into the equation (8). Then,

$$B=C_1+\mathrm{C}\,,\ C_1=\mathrm{B}-\mathrm{C}$$

From,

$$\begin{aligned}W'_{s(t)}&=-\frac{1}{2\mathrm{b}}e^{-\frac{1}{2\mathrm{b}}t}((\mathrm{B}-\mathrm{C})\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1t}+C_2\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1t})\\&+e^{-\frac{1}{2\mathrm{b}}t}(\frac{\mathrm{C}-\mathrm{B}}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1t}+\frac{C_2}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1t})\end{aligned} \tag{9}$$

$W'_{s(0)}=D_{f(0)}=\frac{\mathrm{A}}{\mathrm{b}}$ is into the equation (9). Then,

$$\frac{\mathrm{A}}{\mathrm{b}}=-\frac{1}{2\mathrm{b}}(\mathrm{B}-\mathrm{C})+\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}C_2\,,C_2=\frac{2\mathrm{A}+\mathrm{B}-\mathrm{C}}{\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}}$$

According to the above, the special solution of equation (5) is shown as the follow:

$$W_{s(t)}=e^{-\frac{1}{2\mathrm{b}}t}((\mathrm{B}-\mathrm{C})\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1t}+\frac{2\mathrm{A}+\mathrm{B}-\mathrm{C}}{\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}}\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1t})+\mathrm{C} \tag{10}$$

The equation (10) is the curve of the Water supply capacity.
From: $D_{f(t)} = bW'_{s(t)}$, then the curve of the delay flow can be obtained as equation (11):

$$D_{f(t)} = e^{-\frac{1}{2\mathrm{b}}t}(\mathrm{A}\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}t - \frac{\mathrm{A}+\frac{2\mathrm{b}}{\mathrm{a}}(\mathrm{B}-\mathrm{C})}{\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}}\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}t) \tag{11}$$

The curve of plan for transfer water from other area can also be obtained as equation (12):

$$W_{r(t)} = e^{-\frac{1}{2\mathrm{b}}t}((\mathrm{B}-\mathrm{C})\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}t + \frac{2\mathrm{A}+\mathrm{B}-\mathrm{C}}{\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}}\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}t) + \mathrm{C} \tag{12}$$

The curve of project building can also be obtained as equation (13):

$$W_{br(t)} = \frac{D_{f(t)}}{\mathrm{b}} = e^{-\frac{1}{2\mathrm{b}}t}(\frac{\mathrm{A}}{\mathrm{b}}\cos\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}t - \frac{\mathrm{A}+\frac{2\mathrm{b}}{\mathrm{a}}(\mathrm{B}-\mathrm{C})}{\mathrm{b}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}}\sin\frac{1}{2\mathrm{b}}\sqrt{\frac{4\mathrm{b}}{\mathrm{a}}-1}t) \tag{13}$$

2. Condition 2

When $\mathrm{b} = \frac{\mathrm{a}}{4}$, $\frac{1}{\mathrm{b}^2} - \frac{4}{\mathrm{ab}} = 0$, Then: $\lambda_1 = \lambda_2 = -\frac{1}{2\mathrm{b}}$

The general solution of equation (5) is shown as the follow:

$$W_{s(t)} = e^{-\frac{1}{2\mathrm{b}}t}(C_1 + C_2 t) + \mathrm{C} \tag{14}$$

$W_{s(0)}$ =B will be into equation (14). Then,

$$\mathrm{B} = C_1 + \mathrm{C}\,,\; C_1 = \mathrm{B} - \mathrm{C}$$

From,

$$W'_{s(t)} = -\frac{1}{2b}e^{-\frac{1}{2b}t}[(\mathrm{B\text{-}C}) + C_2 t] + C_2 e^{-\frac{1}{2b}t} \tag{15}$$

$W'_{s(0)} = \frac{D_{f(0)}}{b} = \frac{\mathrm{A}}{b}$ will be into the equation (15). Then,

$$\frac{\mathrm{A}}{b} = -\frac{1}{2b}(\mathrm{B}-\mathrm{C}) + C_2\,,\; C_2 = \frac{2\mathrm{A}+\mathrm{B}\text{-}\mathrm{C}}{2b}$$

According to the above, the special solution of equation (5) is shown as the follow:

$$W_{s(t)} = e^{-\frac{1}{2\mathrm{b}}t}((\mathrm{B}-\mathrm{C}) + \frac{2\mathrm{A}+\mathrm{B}-\mathrm{C}}{2\mathrm{b}}t) + \mathrm{C} \tag{16}$$

The equation (16) is the curve of the water supply capacity.

From $D_{f(t)} = \mathrm{b}W'_{s(t)}$, then

$$D_{f(t)} = -\frac{1}{2}e^{-\frac{1}{2b}t}((B-C)+\frac{2A+B-C}{2b}t)+\frac{2A+B-C}{2b}e^{-\frac{1}{2b}t} \tag{17}$$

The equation (17) is the curve of the delay flow.

The curve of the water transfer rate can be obtained as (18) and the rate curve of Building water supply facilities can be obtained as (19).

$$W_{r(t)} = (C - W_{s(t)})/a = e^{-\frac{1}{2b}t}(\frac{C\text{-}B}{a} - \frac{2A\text{+}B\text{-}C}{2ab}t) \tag{18}$$

$$W_{br(t)} = \frac{D_{f(t)}}{b} = -\frac{1}{2b}e^{-\frac{1}{2b}t}((B-C)+\frac{2A+B-C}{2b}t)+\frac{2A+B-C}{2b^2}e^{-\frac{1}{2b}t} \tag{19}$$

3. Condition 3

when $b < \frac{a}{4}$, $\frac{1}{b^2} - \frac{4}{ab} > 0$, Then, $\lambda_{1,2} = -\frac{1}{2b} \pm \frac{1}{2b}\sqrt{1-\frac{4b}{a}}$

$$W_{s(t)} = C_1 e^{(-\frac{1}{2b}+\frac{1}{2b}\sqrt{1-\frac{4b}{a}})t} + C_2 e^{(-\frac{1}{2b}-\frac{1}{2b}\sqrt{1-\frac{4b}{a}})t} + C \tag{20}$$

$W_{s(0)}$=B will be into the equation (20). Then

$$C_1 + C_2 + C = B\,,\; C_1 = B - C - C_2$$

From

$$W'_{s(t)} = (-\frac{1}{2b}+\frac{1}{2b}\sqrt{1-\frac{4b}{a}})C_1 e^{(-\frac{1}{2b}+\frac{1}{2b}\sqrt{1-\frac{4b}{a}})t} + (-\frac{1}{2b}+\frac{1}{2b}\sqrt{1-\frac{4b}{a}})C_2 e^{(-\frac{1}{2b}+\frac{1}{2b}\sqrt{1-\frac{4b}{a}})t} \tag{21}$$

$W'_{s(0)} = \frac{D_{f(t)}}{b} = \frac{A}{b}$ will be into the equation (21). Then

$$\frac{A}{b} = (-\frac{1}{2b}+\frac{1}{2b}\sqrt{1-\frac{4b}{a}})C_1 + (-\frac{1}{2b}-\frac{1}{2b}\sqrt{1-\frac{4b}{a}})C_2 \tag{23}$$

Because

$$C_1 = B - C - C_2$$

So, $C_1 = \dfrac{(\sqrt{1-\frac{4b}{a}}+1)(B-C)+2A}{2\sqrt{1-\frac{4b}{a}}}$, $C_2 = \dfrac{(\sqrt{1-\frac{4b}{a}}-1)(B-C)-2A}{2\sqrt{1-\frac{4b}{a}}}$

According to the above, the special solution of equation (5) is shown as the follow:

$$\begin{aligned} W_{s(t)} &= \frac{(\sqrt{1-\frac{4\text{b}}{\text{a}}}+1)(\text{B-C})+2\text{A}}{2\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}+\frac{1}{2\text{b}}\sqrt{1-\frac{4\text{b}}{\text{a}}})t} \\ &+\frac{(\sqrt{1-\frac{4\text{b}}{\text{a}}}-1)(\text{B-C})+2\text{A}}{2\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}+\frac{1}{2\text{b}}\sqrt{1-\frac{4\text{b}}{\text{a}}})t}+\text{C} \end{aligned} \tag{22}$$

The equation (22) is the curve of the Water supply capacity.
From $D_{f(t)} = \text{b}W'_{s(t)}$, then

$$D_{f(t)} = \frac{\frac{4\text{b}}{\text{a}}(\text{C-B})+2\text{A}(\text{-}1+\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}})}{4\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}+\frac{1}{2\text{b}}\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}})t} + \frac{\frac{4\text{b}}{\text{a}}(\text{C-B})+2\text{A}(1+\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}})}{4\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}-\frac{1}{2\text{b}}\sqrt{1\text{-}\frac{4\text{b}}{\text{a}}})t} \tag{23}$$

The equation (23) is the curve of the delay flow.
And the curve of the water transfer rate can be obtained as (24).

$$\begin{aligned} W_{r(t)} &= \frac{(\text{C}-W_{s(t)})}{\text{a}} \\ &= \frac{(\sqrt{1-\frac{4\text{b}}{\text{a}}}+1)(\text{C-B})-2\text{A}}{2a\sqrt{1-\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}+\frac{1}{2\text{b}}\sqrt{1-\frac{4\text{b}}{\text{a}}})t} + \frac{(\sqrt{1-\frac{4\text{b}}{\text{a}}}-1)(\text{C-B})+2\text{A}}{2a\sqrt{1-\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}-\frac{1}{2\text{b}}\sqrt{1-\frac{4\text{b}}{\text{a}}})t} \end{aligned} \tag{24}$$

The rate curve of Building water supply facilities can be obtained as (25).

$$\begin{aligned} W_{br(t)} &= \frac{D_{f(t)}}{\text{b}} = \frac{\frac{4\text{b}}{\text{a}}(\text{C}-\text{B})+2\text{A}(-1+\sqrt{1-\frac{4\text{b}}{\text{a}}})}{4\text{b}\sqrt{1-\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}+\frac{1}{2\text{b}}\sqrt{1-\frac{4\text{b}}{\text{a}}})t} \\ &+\frac{\frac{4\text{b}}{\text{a}}(\text{B}-\text{C})+2\text{A}(1+\sqrt{1-\frac{4\text{b}}{\text{a}}})}{4\text{b}\sqrt{1-\frac{4\text{b}}{\text{a}}}} e^{(-\frac{1}{2\text{b}}-\frac{1}{2\text{b}}\sqrt{1-\frac{4\text{b}}{\text{a}}})t} \end{aligned} \tag{25}$$

From above deduction, we can know that although SD equations are linearity, they simulating in computer can describe nonlinear characteristics produced by multi-feedback when consider temporal dynamic affection.

4. Decision-making system based on SD-MOP integrated model for urban water resources demand forecasting and planning

From above analysis, we can know that urban water resources demand forecasting is the key procedure of urban water system management. In different scenarios, the forecasting outcomes may be different, which result in different corresponding planning. From above deduction, we can also get the conclusion that SD model can be applying to simulate nonlinear and complex system behavior though the basic equations are linear and simple. Hence, we introduce a decision-making system which core in SD-MOP integrated model for urban water resources demand forecasting and planning. The procedure of applying SD-MOP integrated model as Fig.5.

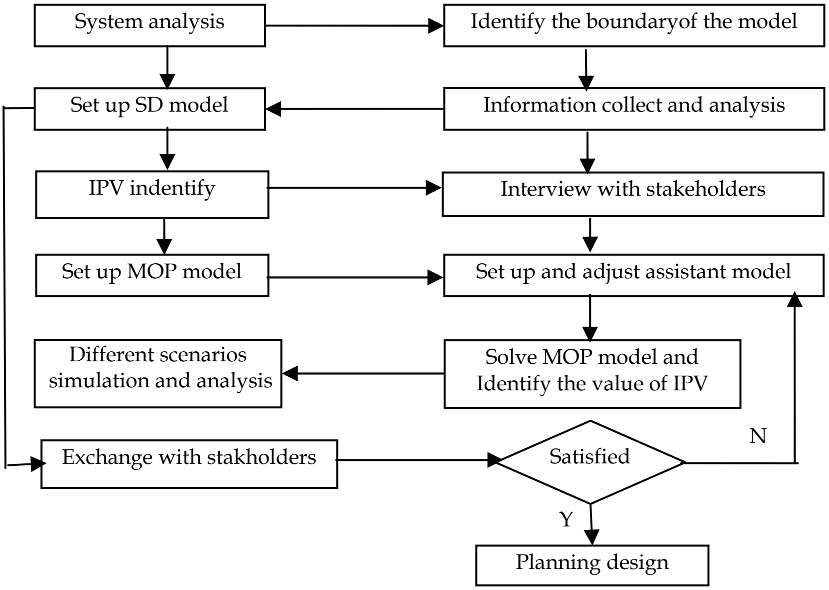

System analysis
Identify the boundaryof the model
Set up SD model
Information collect and analysis
IPV indentify
Interview with stakeholders
Set up MOP model
Set up and adjust assistant model
Different scenarios simulation and analysis
Solve MOP model and Identify the value of IPV
Exchange with stakholders
Satisfied
N
Y
Planning design

Fig. 5. The procedure of SD-MOP integrated model applying

In SD-MOP integrated model, SD is used for water resources system dynamics nonlinear behavior simulation, and MOP is used for optimal policy choice and optimal design forming.

4.1 Setting up SD model

The first step of SD-MOP applying is constructing SD model based on information collection system analysis. The procedures of constructing SD model are the follows:

1. identify the boundary of SD model;
2. classify sub-systems of urban water system;
3. determine the set of main level variables;
4. analysis the realtions of system parameters and variable;
5. design the flow diagram;
6. determine the basic value of parameters by mathmatic forecasting method both in statistical method and experience according to current and historical imformation of the target system;
7. set up basic mathmatic equations which consist of SD model;
8. test SD model validity and adjust it accoding testing results until it can be used in realistic system simulation.

4.2 Analyzing IPV

Analyzing the sensitivity by sensitivity test and original run (run in the condition which the system keep current behavior and tendency without any policy adjustment), the sensible parameters and the closed relating variables can be identified, which are named as IPV (Important Parameter and Variable).

IPV aggregation includes controllable factors and non-controllable two types. Non-control lable factors can become system development neck, while adapting controllable factors in suitable way could exploit urban development.

4.3 Setting up MOP model

Running the SD model based on the current situations (called original run). The gap between the original run results and ideal level of the system can be identified. In order to obtain optimal programme design and adjust the system function and behavior, MOP model cored in IPV is set up. In the MOP model the controllable factors of IPV become decison variables and non-controllable factors of IPV become constrains, while some level variable which closely related to IPV become maximum or minimum aim.

General format of MOP model as follow:

$$\max\left| f_k(x) / \forall_k \right| \tag{26}$$

$$\text{s.t. } g_i(x) \le b_i, \forall_i \tag{27}$$

$$x_j \ge 0, x_j \in x \tag{28}$$

Where, x is decision variable (a set of real number in a closed boundary limit and is the value of IPV or value of variable that are related to IPV), equation (26) is objective function, (27) and (28) are the limiting conditions.

4.4 Setting up assistant model to solve MOP

Applying ODTL (Objective Deviation Tolerance Level) method (Zhou, 1998) to solve MOP model. Here, there is some different from Zhou in interview process. First, we determin each goal ODTL by interview with stakeholder based on giving them original run results and the ideal goals. Second, the decision is not finished in one time, but in several times based on showing them the former scenarios SD model simulation results which corresponding to their choice of each goals ODTL, and the stake holders can adjust there decision by comparing and discussing the former results. Finally, the optimal IPV can be determined by several adjust assistant model, solve MOP, simulation corresponding system tendency, and compare and selecte the desirable scenario.

4.5 Planning

Based on the optimum values of IPV, the proposals for running the model can be designed. Accordingly the final plan proposal can be formulated.

5. Case study

Applying SD-MOP integrated model in a real urban system to test its validity [Zhang 2010].

The boundary of the target system is the urban area of Qinhuangdao, which is a city of Hebei province, located at latitude 39°22′-40°37′N and longitude 118°33′-119°51′E, and covers an area of 7,812 km^2. Qinhuangdao has jurisdiction over three districts (Shanhaiguan, Beidahe, and Haibin) and four countries (Lulong, Qinglong, Funing, and Changli). The annual rainfall in Qinhuangdao is about 670mm, with the water resource per capita in Qinhuangdao is 600m^3/a, which is 1/4 the average level in China. The system is composed of population subsystem, industry subsystem, services subsystem, water supply subsystem and water environmental protection subsystem. The planned period is 15 years (2006 - 2020). It is divided into two phases, i.e., 2006-2010 and 2010 - 2020. The base year is 2000.

5.1. Constructing SD model

Based on the analysis of the target system, SD model of Qinhuangdao (QHDWSD) can be constructed, and thus the sensibility of the model can also be tested. There are more than 110 variables in SD model, in which there are more than 110 system dynamic equations. Fig. 6 is the flow chart of QHDWSD.

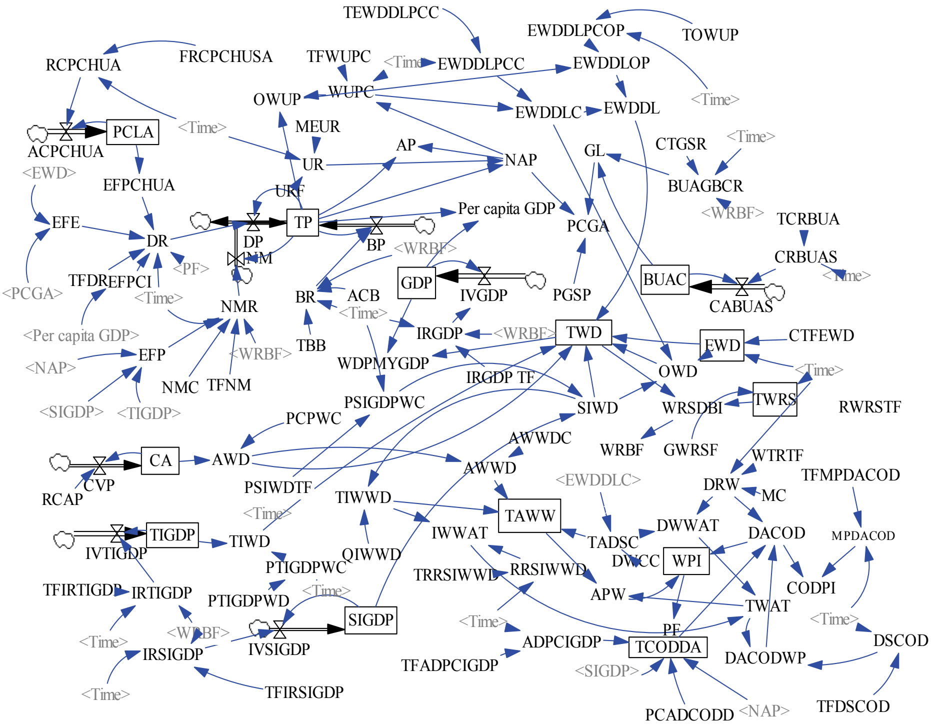


Fig. 6. QHDSD diagram

5.2 Identifying IPV

Based on original running and putting eight variables and fourteen parameters into sensitivity analysis, IPV were identified. Those are: Increase rate of second industrial GDP, per second industrial GDP water consumption, Per capita plow land water consumption.

5.3 Setting up and solving MOP model

In the original simulation, when GDP getting in the aim scale' water resource supply and demand balance index (water available supply to human social and economic activities divided by water demand human social and economic activities) will be lower than 0.6 in 2020 (Fig. 2). The consequence will be that people active's water consumption invade and occupy eco-environmental share and lead to water ecosystem quality degradation and water resource sustainable supply capability decrease. According above analysis, the key issue is the structure of the economic, thus MOP model is setting up as follow.

$$\mathrm{Z}_1(X) = \max\sum_{i=1}^{3} X_i \tag{29}$$

$$\mathrm{Z}_2(X) = \min\sum_{i=1}^{3} \mathrm{q}_i \cdot X_i \tag{30}$$

$$\sum_{i=1}^{3} \mathrm{q}_i \cdot X_i \le \mathrm{Q} \tag{31}$$

$$\mathrm{Ymin}_i \le X_i / (\sum_{i=1}^{3} X_i) \le \mathrm{Ymax}_i \tag{32}$$

$$X_i \ge 0 \tag{33}$$

where: X_i=GDP of three industry (10^8￥); q_i=per GDP water consumption of three industry (t/10^8￥); Q=water resource amount could be supplied to human economic activities (t); Ymin$_i$= the lower bound of three industry proportion in total GDP; Ymax$_i$= the higher bound of three industry proportion in total GDP. Then set up assitant model and solved it based on interaction with stakeholders who consists of the staff of water resources bureau, the staff of the environmental protection agency, the staff of regional development and reform Commission the staff of related bureaus, the staff of water supply and wastewater treatment firms, delegates of the three industries, and representatives from the public.

5.4 Obtaining relative optimal programme

According IPV solution, the optimal design could be obtained and the corresponding water resources plan of Qinhuangdao city was formulated. Table 1 shows the comparison of different industry ratio in the total gross domestic production (GDP) respectively between optimal solutions and original tendency. The comparison results of the water supply-demand balance, GDP, population scale and water pollution index between the feasible programme simulations with the original simulation as Fig. 7.

year	item	Primary industry (%)	Secondary industry (%)	Tertiary industry (%)
2010	optimal designs	62	356	457.5
	original tendency	65	370	440.5
2020	optimal designs	102	1220	1397.2
	original tendency	107	1256	1357.2

Table 1. Industrial structure (different industry ratio in GDP)

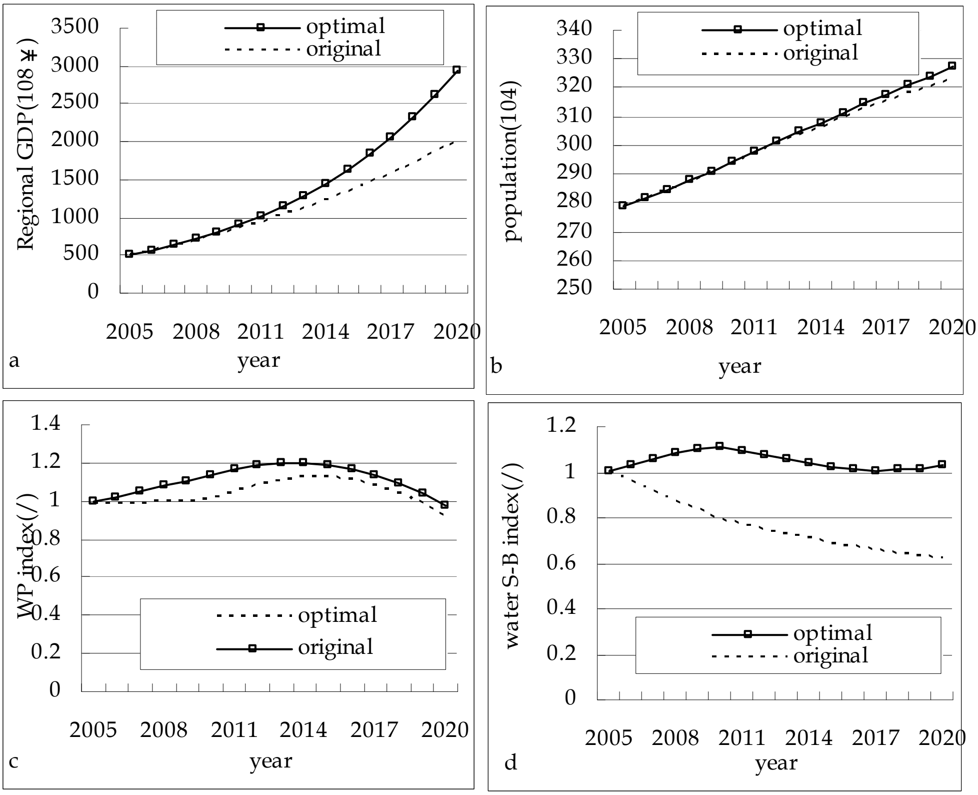


Fig. 7. Main level variable comparing between optimal design and original tendency

In Fig. 7, sub Fig-a is for gross domestic production, sub Fig-b is for total population scale, sub Fig-c is for water pollution index (WPI-the ratio of simulating year water contamination discharge to base year water contamination discharge) contamination, and sub Fig-d is for water resources supply-demand balance index (WRSDBI-the ratio of water supply quantity to water demand quantity).

Fig. 7 and table 1 indicate that through adjusting system structure can realize water sustainable utilization while not decreasing the speed of economic development. The water resource strategy plan is based on nonlinear dynamics forecasting approach for water resource demand.

5.5 Nonlinear dynamics approach validity test in practice

Follow is an example of Qinhuangdao water resource plan of 2000 to 2005. And it was researched by our group during 1998 to 2000. In the plan, we used two methods, nonlinear method and trend extending method, to forecast urban water resources demand. Fig. 8 shows the comparative errors for forecasting data and actual data between SD nonlinear method and trend extending method. From Fig. 8, we can know that nonlinear forecasting is more accurate with can give support to water resources plan.

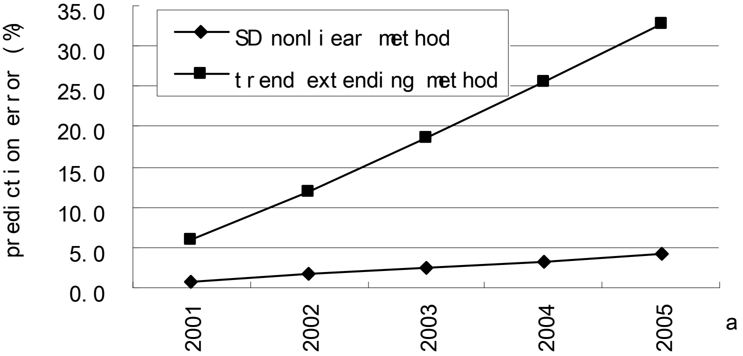


Fig. 8. The comparative analysis results

6. Conclusion

From above study, we can get the conclusion : (i) complex system analysis and nonlinear dynamics simulation are very useful for urban water resource demand forecasting and planning, (ii) the integrated model of SD-MOP can avoid the randomness of proposal designed by experiences of planners and decision-makers, which results in the generated planning proposal has high reliability.

7. Acknowledgements

This work is supported by Hebei Education Department Natural Science Foundation (No.2008324), and Tianjin philosophy and social sciences key project Foundation (No. TJYY08-1-078).

8. References

Abbott MD, Stanley RS. (1999). Modeling groundwater recharge and flow in an upland fractured bedrock aquifer. *Syst Dyn Rev*, 15, pp. 163-184

Ahmad S, Simonovic SP. (2004). Spatial system dynamics: new approach for simulation of water resources systems. *J Comput Civ Eng*, 18, pp. 331-340

Biswas AK. (1976). Systems approach to water management. McGraw Hill, London. ISBN: 0070054800

Claudia Pahl-Wostl. (2009). A conceptual framework for analyzing adaptive capacity and multi-level learning processes in resource governance regimes, *Global Environmental Change,* 19, pp. 354-365.

Forrester JW, 1968. Principles of systems. Productivity, Portland

Guo HH, Xu YL, Zou R. (1999). Study on the environmental system planning method for watershed under incomplete information. *J Environ Sci*, 19, pp. 421-6 (in Chinese)

Huang G H, Chang N B. (2003). The perspectives of environmental informatics and systems analysis. *J Environ Inform,* 1 (1), pp. 1-6

Jamieson D. (1996). Scientific uncertainty: How do we know when to communicate research findings to the public? Science of the Total Environment 184, pp. 103-107

Mashayekhi AN. (1990). Rangelands destruction under population growth: the case of Iran. *Syst Dyn Rev,* 6, pp. 167-93

Meadows DL, Meadows DH. (1973). Toward global equilibrium: collected papers. Cambridge, Mass: Wright-Allen Press

Roberts N, Andersen D, Deal R, Garet M, ShafferW. (1983). Introduction to computer simulation: a system dynamics modeling approach. Productivity, Portland

Saeed K. (1994). Development planning and policy design: a system dynamics approach, Brookfield: Avebury

Yong Liu, Huaicheng Guo, Zhenxing Zhang, Lijing Wang, Yongli Dai, Yingying Fan. (2007). An Optimization Method Based on Scenario Analysis for Watershed Management under Uncertainty. *Environ Manage*, 39, pp. 678-690

Zhou Rui, Guo Huaichen, Li lei. (1998). A new method based on the objective-deviationtolerance-level from multi-objective-decision making, *Journal of system engineering*, 13(3), pp. 41-47(in Chinese)

Zhang Xuehua, Guo Huaichen. (2002). Application of SD-MOP integrated model in urban eco-environmental planning of Qinhuangdao. *J Environ Sci*, 22, pp. 92-97(in Chinese)

Zhang Xuehua, Guo Huaichen, Zhang Baoan. (2002). Application of System Dynamics-Multi Objedtive Programme integrated model in urban water resources planning of Qinhuangdao.*Advance in water seience*, 13, pp. 351-357(in Chinese)

Zhang Xuehua, Zhang Hongwei, Zhang Baoan. (2010). SD-MOP integrate model and its application in water resources plan: A case study of Qinhuangdao. *Kybernetes*, 39 (in press)

15

A Detection-Estimation Method for Systems with Random Jumps with Application to Target Tracking and Fault Diagnosis

Yury Grishin and Dariusz Janczak
Bialystok Technical University, Electrical Engineering Faculty
Poland

1. Introduction

Methods for detection and estimation of the structure or parameters of abrupt changes in dynamic systems play an important role for solving a number of problems encountered in practice. They have an important significance in different fields of telecommunications and control applications, such as radar tracking of maneuvering targets, fault diagnosis and identification (FDI), speech analysis, signal processing in geophysics and biomedical systems. Most of these applications belong to the class of problems with nonlinear dynamics. Among them an important role is played by a wide class of systems with abrupt random jumps of parameters or structure.

A dynamic system with jumps of this kind can be defined as a system in which the structure or parameters can change at any random time. Therefore, in order to describe such a system, it is convenient to introduce an unknown random vector $\vartheta(k)$ that determines the current system structure and parameters. Then the system state and observation equations are dependent on this changing vector. The general case then is described as follows:

$$x(k+1) = F[x(k), \vartheta(k), w(k)]\,, \tag{1}$$

$$y(k) = h[x(k), \vartheta(k), v(k)]\,, \qquad \vartheta(k) \in \Omega \quad , \tag{2}$$

where F and h are known nonlinear functions, $w(k)$ and $v(k)$ are system and measurement noises respectively and Ω is the space of possible values of the vector $\vartheta(k)$.

The space Ω can consist of finite or infinite sets of elements. The structure of the space Ω and evolution of the vector $\vartheta(k)$ in time determine the main approaches to solving the problem of detection-estimation in a dynamic system with jump structure. The classification of the statistical characteristics of the parameter vector $\vartheta(k)$ is presented in Fig. 1. According to this classification, after the jump the parameter vector $\vartheta(k)$ can take on finite or infinite sets of values. In the case of the former the dynamic system can be in one of N possible structures. It has been shown that a model of this kind (Willsky, 1976) is the most comprehensive description of system jump changes. Such models demand a considerable amount of prior information on probable jump changes in the system. At the same time, they require a great deal of computation when used for state estimation or jump detection in real-time systems. Modifications to these models are often used for solving problems related to tracking maneuvering targets in radars (Gini & Rangaswamy, 2008) and in designing reliable dynamic systems (Patton et al., 1989). Usually in these cases the multiple model (MM) (Blackman & Popoli, 1999), multiple hypothesis test (MHT) (Bar-Shalom et al., 2001) or interactive multiple model (IMM) approaches are used (Mazor et al., 1998).

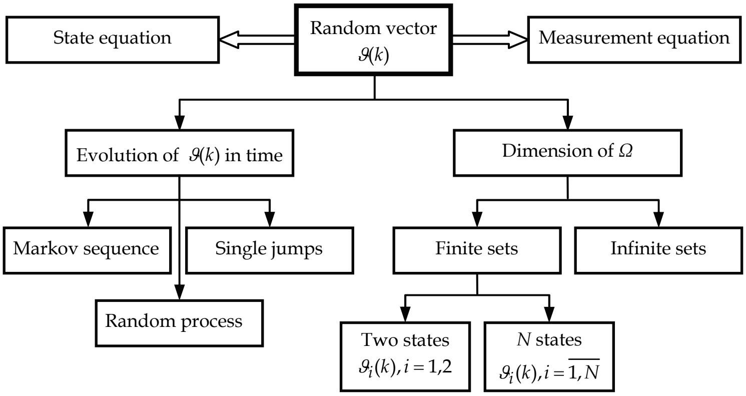

State equation
Random vector $\vartheta(k)$
Measurement equation
Evolution of $\vartheta(k)$ in time
Dimension of Ω
Markov sequence
Single jumps
Finite sets
Infinite sets
Random process
Two states $\vartheta_i(k), i = 1,2$
N states $\vartheta_i(k), i = \overline{1,N}$

Fig. 1. Classification of the parameter vector $\vartheta(k)$

Evolution of the vector $\vartheta(k)$ in time can be described in terms of a random process with a known multidimensional probability density function (pdf), by the Markov sequence or by single jumps. In practice it is difficult to obtain a priori information about the multidimensional probability density function of the process. Therefore a model based on these criteria is not readily applicable to solving the problem of detecting jumps in dynamic systems.

Models in which the vector $\vartheta(k)$ is defined by Markov properties can describe a broad variety of jump changes and hence they are widely used in radar applications and FDI theory (Grishin, 1994). Another class of system models with a jump structure is represented by systems with single jumps that can occur at random time, the pdf of these moments being unknown. This approach assumes that after the jump, the system parameters and structure remain unchangeable. The latter assumption is often unjustified in practice because after the jump the system may be non-stationary. More adequate models are required in order to describe situations in which following the jump the parameter vector $\vartheta(k)$ changes according to the Markov sequence. A model of this kind will be considered below.

For a solution to the problem in a real-time system with a minimum computational burden it is desirable to have simple but adequate models of the jumps. A method for modelling jumps in dynamic systems by means of additive Gauss-Markov sequences with random time rises in the state and observation equation is proposed in (Grishin, 1994). Nevertheless such models also require a relatively large amount of prior information on the structure and parameter of the jumps.

In order to resolve these difficulties a mixed multiple additive Gauss-Markov model is proposed. For this model far less a priori information on probable system jumps is required and it can be applied to a broad class of dynamic systems for which relatively simple models can be used.

Using such models and a generalized likelihood ratio approach (GLR) (Katayama & Sigimoto, 1997) it is easy to obtain suboptimal algorithms for state estimation and jump detection. Such algorithms in comparison with the multiple model estimation algorithms have relatively moderate computation requirements. They can be obtained in recursive form and realized in real-time systems.

In the following section of this chapter we outline the applications of models of this kind to the problem of radar maneuvering target tracking and failure detection.

2. The system model

The system and measurement equations are described by one of the following models:

$$\begin{aligned} &x(k+1) = \Phi(k+1,k)x(k) + w(k) + G_S(k+1)\vartheta_j(k+1,t_i)1(k+1,t_i), \\ &y(k) = H(k)x(k) + v(k), \qquad\qquad j = 1,\dots,N, \end{aligned} \tag{3}$$

or:

$$\begin{aligned} &x(k+1) = \Phi(k+1,k)x(k) + w(k), \\ &y(k) = H(k)x(k) + v(k) + H_0(k)\vartheta_j(k,t_i)1(k,t_i), \qquad j = 1,\dots,N, \end{aligned} \tag{4}$$

where $x(k)$ is the state vector, $w(k), v(k)$ are white Gaussian sequences with zero mean and covariance matrices $Q(k)$ and $R(k)$ respectively, $\vartheta_j\ (k,t_i\ )$ - an unknown Gauss-Markov state vector modelling changes in the system after the jump at the time t_i and $1(k,t_i\ )$ is the unit step function that is zero when $k < t_i$.

The vector $\vartheta_j\ (k,t_i\ )$ can be written in the general case as follows for a dynamic system driven by the random signal $\xi_j(k)$:

$$\vartheta_j(k+1,t_i) = \varphi_j(k+1,k)\ \vartheta_j(k,t_i) + \xi_j(k), \quad j = 1,...,N, \tag{5}$$

where $\varphi_j(k+1,k)$ - a transition matrix, $\xi_j(k)$ is a white Gaussian sequences with zero mean and covariance matrix $Q_{oj}(k)$, j - a number of possible jump models of which prior probabilities $P_j(t_i)$ can be given or not. The other notations specified are commonly used (Sorenson, 1985). The a priori distributions of a random value t_i are assumed to be unknown.

Thus the additional dynamic system can be described by a set of equations of the form (5) with different transition matrices. The choice of a corresponding model can be carried out in real time by an adaptive processing algorithm. The case of one of N possible models will be considered below.

Depending on the nature of the parameter vector $\vartheta_j(k,t_i)$ the model of changes may be classified (Grishin & Janczak, 2006) as deterministic ($\xi_j(k) = 0$), stochastic ($\varphi_j(k+1,k) = 0$) or mixed ($\varphi_j(k+1,k) \neq 0,\ \xi_j(k) \neq 0$).

It is easy to demonstrate that the equations (3) - (5) describe a wide variety of system jumps which take place in different parts of the system such as jump changes of the state vector and its dimension, jumps of the system transition matrix elements, the covariance matrices of observation and system noises. Let us consider a description of different jumps in the system with the additive Gauss-Markov models.

Jump changes of the state vector dimension

For $k > t_i$ equation (3) can be rewritten as

$$x(k+1) = \Phi(k+1,k)x(k) + G_S(k+1)\varphi(k+1,k)\vartheta(k,t_i) + G_S(k+1)\xi(k) + w(k) \tag{6}$$

Defining the augmented state vector as $x_a(k+1) = \left[x(k+1) \quad \vartheta(k+1,t_i)\right]^T$, from (5) and (6)

$$x_a(k+1) = \Phi_a(k+1,k)x_a(k) + \Gamma(k+1)w_a(k), \tag{7}$$

where

$$\Phi_a(k+1,k) = \begin{bmatrix} \Phi(k+1,k) & G_S\,(k+1)\varphi(k+1,k) \\ 0 & \varphi(k+1,k) \end{bmatrix}, \qquad \Gamma(k+1) = \begin{bmatrix} 1 & G_S\,(k+1) \\ 0 & 1 \end{bmatrix}$$

are transition and input matrices, $w_a(k) = \left[w(k)\ \xi(k)\right]^T$ - the augmented input noise vector. Thus equation (3) may be used for modelling the jumps in the system dimension. As the dimension of the observation vector is the same, the observation matrix for $k > t_i$ must be altered, such that $H_a(k) = \left[H(k) \quad 0\right]$.

Jump changes of the state vector variables

If in equation (3) the input matrix is:

$$G_S(k+1) = \begin{cases} I\,,\ k+1 = t_i\,, \\ 0\,,\ k+1 \neq t_i\,, \end{cases} \tag{8}$$

then the state equation of the system will be:

$$x(k+1) = \Phi(k+1,k)x(k) + w(k) + \vartheta(k+1,t_i)\delta(k+1,t_i)\,. \tag{9}$$

Thus every variable of the state vector at time $k+1 = t_i$ changes abruptly. The values of these changes are equal to the values of the corresponding variables of the random vector $\vartheta(k+1,t_i)$. If for $k > t_i$ $G_S(k+1) = I$ and the parameters of equation (5) are chosen as $\varphi(k+1,k) = I\,, \vartheta(t_i,t_i) = \vartheta_0\,, \xi(k) = 0\ (Q_0 = 0)$, then one has:

$$x(k+1) = \Phi(k+1,k)x(k) + w(k) + \vartheta_0 1(k+1,t_i)\,. \tag{10}$$

The preceding equation shows, that state variable bias appears at time t_i.

Abrupt changes of the observation matrix

In considering jumps of the observation matrix elements it is necessary to restrict our discussion to equation (4). If for $k > t_i$ the identity $\vartheta(k,t_i) = x(k)$ is valid, that is $\varphi(k+1,k) = \Phi(k+1,k)\,, \xi(k) = w(k)\,, \vartheta(t_i,t_i) = x(t_i)$, then the observation equation is:

$$y(k) = H(k)x(k) + v(k) + H_0(k)x(k) = [H(k) + H_0(k)]x(k) + v(k)\,, k > t_i\,. \tag{11}$$

3. Detection-estimation algorithms in the systems with the additive Gauss-Markov jumps

To design an appropriate detection-estimation algorithm for a system in which parameters can be abruptly changed, it is necessary to detect the changes, to isolate them (that is to determine the system element in which these changes take place) and then to estimate theirs value. The main approaches to the design of such algorithms include the following:

- change-sensitive filters (Limit Memory Filters) (Willsky, 1976),
- an innovation-based approach that uses the generalized likelihood ratio (GLR) (Gertler, 1998),
- the multiple hypothesis test (Katayma & Sugimoto, 1997),
- an artificial neural network approach (Patton et al., 1989).

In this section we focus on the GLR approach. An approach of this kind involves the use of the basic Kalman filter which is matched with the normal mode of the input process and the GLR computation of the innovation process to detect the parameter or structure jumps (Whang et al., 1994).

When the system changes have occurred, the innovation process is no longer zero mean and it carries information about changes in the system.

3.1 Synthesis of the detection-estimation algorithm

Let us consider the system for which state and measurement equations are given by the model (3). Then, calculating the propagation of all signals through the Kalman filter that is matched with a system without jumps, we obtain that the innovation process $z(k/k-1)$ of the filter in this case can be presented in the following form (Grishin, 1994):

$$z_S(k/k-1) = T_S(k,t_i)\varepsilon(k,t_i) + z_1(k/k-1). \tag{12}$$

where $z_1(k/k-1)$ is the innovation process of the matched Kalman filter

$$z_1(k/k-1) = y(k) - H(k)\hat{x}(k/k-1) \tag{13}$$

and

$$T_s(k,t_i) = [\Psi_{c1}(k,t_i)\ \ \psi_{c2}(k,t_i)\ \ H(k)\Phi(k,k-1)], \tag{14}$$

$$\psi_{C1}(k,t_i) = \begin{cases} H(t_i)G_S(t_i), & k = t_i, \\ H(k)[\Phi_C(k,t_i) - \Phi(k,k-1)F_{C1}(k-1,t_i)\varphi^{-1}(k,k-1)], & k > t_i; \end{cases} \tag{15}$$

$$\Phi_c(k,t_i) = \begin{cases} G_S(t_i), & k = t_i, \\ G_S(t_i) + \Phi(k,k-1)\Phi_c(k-1,t_i)\varphi^{-1}(k,k-1), & k > t_i \end{cases} \tag{16}$$

$$F_{c1}(k,t_i) = \begin{cases} K(t_i)H(t_i)G_S(t_i), & k = t_i, \\ K(k)\psi_{c1}(k,t_i) + \Phi(k,k-1)F_{c1}(k-1,t_i)\varphi^{-1}(k,k-1), & k > t_i, \end{cases} \tag{17}$$

$$\psi_{C2}(k,t_i) = \begin{cases} H(t_i), & k = t_i, \\ H(k)[I - \Phi(k,k-1)F_{C2}(k-1,t_i)\Phi^{-1}(k,k-1)], & k > t_i, \end{cases} \tag{18}$$

$$F_{C2}(k,t_i) = \begin{cases} K(t_i)H(t_i), & k = t_i, \\ K(k)\psi_{C2}(k,t_i) + \Phi(k,k-1)F_{C2}(k-1,t_i)\Phi^{-1}(k,k-1), & k > t_i. \end{cases} \tag{19}$$

$$\varepsilon(k,t_i)=[\vartheta^T(k,t_i)\quad \varepsilon_2^{(1)T}(k,t_i)\quad \varepsilon_2^{(2)T}(k,t_i)]^T, \tag{20}$$

$$\varepsilon_2^{(1)}(k,t_i)=\Phi(k,k-1)\varepsilon_2^{(1)}(k-1,t_i)+L(k,t_i)\xi\,(k-1), \tag{21}$$

$$\varepsilon_2^{(2)}(k,t_i)=C(k,k-1)\varepsilon_2^{(2)}(k-1,t_i)+N(k,t_i\,)\,\xi\,(k-1) \tag{22}$$

$$L(k,t_i)=\begin{cases}0, & k=t_i\,,\\ \Phi(k,k-1)[L(k-1,t_i)-G_S(k-1)]\varphi^{-1}(k,k-1), & k>t_i,\end{cases} \tag{23}$$

$$N(k,t_i)=N_1(k,t_i)+N_2(k,t_i)\,, \tag{24}$$

$$N_1(k,t_i)=\begin{cases}0, & k=t_i\,,\\ [K(k-1)H(k-1)\Phi_C(k-1,t_i)+C(k,k-1)N_1(k-1,t_i)]\times\varphi^{-1}(k,k-1)\,, & k>t_i\,,\end{cases}$$

$$N_2(k,t_i)=\begin{cases}0\,, & k=t_i\,,\\ [K(k-1)H(k-1)+C(k,k-1)N_2(k-1,t_i)]\times\Phi^{-1}(k,k-1)L(k,t_i)\,, & k>t_i\,.\end{cases}$$

It follows from equations (14) and (22) arising at time t_i that the additive gauss-Markov jump changes in the system dynamics result in the appearance of the random vector $\varepsilon(k,t_i)$ of which one of components is the vector $\vartheta\,(k,t_i\,)$, in the innovation process of the matched Kalman filter. When deducing expressions (14)-(22) we used the assumption that the transition matrix $\varphi_j(k+1,k)$ from (5) is non-singular. This assumption is usually feasible in engineering practice. The block diagram representation of the innovation process for the system (3) is presented in Fig. 2.

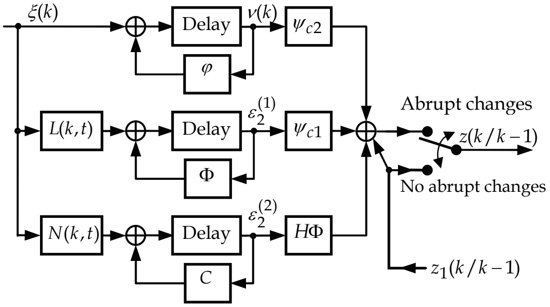


Fig. 2. Block diagram representation of the innovation process for the system with structure or parameters jumps in the system equation

Taking into consideration formulae (13) - (22) the system presented in Fig. 2 can be written in the augmented form as follows:

$$\varepsilon(k+1,t_i)=\Theta(k+1,k)\varepsilon(k,t_i)+J(k+1,t_i)\xi(k) \tag{25}$$

where the state transition and input matrices of the augmented system are calculated as: $\Theta(k+1,k)=diag\big(\varphi(k+1,k)\quad \Phi(k+1,k)\quad C(k+1,k)\big)$ and $J(k+1,k)=[I\;\; L^T\;\; N^T\;]^T$.

When the system jumps take place in the observation channel described by equation (4) the innovation process $z(k/k-1)$ has similar form to (12) :

$$z_o(k/k-1) = T_o(k,t_i)\varepsilon(k,t_i) + z_1(k/k-1)\,, \tag{26}$$

where all components of equation (26) can also be obtained in recursive form taking into consideration propagation of the signals through the Kalman filter matched with the undisturbed system :

$$T_0(k,t_i) = \left[\psi_0(k,t_i) \quad H(k)\Phi(k,k-1)\right], \tag{27}$$

$$\varepsilon\,(k,t_i) = [\vartheta^T(k,t_i)\;\varepsilon\,_1^T(k,t_i)]^T\,, \tag{28}$$

$$\psi_0(k,t_i) = H_0(k) - H(k)\Phi(k,k-1)F_0(k-1,t_i)\varphi^{-1}(k,k-1), \tag{29}$$

$$F_0(k,t_i) = K(k)\psi_0(k,t_i) + \Phi(k,k-1)F_0(k-1,t_i)\varphi^{-1}(k,k-1), \tag{30}$$

$$\varepsilon\,_1(k+1,t_i) = C(k+1,k)\varepsilon\,_1(k,t_i) + D(k+1,t_i)\xi(k), \tag{31}$$

$$D(k+1,t_i) = [K(k)\,H_0(k) + C(k+1,k)\,D(k,t_i)]\,\varphi^{-1}(k+1,k)\,, \tag{32}$$

$$C(k+1,k) = [I - K(k)H(k)]\,\Phi(k,k-1)\,. \tag{33}$$

Thus the problem under consideration can be formulated as a test of two hypotheses – the simple hypotheses H_o with respect to the composite alternative H_1 :

$$\begin{aligned} &H_0 : z(k/k-1) = z_1(k/k-1) \\ &H_1 : z(k/k-1) = T(k,t_i)\varepsilon(k,t_i) + z_1(k/k-1)\,, \end{aligned} \tag{34}$$

where $T(k,t_i)$, $\varepsilon_1(k,t_i)$ are described by (14) and (20) or (27) and (28).

Since the a priori distributions for t_i and $\vartheta(k,t_i\,)$ are unknown we have to use the generalized likelihood ratio (GLR) test. The GLR for the hypotheses (34) for $k \geq t_i$ can be written as follows (Grishin & Janczak, 2006):

$$\Lambda(k,t_i) = \Lambda(k-1,t_i\,)\frac{f[z(k/k-1)/z_{ti}^{k-1}, H_1(t_i, \varepsilon(k,t_i))]}{f[z(k/k-1)/H_0]} \tag{35}$$

Since the vector $z(k/k-1)$ in (34) is Gaussian the probability density functions $f[\cdot]$ in this expression are also Gaussian. Thus the likelihood ratio can be written in the logarithmic form:

$$\begin{aligned} &\lambda(k,t_i) = \ln\Lambda(k,t_i) = \lambda(k-1,t_i) + \ln\det P_{z1}(k) - \ln\det P_{zo}(k) + \\ &+ z^T(k/k-1)P_{z1}^{-1}(k)z(k/k-1) - [z(k/k-1) - \tilde{z}(k/k-1)]^T P_{zo}^{-1}(k)[z(k/k-1) - \tilde{z}(k/k-1)], \\ &\lambda(t_i-1,t_i) = 0\,, \end{aligned} \tag{36}$$

where $P_{z1}(k)$ is the covariance matrix of the innovation process in the matched Kalman filter (hypothesis H_o), the value

$$\tilde{z}(k/k-1)=E[z(k/k-1)/z_{ti}^{k-1},H_1]=T(k,t_i)\hat{\varepsilon}(k/k-1,t_i) \quad (37)$$

is the prediction estimate of the innovation process for jumps which have occurred at known time t_i and

$$\hat{\varepsilon}(k/k-1,t_i)=\Theta(k,k-1)\hat{\varepsilon}(k-1/k-1,t_i) \quad (38)$$

is the prediction estimation of the Kalman filter for the system described by the expressions (12) and (25).

The covariance matrix $P_{zo}(k)$ from (36) is given by

$$P_{zo}(k)=T(k,t_i)P_o(k/k-1,t_i)T^T(k,t_i)+P_{z1}(k), \quad (39)$$

where $P_o(k/k-1,t_i)$ is the covariance matrix of the estimate (38).

Therefore if the estimates $\hat{\varepsilon}(k/k-1,t_i)$ for each given t_i are calculated the maximum likelihood estimate is

$$\hat{t}_i=\arg\max_{t_i}\lambda(k,t_i). \quad (40)$$

Then the decision rule is

$$\lambda(k,\hat{t}_i)\ \overset{H_1}{\underset{H_0}{\gtrless}}\ \lambda_0(k,\hat{t}_i), \quad k-M+1\le\hat{t}_i\le k, \quad (41)$$

where $\lambda_0(k,\hat{t}_i)$ is the threshold value and $k-M+1\le\hat{t}_i\le k$ is used to avoid a growing bank of filters.

Thus the system of joint detection - estimation of jumps changes in a dynamic system consists of the basic Kalman filter, which calculates values $z(k/k-1)$, the bank of Kalman filters, which compute the likelihood ratios $\lambda(k,t_i)$ at different moments t_i=$k-M+1,..k$, the logic circuit, which selects the maximum value $\lambda(k,t_i)$ and a threshold circuit for detection of abrupt changes. Such a detection-estimation algorithm demonstrates a moderate computational burden and can be carried out in real-time systems. Its structure is presented in Fig. 3.

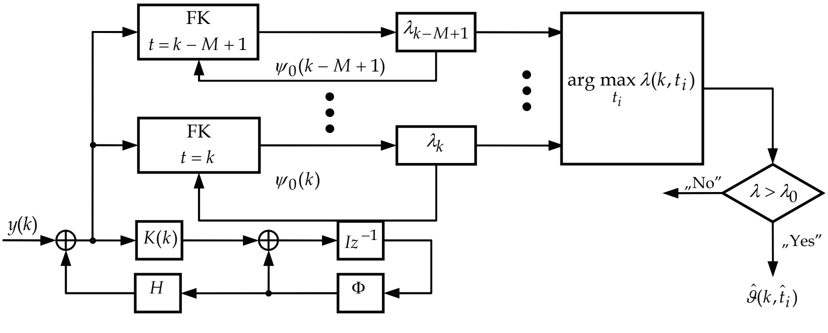


Fig. 3. Detection-estimation algorithm for the system with additive Gauss-Markov jumps

The partial estimates $\hat{\vartheta}(k,t_i)$ are obtained using $N=1 \div M$ samples of the innovation process $z(k/k-1)$ and therefore they can be obtained using the finite memory filters of which weights are calculated recursively.

3.2 Synthesis of the simplified detection-estimation algorithm

The method presented in section 3.1 is effective in supplying reasonably accurate estimates of the state vector $\vartheta(k,t_i)$. Moreover it does not require a priori knowledge of the additional system state vector $\vartheta(t_i-1,t_i)$ initial value. However high order systems results in a relatively high calculation burden. This is a consequence of the high order of the Kalman filter for the system (12)-(33) and the necessity for filter parameter calculations at every time step. To remediate these difficulties some simplifications may be introduced. As will be shown in the following section, assuming an a priori knowledge of the vector initial value $\vartheta(t_i-1,t_i)$, the decision filter equations (12) - (33) may be simplified. In this case the filter parameters may be calculated prior to the estimation process (off line). Of course, a set of adequately spaced initial values $\vartheta_j(t_i-1,t_i)$ should be assumed and the corresponding filters should be applied to the system structure (Fig. 3). Simulation investigations of the detection method have shown it to be reasonably robust to inaccuracy of the vector $\vartheta_j(t_i-1,t_i)$ value and the decision method chooses a filter initialised with $\vartheta_j(t_i-1,t_i)$ that is closest to the real one. The accuracy of the simplified method is not amenable to the method described in the previous section but the calculation burden is smaller.

A detection-estimation algorithm can be obtained in a way similar to that described in section 3.1 but with additional assumption that is known $\vartheta_j(t_i-1,t_i)$. A representation of the residuals $z(k/k-1)$ for $k \geq t_i$ can be divided into two components (one associated with the undisturbed system and the other following a given failure) and has the following form (in the case of system (4)):

$$z(k/k-1,t_i) = z_1(k/k-1) + \Psi_z(k,t_i)\phi(t_i,t_i-1)\vartheta(t_i-1,t_i) + \sum_{n=0}^{k-t_i}\Psi_z(k,t_i+n)\xi(t_i+n-1)\,, \quad (42)$$

where $z_1(k/k-1)$ is the innovation process (zero mean white noise) related to the unchanged system and the remaining elements represent the influence of specific system change on the residuals of the filter matched to the undisturbed model.

All elements $\Psi_z(k,t_i)$ depend on the system matrices, onset time and filter gain and can be calculated in a recursive way. In the case of failure described by the equation (4) these elements can be calculated as follows:

$$\Psi_z(k,t_i) = H_o(k)\Phi_z(k,t_i) - H(k)\Phi(k,k-1)F_z(k-1,t_i)\,, \quad (43)$$

$$\Phi_z(k,t_i) = \phi(k,k-1)\Phi_z(k-1,t_i)\,, \quad (44)$$

$$F_z(k,t_i) = K(k)\Psi_z(k,t_i) + \Phi(k,k-1)F_z(k-1,t_i)\,, \quad (45)$$

with initial conditions: $F_z(t_i-1,t_i)=0$, $\Phi_z(t_i-1,t_i)=\mathrm{I}$ where I is the identity matrix.

Considering equation (42) the detection problem can be formulated as a statistical test of two hypotheses (H_0, H_1), the first of which (H_0) is intended to test the presence of the white noise $z_1(k/k-1)$ and the second (H_1), the presence (H_1) of the signal $\Psi_z(k,t_i)\upsilon_{0\phi}$ to $z_{1\xi}(k/k-1)$ noise background.

$$\begin{aligned} H_0 &: z(k/k-1) = z_1(k/k-1), \\ H_1 &: z(k/k-1) = z_{1\xi}(k/k-1) + \Psi_z(k,t_i)\vartheta_{0\phi}, \end{aligned} \tag{46}$$

where $\vartheta_{0\phi} = \phi(t_i, t_i - 1)\vartheta(t_i - 1, t_i)$ and $z_{1\xi}(k/k-1)$ represents all noise components from equation (42).

Since the distribution of the onset time t_i is unknown a priori, the generalized likelihood ratio (GLR) test is used:

$$\lambda(k,\hat{t}_i) = \frac{\max\limits_{k_i} f[Z_{t_i}^k / H_1(t_i)]}{f[Z_{t_i}^k / H_0]}, \tag{47}$$

where $f[*]$ is the conditional probability density function and $Z_{t_i}^k = \{z(t_i/t_i-1), \dots, z(k/k-1)\}$.

The decision procedure has the form (48) where the generalized likelihood logarithm $\Lambda(k,\hat{t}_i)$ is compared with the threshold $\Lambda_p(k,\hat{t}_i)$. A variable threshold level is applied.

$$\Lambda(k,\hat{t}_i) \underset{H_0}{\overset{H_1}{\gtrless}} \Lambda_p(k,\hat{t}_i), \qquad \hat{t}_i = \arg\max_{t_i}\left(\Lambda(k,t_i)\right), \quad k - M + 1 \le \hat{t}_i \le k \;, \tag{48}$$

where $\Lambda(k,\hat{t}_i)$ is the logarithm of $\lambda(k,\hat{t}_i)$, M is the width of the moving window used to avoid an increasing number of additional filters matched to successive onset moments.

3.3 Threshold determination

The performance of the decision procedure is essential to the efficiency of detection and so to the quality of estimation. The general principles of the applied GLR method are well established (Willsky, 1976), (Sage & Melsa, 1971). Unfortunately, the use of the GLR approach requires knowledge of the resulting probability distributions. For instance in the detection - estimation structure based on the Kalman filter the usually resulting probability distributions are unknown and the threshold value cannot be obtain in an analytical way. The detailed solutions to the problem proposed in the literature are based on simplifications such as the use of simplified statistics (not GLR) or experimental determination. Moreover in numerical examples a constant threshold level is used. This approach is correct under steady state conditions of the object and estimator when the corresponding probability density functions are constant. It is not appropriate in a non-stationary state of the object or filter and leads to permanent additional detection delay under such conditions. The solution to the problem requires that changes in the probability distributions and application of a variable threshold level be taken into consideration. This approach allows the constant probability of false alarm (P_{FA}) to be obtained, i.e. the probability of taking the decision that a fault has occurred while the system is in a normal state. A method for obtaining a non-constant threshold level variable for a simplified filter as described in the previous section will be presented next.

The choice of a decision threshold $\Lambda_p(k,\hat{t}_i)$ can be obtained using the Neyman - Pearson criterion, where a probability P_{FA} of the false alarm level is assumed.

$$P_{FA} = \int_{\Lambda p(k,t_i)}^{\infty} f(\Lambda(k,t_i) = \Lambda_o / H_0)\, d\Lambda_o = 1 - F_{\Lambda(k,t_i)/H_0}(\Lambda_p(k,t_i))\,, \tag{49}$$

where $F_{\Lambda(k,t_i)/H_0}(\Lambda_p(k,t_i))$ is the conditional probability distribution function of $\Lambda(k,t_i)$. As seen in (49), the decision threshold can be determined with the use of $F_{\Lambda(k,t_i)/H_0}(\Lambda_p(k,t_i))$. It can be shown (Grishin, 1994) that the GLR logarithm can be computed in the following way:

$$\begin{aligned}\Lambda(k,t_i) = \frac{1}{2}\sum_{l=k_i}^{k}\Big\{ & [z(l/l-1)]^T P_{z1}^{-1}(l/l-1)[z(l/l-1)] - [z(l/l-1) - \overline{z}_{H_1}(l/l-1,t_i)]^T \times \\ & \times P_z^{-1}(l/l-1)[z(l/l-1) - \overline{z}_{H_1}(l/l-1,t_i)] + \ln[\det(P_{z1}(l/l-1)] - \ln[\det(P_z(l/l-1)]\Big\},\end{aligned} \tag{50}$$

where $P_{z1}(l/l-1)$, $P_z(l/l-1,t_i)$, and $\overline{z}_{H_1}(l/l-1,t_i)$ are covariance matrixes and the expected value of the following conditional probability distributions for the Kalman filter innovation process $z(k/k-1)$:

$$\begin{aligned} & f[z(l/l-1)/Z_{t_i}^{l-1}, H_0] = N[z(l/l-1)/H_0;\ 0,\ P_{z1}(l/l-1)]\,, \\ & f[z(l/l-1)/Z_{t_i}^{l-1}, H_1] = N[z(l/l-1)/H_1;\ \overline{z}_{H_1}(l/l-1,k_i),\ P_z(l/l-1)]\,. \end{aligned} \tag{51}$$

Taking into consideration equation (42), the parameters of the distributions (51) can be calculated as follows:

$$P_{z1}(l/l-1) = H(l)P(l/l-1)H^T(l) + R(l), \tag{52}$$

$$P_z(l/l-1,t_i) = P_{z1}(l/l-1) + \sum_{n=0}^{k-t_i} \Psi_z(l,t_i+n)\, Q_\xi(t_i+n-1)\Psi_z^T(l,t_i+n)\,, \tag{53}$$

$$\overline{z}_{H_1}(l/l-1,t_i) = E\big[z(l/l-1,t_i)/H_1\big] = \Psi_z(l,t_i)\vartheta_{0\phi}\,, \tag{54}$$

where $P(l/l-1)$ is the covariance matrix of the state vector prediction $\hat{x}(l/l-1)$ obtained in the basic Kalman filter.

Unfortunately, as follows from (50) the GLR logarithm $\Lambda(k,t_i)$ is the difference between a random variable with χ^2 distribution (first term) and a random variable with a non-central χ^2 distribution (second term) in summation with the deterministic term (third part), so an appropriate approximation of the distribution should be applied. The following approximation of the sum (50) can be assumed:

$$\Lambda(k,t_i) \approx \hat{\Lambda}(k,t_i) = \alpha_a(k,t_i) \cdot \Lambda_a(k,t_i) + c_{d0}(k,t_i), \tag{55}$$

where $\alpha_a(k,t_i)$, $c_{d0}(k,t_i)$ are coefficients, $\Lambda_a(k,t_i)$ is a random variable with a known and easy to compute distribution that would allow for approximation of the $\Lambda(k,t_i)$ distribution.

The sum (50) can be written as:

$$\Lambda(k,t_i) \approx \Lambda_S(k,t_i) = \frac{1}{2}\sum_{l=t_i}^{k}\left\{\sum_{j=1}^{s}\left(a_{0j}(l)\left[z_{0j}(l/l-1)+b_{0j}(l)\right]^2\right)\right\} + c_{d0}(k,t_i)\,, \tag{56}$$

where:

$$a_{0j}(l) = \frac{\sigma_{1j}^2(l/l-1) - \sigma_{0j}^2(l/l-1)}{\sigma_{1j}^2(l/l-1)}\,, \qquad b_{0j}(l) = \frac{\sigma_{0j}(l/l-1)\overline{z}_j(l/l-1)}{\sigma_{1j}^2(l/l-1) - \sigma_{0j}^2(l/l-1)}\,,$$

$$c_{d0}(k,t_i) = \frac{1}{2}\sum_{l=t_i}^{k}\left\{\sum_{j=1}^{s} c_{0j}(l) + \ln\frac{\det\left(P_{z1}(l/l-1)\right)}{\det\left(P_z(l/l-1)\right)}\right\} \qquad c_{0j}(l) = \frac{\overline{z}_j^2(l/l-1)}{\sigma_{0j}^2(l/l-1) - \sigma_{1j}^2(l/l-1)}\,,$$

$\sigma_{0j}^2(l/l-1)$, $\sigma_{1j}^2(l/l-1)$ are *j-th* elements from the diagonals of matrices $P_{z1}(l/l-1)$, $P_z(l/l-1)$ respectively, $\overline{z}_j(l/l-1)$ is *j-th* element of the vector $\overline{z}(l/l-1)$ and $z_{0j}(l/l-1) = \frac{z_j(l/l-1)}{\sigma_{0j}}$, so $z_{0j}(l/l-1)$ is normally distributed $N[0,\ 1]$.

Defining a new variable $\Lambda_{cd0}(k,t_i)$:

$$\Lambda_{cd0}(k,t_i) = \Lambda_S(k,t_i) - c_{d0}(k,t_i) = \tfrac{1}{2}\sum_{l=t_i}^{k}\sum_{j=1}^{s} a_{0j}(l)\left[z_{0j}(l/l-1) + b_{0j}(l)\right]^2 \tag{57}$$

we can see that $\Lambda_{cd0}(k,t_i)$ is the weighted sum (with weights $\frac{1}{2}a_{0j}(l)$) of squares of $s\cdot(k-k_i+1)$ normally distributed ($N[0,\ b_{0j}]$) variables. This leads to the idea of using the non-central χ^2 distribution as an approximation distribution (the distribution of $\Lambda_a(k,t_i)$). In the case of the non-centrality parameter (β_{nc}), the number of degrees of freedom (N_{nc}) and the coefficient $\alpha_a(k,t_i)$ (α_{nc}) must be determined. Calculation of these parameters is performed by matching three statistical moments (the first non-central, second and third central) of the variable $\alpha_a(k,t_i)\cdot\Lambda_a(k,t_i)$ (see (55)) and the sum $\Lambda_{cd0}(k,t_i)$ (see (57)).

As a result two sets of solutions ($\{\alpha'_{nc}, \beta'_{nc}, N'_{nc}\}$, $\{\alpha''_{nc}, \beta''_{nc}, N''_{nc}\}$) are obtained:

$$\alpha'_{nc} = \frac{S_{\mu2} + S_p}{S_m}\,, \qquad \beta'_{nc} = \frac{-S_m S_p}{2\alpha'_{nc}S_{\mu2} - S_{\mu3}}\,, \quad N'_{nc} = \frac{-\beta'_{nc}\left(3\alpha'_{nc}S_{\mu2} - 2S_{\mu3}\right)}{\alpha'_{nc}S_p}\,,$$

$$\alpha''_{nc} = \frac{S_{\mu2} - S_p}{S_m}\,, \qquad \beta''_{nc} = \frac{S_m S_p}{2\alpha''_{nc}S_{\mu2} - S_{\mu3}}\,, \quad N''_{nc} = \frac{\beta''_{nc}\left(3\alpha''_{nc}S_{\mu2} - 2S_{\mu3}\right)}{\alpha''_{nc}S_p}\,,$$

where

$$S_m = S_m(k,t_i) = \sum_{l=t_i}^{k}\sum_{j=1}^{s} a_{0j}(l)\cdot\left(1 + b_{0j}^2(l)\right), \quad S_{\mu2} = S_{\mu2}(k,t_i) = \sum_{l=t_i}^{k}\sum_{j=1}^{s} a_{0j}^2(l)\cdot\left(1 + 2b_{0j}^2(l)\right),$$

$$S_{\mu3} = S_{\mu3}(k,t_i) = \sum_{l=t_i}^{k}\sum_{j=1}^{s} a_{0j}^{3}(l)\cdot\left(1+3b_{0j}^{2}(l)\right), \qquad S_p = \sqrt{S_{\mu2}^{2} - S_m S_{\mu3}} .$$

The set with $\beta_{nc} \geq 0$ and $N_{nc} > 0$ should be taken as the final solution. Moreover at the beginning the following condition should be checked: $S_{\mu2}^2 - S_m S_{\mu3} \geq 0$

If the condition is not fulfilled the above approximation cannot be calculated. In this case an approximation using the central χ^2 distribution was also derived and tested. However this is less accurate in cases of low value of M (moving widow width) but has no numerical constraint and needs less computation. Two of the required parameters (the number of degrees of freedom and the coefficient $\alpha_a(k,t_i)$) can be determined by matching two distribution parameters (mean value and variance) of the variable $\alpha_a(k,t_i)\cdot\Lambda_a(k,t_i)$ and the sum $\Lambda_{cd0}(k,t_i)$.

In practice, the number of degrees of freedom obtained in both approximations is not usually an integer number, so the distributions cannot be computed as typical central χ^2 or noncentral χ^2 distributions. Instead of the central χ^2 distribution function the Gamma distribution function (with parameters $N_c(k,t_i)/2$ and 2) can be used. The other distribution can be calculated in the following way (modification of the standard numerical procedure):

$$F_{\Lambda(k,t_i)}(x) = \sum_{i=0}^{\infty}\left(\frac{(\beta_{nc}/2)^i}{i!}e^{-\frac{\beta_{nc}}{2}}\right)\cdot P\left\{\chi^2_{N_{nc}+2i} \leq x\right\} = \sum_{i=0}^{\infty} f_{Po}(i;\ \beta_{nc}/2)\cdot F_{\Gamma}(x;\ N_{nc}+2i,\ 2), \tag{58}$$

where f_{Po} - Poisson probability density function, F_{Γ} - Gamma cumulative distribution function.

The performance of the proposed method was tested by means of numerical simulations.

The results presented below were obtained for the first order process model and on the basis of additive changes to the observation equation (see (4)) with the following parameters: $\Phi(k,k-1)=1$, $H(k)=1$, $Q(k)=(0.2)^2$, $H_o(k)=1$, $R(k)=10^2$, $\phi(k,k-1)=1$, $Q_{\xi}(k)=(0.8)^2$, $\upsilon(t_i-1,t_i)=1$, $x(0)=x_0$: $N[x_0;12,10]$. At the beginning the accuracy of the approximations was tested using Monte Carlo simulation (number of simulations $N_s = 100\,000$). In Fig. 4 the distribution of $\Lambda(k,t_i)$ (determined by numerical experiment - "ex") and analytically calculated approximations ("nc" - noncentral, "c" - central χ^2 distribution) are compared for the case of $M=1$ (the smallest width of the moving window) and $M=5$ (medium value of M).

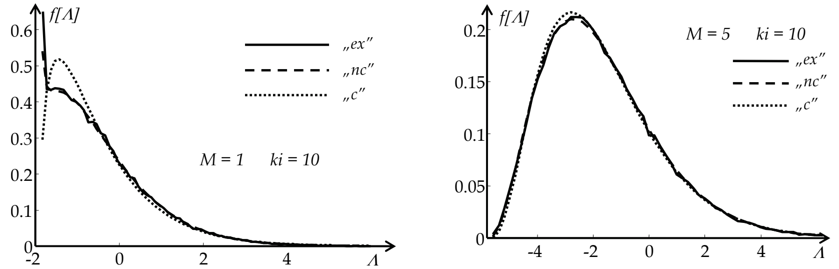


Fig. 4. Distribution of $\Lambda(k,t_i)$ ("ex") and its approximations ("nc", "c") for $M=1$, $M=5$

As can be concluded from Fig. 4 the approximation "nc" is precise for all *M*. The accuracy of approximation "c" is not so exact, especially for low value of *M* and low threshold level (high P_{FA}). These observations were confirmed by analytical measures. The Kullback measure of distances between the distribution of $\Lambda(k,t_i)$ and its approximations were calculated. The results are shown in table 1.

	M=1	M=2	M=3	M=4	M=5
"nc"	0.0018	0.0023	0.0023	0.0024	0.0022
"c"	0.0161	0.0139	0.0110	0.0078	0.0058

Table 1. Kullback measure of distances between the distribution of $\Lambda(k,t_i)$ and its approximations.

The numerical data presented in table 1 confirm that the approximation "c" is far less accurate then "nc" for small *M* but is comparable for higher *M* values ($M \geq 5$).
Next, the threshold level was calculated. A constant probability P_{FA} of false alarm was assumed. This caused a change in the threshold value. The results are shown in Fig. 5. It should be added that the character of the changes depends on system and failure parameters and can vary from that presented.

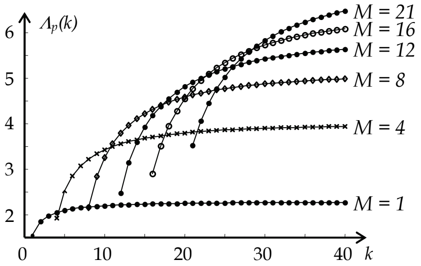


Fig. 5. Variation of threshold level in the case of constant P_{FA}

Finally a check of the validity of the threshold algorithms was performed by testing the outcome probability P_{FA} of false alarm. The results of $N_s = 10^6$ Monte Carlo simulations are shown in Fig. 6. There were two P_{FA} values assumed: $P_{FA} = 0.01$ and $P_{FA} = 0.001$. The parameter is verified for $M = 1,...,5$. The mean value of P_{FA} was calculated and is shown as $\overline{P_{FA}}$.

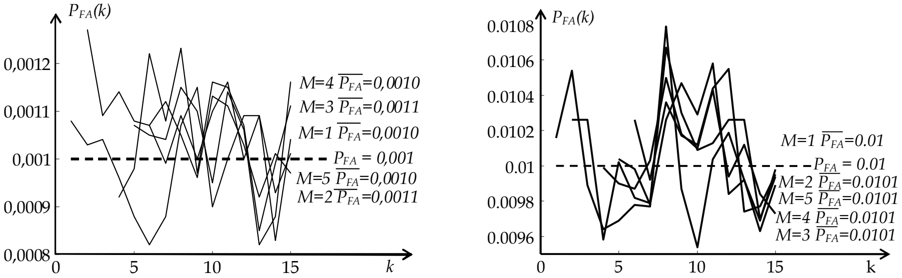


Fig. 6. P_{FA} variation in time when thresholds were calculated for $P_{FA} = 0.001$, $P_{FA} = 0.01$

It can be seen from Fig. 6, that the proposed method demonstrates high accuracy. The maximum difference between the obtained and assumed P_{FA} was less than $\Delta P = 8 \cdot 10^{-4}$. The difference diminishes as the number of simulations increases. Mean values $\overline{P_{FA}}$ are very close to the assumed P_{FA}.

The simulation results demonstrate the effectiveness of the proposed probability distribution approximations. The method allows a constant rate of the probability of false alarm to be obtained in the non-stationary state of the object or filter.

4. Tracking of maneuvering targets

The demands of high precision tracking and guidance systems require accurate state estimation of the targets. A variety of maneuvering target tracking methods have been proposed in the literature. The main principles and techniques used to track target in real situations and a comparative evaluation of some of the algorithms can be found in (Blackman & Popoli, 1999). In recent years a great deal of new maneuvering target tracking algorithms have been proposed. Among them, there are algorithms such as those which use the input estimation (IE) technique, variable dimension (VD) filtering, multiple hypothesis tracking (MHT) and the interacting multiple model (IMM) approach (Blackman & Popoli, (1999), (Bar-Shalom & Fortmann, 1988), (Bar-Shalom et al., 2001), (Li & Bar-Shalom, 1993). Although the structure of many optimal algorithms of maneuvering target tracking is known, the computational complexity often limits theirs practical realization. Many different tracking algorithms have been developed for the purposes of computational feasibility. Some of them use combined techniques such as IMM/IE, IE/VD (Blackman & Popoli, 1999). For a mathematical description of a maneuver the following models are usually used: white noise models, a noisy jerk as a maneuver model, non-random maneuver models and combined target maneuver models. The additive Gauss-Markov Models (AGMM) presented earlier enable a realistic but simple description of quite complex changes in a real process to be obtained. The maneuver of a moving object manifests as a change in acceleration. Usually the change is modelled as a step or ramp function. In most applications this approximation is sufficient but for precise or close distance tracking the change model should be more representative. Reasonably accurate maneuver models incorporate acceleration changes in the form of inertial system step response in the presence of correlated noise. The acceleration dynamics (Blackman & Popoli, 1999) can be described as:

$$\dot{a}(t) = -\frac{1}{\tau}a(t) + \frac{\beta}{\tau}[1(t,t_i) - 1(t,t_j)] + w(t)\,, \tag{59}$$

where $a(t)$ is acceleration, β is acceleration level, τ is correlation time, $w(t)$ is zero mean white noise with covariance Q_w and $1(t,t_i)$ is unit step function with onset time t_i and t_j is a time of maneuver termination.

An example of acceleration ($\beta = 19.6\frac{m}{s^2}$ for $Q'_w = 1\frac{m^2}{s^4}$ and $Q''_w = 9\frac{m^2}{s^4}$) used for simulation is presented in Fig. 7.

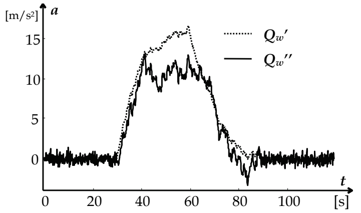


Fig. 7. Realization of an acceleration modelling maneuver

Defining the components of the state vector in terms of position, velocity and acceleration, the target dynamics model on one axis can be written as:

$$\dot{x}(t) = F(t)x(t) + B(t)w(t) + \frac{\beta}{\tau}B(t)[1(t,t_i) - 1(t,t_j)]\,, \tag{60}$$

where matrices $F(t)$, $B(t)$ are defined as:

$$F(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\frac{1}{\tau} \end{bmatrix},\ B(t) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

A discrete form of the model (60) is given by:

$$x(k+1) = \Phi_d(k+1,k)x(k) + w_d(k) + B_d(k+1)[1(k+1,t_i) - 1(k+1,t_j)], \tag{61}$$

where the transition and system input matrices take the values:

$$\Phi_d(k+1,k) = \begin{bmatrix} 1 & T & \tau^2(\frac{T}{\tau} - 1 + \exp(-\frac{T}{\tau})) \\ 0 & 1 & \tau(1 - \exp(-\frac{T}{\tau})) \\ 0 & 0 & \exp(-\frac{T}{\tau}) \end{bmatrix}, \qquad B_d(k) = \begin{bmatrix} \beta\tau^2(1 - \frac{T}{\tau} + \frac{T^2}{2\tau^2} - \exp(-\frac{T}{\tau})) \\ \beta\tau(\frac{T}{\tau} - 1 + \exp(-\frac{T}{\tau})) \\ \beta(1 - \exp(-\frac{T}{\tau})) \end{bmatrix},$$

where T is the sampling time and $w_d(k)$ is zero mean white noise with covariance matrix:

$$Q_d(k) = E\left[w_d(k) * w_d^T(k)\right] = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix},$$

$$q_{11} = \sigma^2\tau^4\left[1 + 2\left(T/\tau\right) + 2\left(T/\tau\right)^2 + 2/3\left(T/\tau\right)^3 - \left(2\left(T/\tau\right) + \exp\left(-T/\tau\right)\right)^2\right],$$

$$q_{12} = q_{21} = \sigma^2\tau^3\left[1 - 2\left(T/\tau\right) - 2\exp\left(-T/\tau\right) + \left(T/\tau + \exp\left(-T/\tau\right)\right)^2\right],$$

$$q_{22} = \sigma^2\tau^2\left[-3 + T/\tau + 4\exp\left(-T/\tau\right) - \exp\left(-2T/\tau\right)\right], \qquad q_{23} = q_{32} = \sigma^2\tau\left[1 - \exp\left(-T/\tau\right) + \exp\left(-2T/\tau\right)\right],$$

$$q_{13} = q_{31} = \sigma^2\tau^2\left[1 - 2\left(T/\tau\right)\exp\left(-T/\tau\right) - \exp\left(-2T/\tau\right)\right], \qquad q_{33} = 1 - \exp\left(-2T/\tau\right).$$

This complex model can be described by means of AGMM additive to the state (63). Maneuver is treated as a change in the order of target dynamics from the second (62) to the third (61) and is modelled by means of vector $\vartheta(k+1, t_i)$ (64):

$$x(k+1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(k) + w_1(k), \tag{62}$$

$$x(k+1) = \Phi(k+1,k)x(k) + w(k) + G(k+1)\vartheta(k+1,t_i), \tag{63}$$

$$\vartheta(k+1,t_i) = \phi(k+1,k)\vartheta(k,t_i) + \xi(k), \tag{64}$$

where corresponding matrices take the following form:

$$\Phi = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \quad \phi(k+1,k) = \begin{bmatrix} e^{-\frac{T}{\tau}} & 1 \\ 0 & 1 \end{bmatrix}, \quad \vartheta(t_i - 1, t_i) = \begin{bmatrix} 0 \\ \beta\left(1 - e^{-\frac{T}{\tau}}\right) \end{bmatrix},$$

$$G = \begin{bmatrix} \tau^2\left(1 + \left(\frac{T}{\tau} - 1\right)e^{\frac{T}{\tau}}\right) & \tau^2\left(-1 + \left(1 - \frac{T}{\tau}\right)e^{\frac{T}{\tau}} + \beta\left(1 - \frac{T}{\tau} - \frac{T^2}{2\tau^2} - e^{\frac{-T}{\tau}}\right)\right) \\ \tau\left(-1 + e^{\frac{T}{\tau}}\right) & \tau\left(1 - e^{\frac{T}{\tau}} + \beta\left(-1 + \frac{T}{\tau} + e^{\frac{-T}{\tau}}\right)\right) \end{bmatrix}.$$

The performance characteristics of the proposed method were compared with the widely used IMM technique (Bar-Shalom at al., 2001), (Blackman & Popoli, 1999), (Li & Bar-Shalom, 1993) using Monte Carlo simulations. Maneuver was modelled as acceleration change described by the scenario shown in Fig. 7 ($t_i/T = 300$, $t_j/T = 600$ - t_i, t_j - onset and termination time). For a simulation of the IMM algorithm three models of the movement have been used: the constant velocity model, Singer's model with a correlation time $\tau = 10s$ and $\sigma_m^2 = 1\frac{m^2}{s^4}$, and model described by Singer's model with constant acceleration of $\beta = 19.6\frac{m}{s^2}$. The elements of transition matrix are equal to $p_{ii} = 0.9$ on the diagonal and $p_{ij} = 0.05$ elsewhere. Initially all models are assumed equiprobable.

In the Fig. 8 the root mean square errors (RMSE) of distance and velocity estimates are shown. As follows from the schedules, the AGMM algorithm demonstrates a better estimation performance in comparison with the IMM method everywhere apart from transient parts of the manouver. Smaller estimation errors are achieved due to adaptation of the AGMM filter dimension with respect to the real process model.

5. Failure detection in a multisensor integrated system

5.1 Fault tolerant airborne navigational system structure

As an example of the application of the methods developed to the problem of fault detection-identification, let us consider reliable data processing in integrated GPS-based

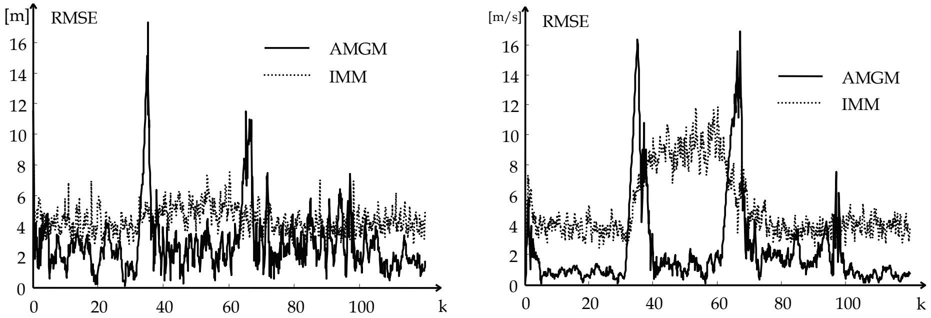


Fig. 8. RMS error of position (left) and velocity (right)

airborne navigational equipment (Brown & Hwang, 1987), (Grishin, 2000). The possible structure of a real airborne navigational aid is presented in Fig. 9. It may consist of a number of radio-navigational and self-contained sensors such as the Microwave Landing System (MLS) or the Instrument Landing System (ILS), the VOR/DME system, the Global Positioning System (GPS), the Inertial Navigation System (INS) and the System of Air Signals (SAS) supplying barometrical and altitude information (Fadden & Schwab, 1989). Each sensor has independent diagnostic facilities (DF) which check the sensor serviceability and control a state matrix circuit (SM). The latter determines the availability of the sensor output data. When a sensor is out of order the integrated filter does not use the sensor's data

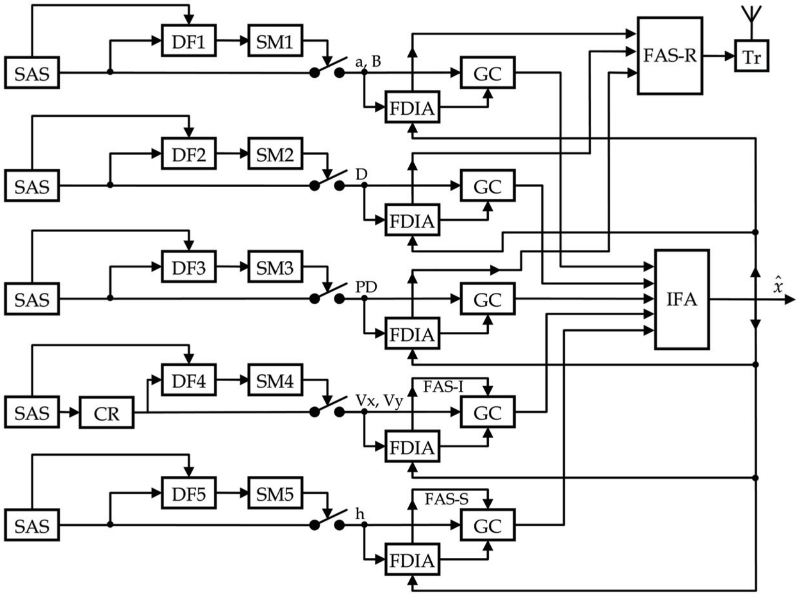


Fig. 9. The structure of the fault-tolerant airborne navigation equipment (DF - diagnostic facilities, SM - state matrix circuit, CR - coordinate recalculation, FDIA - fault detection-identification algorithm, GC - gate circuit, FAS - failure alarm signal, Tr - transmitter, IFA - integrated filtering algorithm)

and the plane state vector estimate is computed with the aid of normally operating sensors only. The corresponding failure alarm signal (FAS) has to be transmitted to the system's users. It should be noted that the diagnostic facilities are able to detect only solid failures in the airborne equipment and cannot determine faults in the ground-based or space-based facilities.

In the absence of failures the integrated algorithm is usually based on non-linear modifications of the Kalman filter (Sage & Melsa, 1971).

The main objective of this section is to present the algorithms for data processing in the multisensor GPS-based airborne navigational equipment which, on the one hand would be tolerant to possible failures of the information sources and on the other hand could enhance the integrity of the whole navigational system. The main complicating factors accompanying the solution to the problem are: rapid changes to the satellite geometry, the presence of receiver clock error, increased dynamics of the aircraft and availability of additional information from a number of the sensors mentioned above. In this case, fault-tolerant signal processing can be based on analytical and/or physical redundancy (Grishin, 2000). One of the main characteristics for a system of this kind is integrity (Brown, 1988) which can be thought of as the ability of the system to provide a timely warning to users as to when the system should not be used for navigation. The integrity performance characteristics such as integrity warning time and accuracy threshold requirements vary with the phase of flight (oceanic en route, domestic en route, terminal area and nonprecision approach). Higher reliability and integrity of airborne equipment may be achieved as a result of the detection of individual sensor failures and computation of the state estimates using data which have their origin in the normal operated sensors only.

For modelling the failures of individual subsystems the additive Gauss-Markov models considered in section 3 were used:

1. jump biases in observations (equation 4) with unknown onset time and value (antenna beam distortion, time jumps in the GPS due to a gradual degradation of the satellite clock, random bias in the INS due to drift of gyroscopes and so on);
2. random drifts (ramp-type incipient failures) which can be caused by multiple path propagation effects in the ILS, frequency shifts in the GPS, soft failures in the INS and a number of other failures that can be described by the equation (3).

Furthermore, it is necessary to take into consideration multiple malfunctions that can arise in the sensors which result in outliers at the input of the integrated estimation filter. These outliers can be caused by pulse interferences, by signal amplitude fluctuations or by clutter or intentional jamming.

It is assumed here that outliers have a normal pdf $N(0,\tilde{R}_{ki})$ with a covariance matrix $\tilde{R}_{ki}=\sigma_{ki}^2 R(k)$, where $\sigma_{ki}^2 >> 1$ depending on the signal amplitude A_i. This means that when the outliers occur the pdf of measurements changes and their variances take on M different values.

Thus the observation equation can be written as follows:

$$y(k)=H(k)+\gamma_i(k)\upsilon(k), \tag{65}$$

where $H(k)$ is the observation matrix, the switching function $\gamma_i(k)$ takes the value 1 when the outliers and multiplicative interferences are absent (normal measurement process) and $\gamma_i(k)=\sigma_{ki}^2$, under abnormal measurement conditions and $v(k)$ is the normal measurement noise with the covariance matrix $R(k)$ and zero mean vector.

In the general case, the switching function can be modeled by the finite state Markov chain of which initial probabilities and the transition matrix are known or unknown depending upon a priori information about the spectral characteristics of the outliers.

In the situation when not all sensors have failed, using the integrated filter estimates makes it possible to detect failures of the individual sensors and to inform the user about them.

Our aim is to develop an integrated filter algorithm which would be fault-tolerant in the presence of the failures and outliers mentioned above. Such an algorithm has been developed for the aircraft state vector which contains nine components such as the *x, y, z* - position, *ΔVx, ΔVy, ΔVz* - INS velocity errors, an altimeter bias and the GPS clock's shift and velocity. But the above mentioned limitations concerning the state vector are not fundamental and all the results can be applied to an arbitrary case.

The state and measurement equations in our case can be written in the following form:

$$x(k+1) = \Phi(k+1)x(k) + U(k) + w(k) + \vartheta_s(k,t_i)\,1(k,t_i), \tag{66}$$

$$y(k) = h\left[x(k)\right] + b(k) + \gamma(k)v(k) + \vartheta_o(k,t_i)1(k,t_i), \tag{67}$$

where $x(k)$ is the aircraft state vector, $U(k)$ is the input control vector, $\vartheta_s(k,t_i)$ is a failure bias of the state vector arising at random time t_i, $1(k,t_i)$ is the unit step function, $w(k)$ is the system input noise vector, $y(k)$ is the measurement vector, $b(k)$ is the unknown constant bias vector, $\vartheta_0(k,t_i)$ is the Markov drift which models incipient failures of such sensors as INS, SAS and errors due to the influence of multipath effects in the ILS, $v(k)$ is zero mean observation noise with covariance matrix $R(k)$, and $\gamma(k) = \{1, \sigma > 1\}$ is a multiplier which describes the outliers in the observation channel.

The incipient failure model is described by (66). The a priori distributions of a random value t_i are assumed to be unknown.

The time dependence of the sequence $\gamma(k)$ can be described by a stationary Markov chain, for which the initial probability vector $P^{\sigma}(0)$ and transition matrix $P\gamma^{ij}$ are

$$P^{\sigma}(0) = \begin{bmatrix} P_1(0) \\ P_{\sigma}(0) \end{bmatrix}, \; P_{\gamma}{}^{ij} = \begin{bmatrix} P_{11} & P_{1\gamma_0} \\ P_{\sigma 1} & P_{\sigma\sigma} \end{bmatrix}. \tag{68}$$

Thus the system and failure model described by (66)-(67) differ from those proposed in (Patton et al., 1989). Firstly, the failures here are treated as an additive Markov process in the dynamic or observation equations with an unknown onset time and can describe both deterministic and stochastic failure models. Secondly the outliers in the observation channels are present at the system input simultaneously with possible failures. Thus, such an approach makes it possible to describe both types of failures models - deterministic and stochastic.

5.2 Algorithms for fault-tolerant data processing

As it follows from (66)-(67), the development of a reliable integrated filter can be advanced by using non-linear filtering theory (Ristic et al., 2004). However, immediate application of this theory yields too complicated an algorithm to use in real-time systems because of the requirement for an infinite amount of memory. To overcome these difficulties it is necessary to decompose the algorithm and to introduce the fault detection procedure as inherent part of the process. Therefore it is necessary to modify the problem in the direction of simplification. A simplification of this kind leads to a suboptimal algorithm which can be applied to a real time system with limited memory requirements.

The first step in this direction is to separate the failure detection - estimation problem into an independent task. A solution can be found if one knows the sensor error statistical models and the integrated filter estimates. Using the approach presented in section 3 it is possible to estimate failure onset time $\hat{t}_i$ and the value of the vector $\hat{\vartheta}(k,t_i)$. So in observation equation (67) vector $\hat{\vartheta}(k,t_i)$ can then be considered to be a known value.

The second step in solving the problem is synthesis of the integrated filtering algorithm so that it will be sufficiently robust with respect to the presence of malfunctions (outliers) in the observation channels.

In order to cope with this problem for the system described by equations (66) and (67), it is necessary to use a general nonlinear filtering theory approach (Ristic et al., 2004). In this case the estimates of the dynamic system state vector can be found as a conditional mean of the following form (Janczak & Grishin, 2008):

$$\hat{x}(k / k) = E[x(k) / Y_1^k] = \sum_{i\in 2^k} \hat{x}^i(k / k)P(\bar{\Gamma}_k^i / Y_1^k), \tag{69}$$

where $Y_1^k = \{y(1), y(2), \dots, y(k)\}$ is the sequence of the input data, $\bar{\Gamma}_i^k = \{\gamma(1), \gamma(2), \dots, \gamma(k)\}$ denotes the realization of the switching function and

$$\hat{x}^i(k/k) = E[x(k)/Y_1^k, \bar{\Gamma}_i^k], \tag{70}$$

are partial estimates that are calculated for each realization of the switching function. Thus the optimal estimation algorithm requires infinitely increasing memory and cannot be realized in practice. Practical realization can only be achieved by using different approximations of the pdf of the estimates (69). One of the possible approaches to solving this problem is using the Gaussian approximation method (Ristic et al., 2004). In such an approach the state vector estimates $\hat{x}(k / k)$ can be expressed as the weighted sum of the partial estimates $\hat{x}^i(k / k)$ corresponding to the presence and absence of the outliers in the measurements:

$$\hat{x}(k/k) = \sum_{i\in 1,\sigma} \hat{x}(k/k, \gamma_i(k) = \sigma_{ki}^2)P(\gamma_i(k) = \sigma_{ki}^2 / Y_1^k). \tag{71}$$

The posterior probability of the measurement channel state $P(\gamma_i(k) = \sigma_{ki}^2 / Y_1^k)$ depends on the outlier stochastic characteristics. If the outliers are statistically independent, the probability can be found from:

$$p_{i/k} = \frac{f(y(k) / \gamma_i(k) = \sigma_{ki}^2, Y_1^{k-1})p_{i/k-1}}{\sum_{j=1,\sigma} f(y(k) / \gamma_j(k) = \sigma_{ki}^2, Y_1^{k-1})p_{j/k-1}}, \tag{72}$$

where $p_{i/k}$ is the a posteriori probability of the measurement noise covariance matrix $\widetilde{R}_{ki} = \sigma_{ki}^2 R(k)$.

These probabilities can be calculated in real time using current data at the filter input based on the pdf $f(y(k)/\gamma_i(k)=\sigma_{ki}^2,Y_1^{k-1})$ of predicted estimates (Bar-Shalom et al., 2001). When the fluctuations and outliers are independent in time, the probability $p_{i/k-1}=q_{ki}$, where q_{ki} are the a priori probabilities.

It can be shown that for a system which contains N observation channels with outliers, this method yields the following expression for the state vector estimate (Grishin, 2000):

$$\begin{aligned}\hat{x}(k/k)&=\sum_{i_1}^{\sigma_1}\sum_{i_2}^{\sigma_2}\dots\sum_{i_N}^{\sigma_N}\hat{x}_{i_1,\dots i_N}(k/k)\cdot p(i_1,\dots i_N/k)=\Phi\hat{x}(k-1/k-1)+\\&+\sum_{j=1}^{N}\left\{\sum_{i_1}^{\sigma_1}\sum_{i_2}^{\sigma_2}\dots\sum_{i_N}^{\sigma_N}P_{i_1,\dots i_N}(k/k)H_j^T(k)\cdot\left[i_j^2R_{jj}(k)\right]^{-1}p(i_1,\dots i_N/k)\times\right.\\&\left.\times\left[y_j(k)-H_j(k)\Phi_j\cdot\hat{x}(k-1/k-1)\right]\right\},\qquad i_j=1,\sigma_j,\ \ j=1,\dots,N,\end{aligned}\tag{73}$$

where $\hat{\mathrm{x}}_{i_1,i_2,\dots,i_N}(k/k)$ is a partial estimate of the state vector for certain failure realisation in the observation channels (sensors of navigational information), $p(i_1,i_2,\dots,i_N/k)=p(\gamma(1)=i_1,\gamma(2)=i_2,\dots,\gamma(N)=i_N/y(1),\dots,y(k))$ are the a posteriori probabilities of these realisations, $P_{i_1,i_2,\dots,i_N}(k/k)$ is the update covariance matrix of the partial estimate, $i_j=1,\ \sigma$ are values of the multiplier $\gamma(k)$ in the j-th channel for a normal and failure state of performance, and $y_j(k)$ measurements at the output of the j-th navigational information source.

It can be shown that a posteriori probabilities are calculated in real time as follows:

$$\begin{aligned}p(i_1,i_2,\dots,i_N/k)&=f\left[y(k)/\gamma^*_{i_1,i_2,\dots,i_N}(k),Y_1^{k-1}\right]\times p\left[\gamma^*_{i_1,i_2,\dots,i_N}(k)/\mathrm{Y}_1^{k-1}\right]\times\\&\times\left\{\sum_{i_1=1}^{\sigma_1}\sum_{i_2=1}^{\sigma_2}\dots\sum_{i_N=1}^{\sigma_N}f\left[y(k)/\gamma^*_{i_1,\dots,i_N}(k),\mathrm{Y}_1^{k-1}\right]\times p\left[\gamma^*_{i_1,i_2,\dots,i_N}(k)/\mathrm{Y}_1^{k-1}\right]\right\}^{-1},\end{aligned}\tag{74}$$

where $f\left[y(k)/\gamma^*_{i_1,\dots,i_N}(k),Y_1^{k-1}\right]$ is a value of the likelihood function at the point y(k), $p\left[\gamma^*_{i_1,i_2,\dots,i_N}(k)/Y_1^{k-1}\right]$ - a priori probability of a certain combination of channel observation serviceability, which can be calculated on the basis of a previous value of p and the Markov chain characteristics:

$$p[\gamma^*_{i_1,\dots,i_N}(k)/\mathrm{Y}_1^{k-1}]=\prod_{j=1}^{N}\ \sum_{n=1}^{\sigma}P_{n\,ij}^{(j)}\mathrm{P}[i_1,\dots,i_N/k-1],\qquad i_j=1,\sigma\ ,\tag{75}$$

where $P_{nij}^{(j)}$ is the transition matrix elements of the Markov chain $\gamma^{(j)}(k)$ in the j-th observation channel. The algorithm described by (73) - (75) can be thought of as a soft multichannel outlier screening procedure which is correct for arbitrary values of $\sigma>1$ (not necessarily for large ones).

Let us consider then, the part of the system structure (Fig. 9) which is responsible for a decision of the failure detection-estimation problem in each information channel (sensor).

All of them contain a fault detection-identification algorithm (FDIA), which is used for estimating the failures and for generating the failure alarm signal (FAS) to inform the user.

The failure detection-identification algorithm is designed on the basis of the GLR approach for an additive Gauss-Markov model of the system failures. It can be constructed on the assumption that no a priori information about failure onset time and the initial conditions of vector $\vartheta(k,t_i)$ exists.

Since the failure vector $\vartheta(k,t_i)$ is part of $\varepsilon(k,t_i)$ its estimate is also known. This estimate can be used to cancel the input data biases, for example. The block diagram for a cancellation of this kind is presented in Fig. 10.

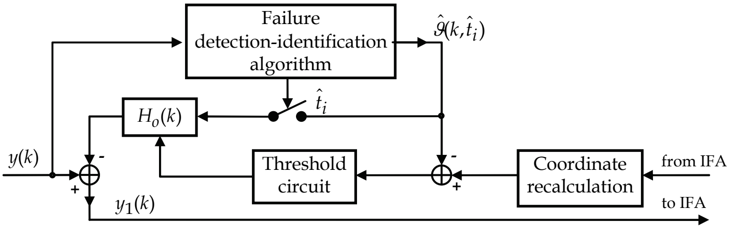


Fig. 10. The fault bias cancellation method

After detecting abrupt changes to the sensor output, it is necessary to control the presence of biases in the output estimates of the IFA to distinguish sensor failures from aircraft manoeuvres.

It should be noted that the proposed structure also makes it possible to isolate failures, that is, to determine if failures have occurred in the airborne navigation equipment or in the space-based facilities. This can be realised by comparing the data of the FDIA and content of the state matrix circuits. Following this, the failure alarm signal should be generated and transmitted to the users.

6. Conclusion

We have presented a new recursive algorithm for joint detection and estimation of jump changes in the dynamics and measurements of linear discrete-time systems in the presence of outliers in observations. The algorithm has been developed on the basis of the GLR method. The jumps were modelled as Gauss-Markov biases in state and observation equations. The structure of the algorithm is sufficiently simple to enable it to be applied in real-time systems with a relatively limited computational burden. The proposed models describe a wide class of dynamic systems with jump parameters. The detection-estimation algorithm developed, was successfully applied to the problem of radar maneuvering target tracking and fault-tolerant signal processing for enhancing the integrity and reliability of airborne navigation equipment. Simulation results revealed good estimation properties for the algorithm.

7. References

Bar-Shalom, Y. & Fortmann, T. (1988). *Tracking and data association"*, Academic Press, N.Y.

Bar-Shalom, Y.; Li, X. & Kirubarajan, R. (2001). *Estimation with applications to tracking and navigation*, John Wiley & Sons, New York.

Blackman, S. & Popoli, R. (1999). *Design and analysis of modern tracking systems.* Artech House, Boston.

Brown, G. & Hwang, P. (1987). GPS failure detection by autonomous means within cockpit. *Navigation*, Vol. 33, No. 4, 1987, pp. 335-353.

Brown, A. (1988). Civil Aviation Integrity Requirements for the Global Positioning System. *Navigation*, Vol. 35, No. 1, 1988, pp. 23-40.

Fadden, D. & Schwab, R. (1989) Aircraft Interface with Future ATC System, *Proc. IEEE,* Vol. 77, No 11, pp. 1745-1751.

Gertler, J. (1998). *Fault Detection and Diagnosis in Engineering Systems*, Marcel Dekker, Inc,.N.Y.

Gini, F. & Rangaswamy, M./Ed. (2008). *Knowledge-based radar detection, tracking and classification*, John Wiley&Sons, ISBN 978-0-470-14930-0, N. Y.

Grishin, Yu. (1994). An Application of the Additive Gauss-Markov Models of Discrete-Time Dynamic Systems to the Problem of Abrupt Changes Detection, *Proceedings of Int. AMSE Conf. Systems: Analysis, Control & Design*, pp. 211-220, v.1, July 1994, Lyon.

Grishin, Yu. (2000). Reliable data processing in an integrated GPS-based airborne navigational equipment, *Proceedings of the European Communication Conference EUROCOMM 2000*, pp 91-94, ISBN 0-7803-6323X, Munich (Germany), May 17, 2000, IEEE, NJ.

Grishin, Yu. & Janczak, D. (2006). Joint Adaptive Detection - Estimation Algorithm for Maneuvering Target Tracking, *Proceedings of the International Radar Symposium IRS 2006*, pp. 375-378, ISBN 83-7207-621-9, Krakow (Poland), 24-26 May 2006, PIT, Warszawa.

Janczak, D. & Grishin, Yu. (2008). The GLR failure detection algorithm for a class of nonlinear dynamic models with application to radar tracking problems, *Proceedings of the International Radar Symposium IRS 2008*, pp. 233-236, ISBN 978-83-7207-757-8, Wroclaw (Poland), 21-23 May 2008, PIT, Warszawa.

Katayma, T. & Sugimoto, S. (ed) (1997). *Statistical methods in control and signal processing*. Marcel Dekker, Inc., N.Y.

Li, Rong. & Bar-Shalom, Y. (1993). Performance prediction of the interacting multiple model algorithm, *IEEE Trans., Vol. AES-29*, No 3, 1993, pp. 755-771.

Mazor, E.; Dayan, J.; Averbuch, A. & Bar-Shalom, Y. (1998). Interacting multiple model methods in target tracking: a survey, *IEEE Trans., Vol. AES-34*, No 1, 1998, pp. 103-123.

Patton, R.; Frank, P. & Clark, R. (1989). *Fault diagnosis in dynamic systems. Theory and applications*, Prentice Hall, N.Y.

Ristic, B.; Arulampalam, S. & Gordon, N. (2004). *Beyond the Kalman filter-Partical filters for tracking applications*, Artech House, ISBN 1-58053-631-x, Boston.

Sage, A. & Melsa, I. (1971). *Estimation Theory with Applications to Communication and Control*, Mc Graw-Hill, N. Y.

Sorenson, H. (ed) (1985). *Kalman filtering: theory and application,* IEEE Press, Piscataway, NJ.

Whang, I.; Lee, J. & Sung, T. (1994). Modified input estimation technique using pseudoresiduals, *IEEE Trans., Vol. AES-30*, No 1, 1994, pp.220-227.

Willsky, A. (1976). A Survey of Design Methods for Failure Detection in Dynamic Systems, *Automatica*, Vol. 12, 1976, pp. 601-611.