

ISSN: 1813-3894

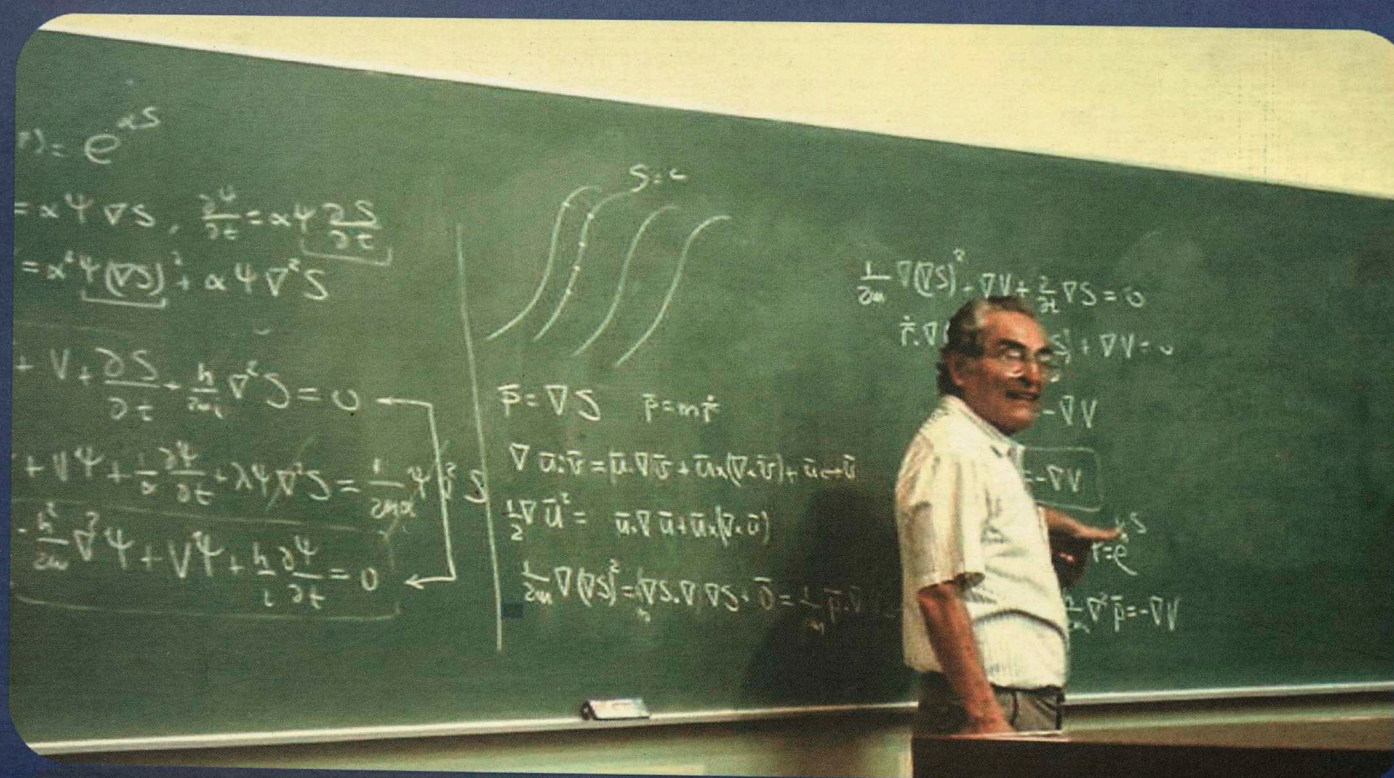
REVCIUNI

VOLUMEN 15

NÚMERO 1

Enero - Diciembre 2012

Revista de la Facultad de Ciencias de la UNI



Proceedings XV MEETING OF PHYSICS

Festschrift for
HOLGER G. VALQUI

Edited by
Orlando Pereyra, Ener Salinas and Harold Blas



UNIVERSIDAD NACIONAL DE INGENIERÍA

LIMA - PERÚ

AUTORIDADES UNIVERSITARIAS

RECTOR

Dr. Aurelio Padilla Ríos

PRIMER VICERRECTOR

Geol. José Martínez Talledo

SEGUNDO VICERRECTOR

Mag. Walter Zaldívar Álvarez

DECANO DE LA FACULTAD DE CIENCIAS

Dr. Walter Estrada López

Carátula: *El profesor Valqui explicando - a su manera - nuevas ideas de la Cuántica durante su conferencia en el II Meeting of Physics en el año 2000. En el ángulo superior izquierdo hay una micrografía realizada con la punta de un AFM raspando una película de polímero blando.*

Revista de la Facultad de Ciencias de la UNI – REVCIUNI

Volumen 15, Número 1, enero-diciembre 2012
ISSN: 1813-3894

Publicada por el Instituto de Investigación de la Facultad de Ciencias
de la Universidad Nacional de Ingeniería

Director del Instituto de Investigación:
Dr. Carlos Javier Solano Salinas

Comité Editorial:

- Dr. Orlando Pereyra (Universidad Nacional de Ingeniería, UNI - Perú)
- Dr. Ener Salinas (ABB - Suecia, LSBU - UK)
- Dr. Harold Blas (Universidade Federal de Mato Grosso, UFMT - Brasil)

Comité Científico:

- Dr. Javier Solano (Universidad Nacional de Ingeniería, UNI - Perú)
- Dr. Héctor Loro (Universidad Nacional de Ingeniería, UNI - Perú)
- Dr. Armando Bernui (Universidade Federal de Itajubá, UNIFEI - Brasil)

La revista se distribuye en la Facultad de Ciencias – UNI
Av. Túpac Amaru 210 - Rimac
Lima - Perú

Página web: <http://fc.uni.edu.pe/revciuni>

E-mail: postgradofc@uni.edu.pe

Frecuencia de publicación: Anual

Registro de Depósito Legal N° 15010199-3597
Revista indexada en el LATINDEX

Impreso en los Talleres Gráficos de
CONTACTO GRÁFICO PERÚ S.R.L.
482 5252 / 381 3992 Fax: 4827 055
E-mail: contacto.graficoperu@gmail.com

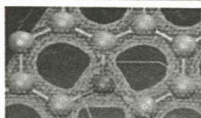
Poster of the event



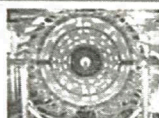
UNIVERSIDAD NACIONAL DE INGENIERÍA, FACULTY OF SCIENCES

XV MEETING OF PHYSICS

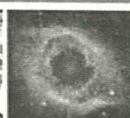
August 1st-3th, 2011



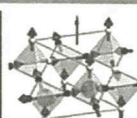
Graphene and nanotubes



Searching the Higgs



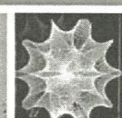
Helix Nebula



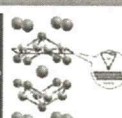
Monopoles in "spin ice"



Nano propellers



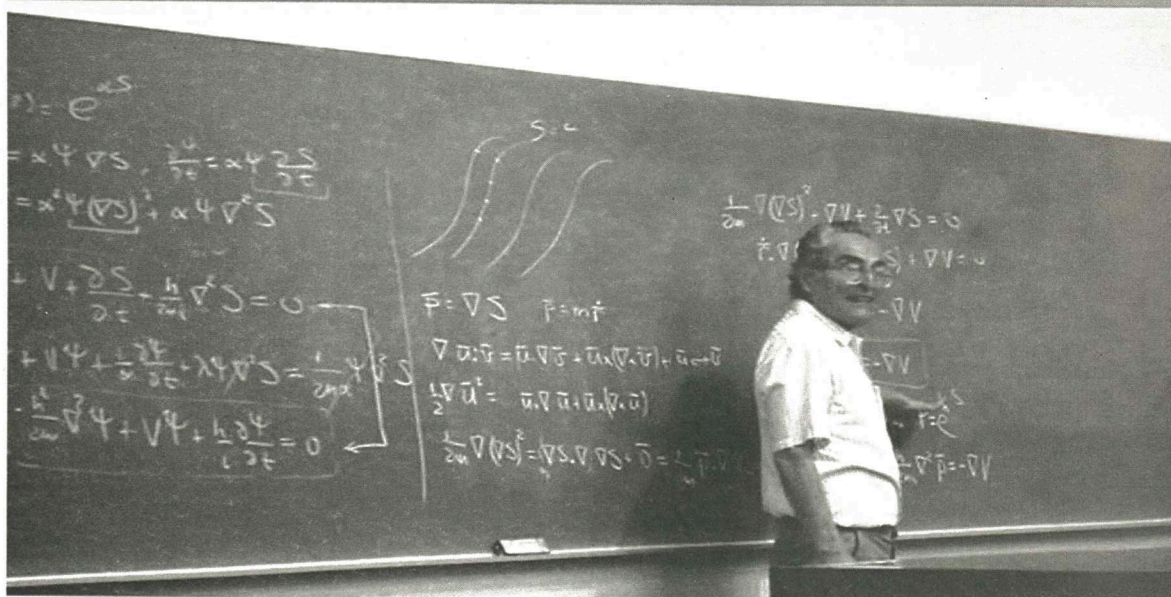
Calabi-Yau Manifold



Hidden Order



Nano Casimir effect



Organizing Committee

Orlando Pereyra (*)
Harold Blas (**)
Ener Salinas (**)
Arminda Tirado
Javier Solano
Rosendo Ochoa
Héctor Loro

(*) Local Coordinator

(**) International Coordinators

Speakers

Holger Valqui (UNI, Peru)
Carmen Gonzales (Alcatel, France)
Andrés la Rosa (PSU, USA)
Fernando Ponce (ASU, USA)
C.A. Tello (UNESP, Brazil)
Harold Blas (UFMT, Brazil)
Ener Salinas (ABB-Sweden, LSBU-UK)
José Antonio Alarco (QUT, Australia)
Antonio Pereyra (ON, Brazil)
Manfred Horn (UNI, Peru)
Karim Salazar Salinas (CTIC-UNI, Peru)
Teófilo Vargas (UNMSM, Peru)
Walter Estrada (UNI, Peru)
Agustín Zúñiga (IPEN, Peru)
Modesto Montoya (ECI, Peru)
J. Anibal Paulyac (FIEE-UNI, Peru)
Carlos Landauero (UNMSM, Peru)
Benjamin Marticorena (UNI, Peru)
Jaime Ávalos (OEI, Peru)
Victor Florian Sales (SCITECH, Peru)

Research works in the areas of Theoretical, Experimental, Educational and Applied Physics will be presented by international and local scientists as well as students. The main purpose of this event is to continue or initiate new collaborations between these groups of scientists.

Registration: it is free.

Place: Auditorium of the Faculty of Electric and Electronic Engineering and Auditorium of the Faculty of Sciences of UNI.

Information: Faculty of Sciences, UNI, Av. Túpac Amaru 210, Rimac. 5th door of UNI.

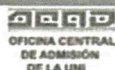
Reception of contributions: opereyra@uni.edu.pe

Deadline for contributions: July 15th, 2011.

Web page:

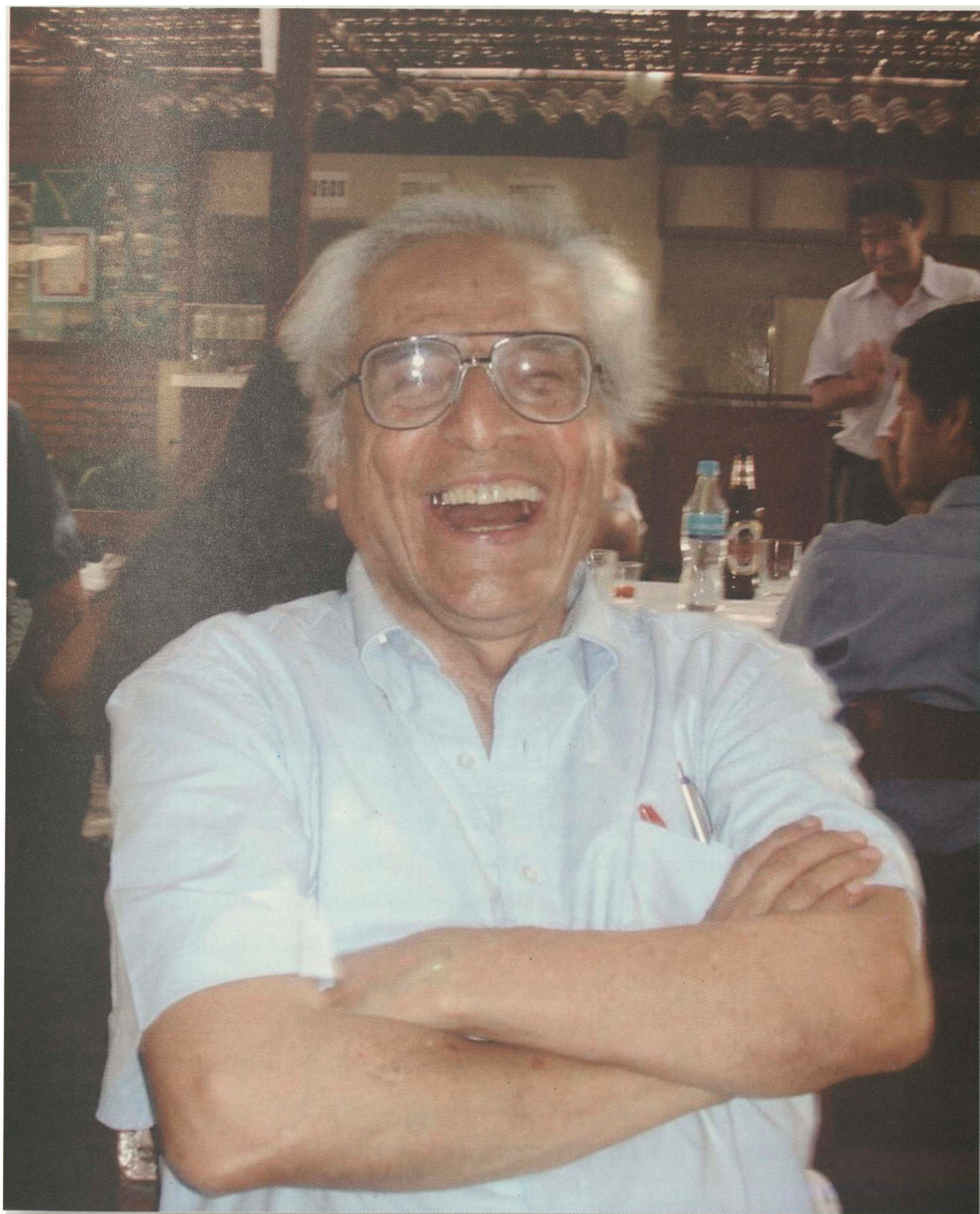
<http://fc.uni.edu.pe/portal/pwencuentro/encuentro.html>

Sponsors:



CEPRE-UNI





ORGANIZING COMMITTEE AND SPEAKERS

ORGANIZING COMMITTEE

Orlando Pereyra	<i>Universidad Nacional de Ingeniería, UNI, Peru (Local Coordinator)</i>
Harold Blas	<i>Universidade Federal de Mato Grosso, UFMT, Brazil (International Coordinator)</i>
Ener Salinas	<i>ABB - Sweden, LSBU - UK (International Coordinator)</i>
Arminda Tirado	<i>Universidad Nacional Federico Villarreal, Peru (Local Coordinator)</i>
Javier Solano	<i>Universidad Nacional de Ingeniería, UNI, Peru (Local Coordinator)</i>
Rosendo Ochoa	<i>Universidad Nacional de Ingeniería, UNI, Peru (Local Coordinator)</i>
Héctor Loro	<i>Universidad Nacional de Ingeniería, UNI, Peru (Local Coordinator)</i>

SPEAKERS

Holger Valqui	<i>UNI – Peru</i>
Carmen Gonzales	<i>Alcatel – France</i>
Andrés La Rosa	<i>PSU – USA</i>
Fernando Ponce	<i>ASU – USA</i>
C.A. Tello	<i>UNESP – Brazil</i>
Harold Blas	<i>UFMT – Brazil</i>
Ener Salinas	<i>ABB – Sweden, LSBU – UK</i>
José Antonio Alarco	<i>QUT – Australia</i>
Antonio Pereyra	<i>ON – Brazil</i>
Manfred Horn	<i>UNI – Peru</i>
Karim Salazar Salinas	<i>CTIC-UNI – Peru</i>
Teófilo Vargas	<i>UNMSM – Peru</i>
Walter Estrada	<i>UNI – Peru</i>
Agustín Zúñiga	<i>IPEN – Peru</i>
Modesto Montoya	<i>ECI – Peru</i>
J. Aníbal Pauyac	<i>FIEE-UNI – Peru</i>
Carlos Landauro	<i>UNMSM – Peru</i>
Benjamín Marticorena	<i>UNI – Peru</i>
Jaime Ávalos	<i>OEI – Peru</i>
Victor Florian Salas	<i>SCITECH – Peru</i>

FOREWORD BY THE EDITORS

Since the first UNI meeting in January 1999, the Meetings of Physics have been organized annually. Their main purpose is to foster the exchange of ideas and experiences of Peruvian physicists working in diverse areas such as condensed matter, particle physics, applied physics, gravitation and mathematical physics. Their history provides a view on the development of a rich collaboration between the Peruvian physicist community working abroad and its local counterpart.

The 2011 edition of the meeting, the 15th of its kind, celebrated Prof. H. Valqui's 80th birthday and took place in August 2011 (1st - 3rd) at the National University of Engineering (UNI) in Lima, Peru. Prof. Valqui has served as an inspiration to various generations of Peruvian physicists for his teaching dedication, enthusiasm and ethical attitudes towards diverse issues in the Faculty of Sciences, FC-UNI.

Prof. Valqui has played an important role in the education and as a mentor of several Peruvian physicists in the last decades, as illustrated by this festschrift containing articles of Prof. Valqui's former students, colleagues and friends. Among his former students many are PhD's in Physics working in several institutions in Peru and abroad, some followed the same research areas initiated in their undergraduate period and all have been influenced by his vision and unique attitude towards science. He introduced new lines of research in the Peruvian mathematical physics community such as soliton physics and nonlinear science, in the 80's. Moreover, he has also been interested in the transition formulations from classical to quantum; and recently he has actively been investigating the EPR paradox and virgin photons. His rigorous treatment of physics and mathematics disciplines, in combination with his enthusiasm, has made him a source of inspiration to all his students.

TABLE OF CONTENTS

Virgin photons do not exist <i>Holger G. Valqui</i>	1-6
Direct profiling of polarization fields in nitride semiconductors at nanometric scale using electron holography in the transmission electron microscope <i>F. A. Ponce</i>	7-15
Semiconductor Lasers <i>Carmen Gonzales</i>	16-17
Proton-fountain Electric-field-assisted Nanolithography (PEN) <i>A. La Rosa and M. Yan</i>	18-44
Perspectives in Industrial Nano-materials <i>Jose A. Alarco</i>	45-56
Transmission of electric energy without using wires <i>Ener Salinas</i>	57-65
Fission-track dating applied to Peruvian volcanic glasses <i>C. A. Tello et al.</i>	66-73
Reactivity, energetics and molecular structure: A theoretical and experimental approach <i>Juan Z. Dávalos Prado</i>	74-81
Proton-fountain Electric-field-assisted Nanolithography (PEN): Fabrication of polymer nanostructures that respond to chemical and electrical stimuli. <i>Andres La Rosa, Mingdi Yan, Rodolfo Fernandez, Xiaohua Wang and Elia Zegarra</i>	82-123
Two dimensional BF gravity: A Hamilton-Jacobi analysis <i>M.C.Bertin, B.M.Pimentel and C.E.Valcarcel</i>	124-132
Higgs-Strahlung process in the $SU(3)_L \otimes U_M(1)$ Model with Heavy Leptons <i>D. Romero and O. Pereyra</i>	133-143
Non commutative (generalized) sine-Gordon / massive Thirring correspondence, integrability and solitons <i>H. Blas and H.L. Carrion</i>	144-181
The variable mass Thirring / sine Gordon duality and continuous topological configurations <i>H. Blas and J.M. Jaramillo</i>	182-195
Breaking the Λ CDM degeneracy using CMB data <i>A. Bernui</i>	196-203
Generación de soluciones exactas de las Ecuaciones de Einstein con ayuda de euclídones <i>Páuyac Huamán, J. Anibal</i>	204-208
OA-UNI: an Astronomical Observatory at UNI <i>A. Pereyra, W. Cori, E. Meza, J. Ricra, and G. Granda</i>	209-211
Generalized AKNS System, Non-vanishing Boundary Conditions and N-Dark-Dark Solitons <i>A.de O. Assuncao, H. Blas and M. J. B. F. da Silva</i>	212-241
AKNS solitons with constant boundary conditions via dressing transformation and tau function approach <i>H.Blas and M.Zambrano</i>	242-250
Formación de Comunidades Latinoamericanas de Investigación, La experiencia de CLARA y las posibilidades de la RAAP <i>Benjamín Marticorena</i>	251-256
Pictures of professor Valqui friends and colleagues	257-274

VIRGIN PHOTONS DO NOT EXIST (*)

HOLGER G. VALQUI

Faculty of Science, Universidad Nacional de Ingeniería

Scientific models for microscopic objects (not directly observable) should satisfy three basic conditions:

- i) Be logical and mathematically consistent,
- ii) Be constructed from certain experimental "evidence", which implies that actual measurements have been done to the properties of some of the involved objects,
- iii) Explain the corresponding macroscopic phenomena, and predict new macroscopic effects.

None of these conditions is trivial, as it is not possible to construct a structural model without resorting to various assumptions, some of them quite 'obvious', some accepted naively- or recklessly. On the other hand, the result of any measurement must inevitably be interpreted and that interpretation depends partly on the model that we want to justify. This seems to be the case of photon entanglement, a phenomenon that is explained and justified under the assumption that the photons acquire their polarization just after they have passed through a polarizer. This article aims to show that the existence of photons without polarization (virgin photons) apart from being inconsistent, contradicts a number of experimental facts.

Keywords: Scientific models, measurement, polarization, entanglement.

1 Introduction

By observing the manifestations of a physical object or a set of objects, scientists build a model of such objects and their interactions, so that the model can adequately represent such objects and their corresponding interactions.

All objects and all interactions?

It is necessary to emphasize that any model is limited to only represent the most significant objects (for example, to describe the interaction between two planets one usually ignores the presence of the remaining planets, and any asteroids) and also their more significant interactions (for example, when using an electron gun one often ignores the interactions between the involved electrons)

What are the objects and interactions that are significant?

This is a problem for which the solution heavily depends on the views and interests of the scientists working with the model under consideration.

Are the measurements made by scientists not objective?

One of the requirements for a measurement to qualify as scientific is replicability. For example, measurements of paranormal phenomena do not meet this requirement, hence its alleged results are not considered scientific results. In this regard, science has found that a single measurement of some phenomenon can only be considered as a precarious reference, while not confirmed or ruled out by other "independent" measures. This leads us to distinguish between what we call singular measurement (done once) and a measurement process. To get the "objective value" of the measurement of some property of a system,

we consider two types of measurement processes which are (ergodically) equivalent:

- i) Repetition, N times, of the same singular measurement, each time in 'identical' conditions, or
- ii) The realization of a singular measurement on N "identical" sub-systems that do not interact.

The results of the corresponding singular measurements are probabilistic in nature, but are not arbitrary. [Ignorance of the features that have probabilistic outcomes of a measurement process helped to clarify some fraud committed by scientists]. Of course, the measurements are made under certain boundary conditions and eventually other special conditions, the same which should be specified to allow replicability. The probabilistic nature of the measurements (i.e. measurement process), that always has been accepted by experimentalists, is one of the most striking differences between the classical (theoretical) model and the quantum model of Physics. [In the case of microscopic interactions between objects should be kept in mind that the measurements must be interpreted according to the theoretical model that has been adopted. This carries the danger of vicious circles]

First the physician T. Young (1773-1829), in order to explain the experimental phenomenon of double refraction of light passing through a rhombohedron of Iceland spar (calcite, CaCO_3), then the mathematician A. Fresnel (1788-1827), established the transversal character of light waves. Then Maxwell (1831-1879) established that light was an electromagnetic waveform, and in 1905 Einstein (to explain the photoelectric effect) proposed a model in which light is composed of microscopic "particles" with well-defined energy, the same which later be-

came known as photons. It must be accepted that photons are not usual particles, but special particles that, among other properties, are diffracted when they pass close to the edge of a surface (which apparently does not happen with “normal” particles). An experiment to verify that the particle model does not contradict the photon wave character in its most characteristic manifestation, was conducted by G.I Taylor in 1909 who, with a very weak light bulb, lit up a needle for 2 000 hours (≈ 3 months) which allowed the photons arrive virtually isolated. The image formed on a screen coincided with the diffraction image produced, in a very short time interval by an intense light spot.

2 Polarizers

For electromagnetic waves, particularly light in the visible range, one knows two kinds of light: “natural light”, $n\ell$, and polarized light, $p\ell$.

$n\ell$ is the light emitted by the sun or any incandescent material, while $p\ell$ is the light that has certain transverse characteristics, which become evident when it is directed at and passes through certain objects called polarizers, which can be of different kinds: linear polarizer, LP, right-handed circular polarizer, RHCP, left-handed circular polarizer, LHCP, birefringent crystals, including a rhombohedral calcite prism, *RoCa*.

The LP, RHCP, LHCP, have the virtue of letting pass half the natural light that strikes them, and absorb the other half. Additionally, the light that has passed a polarizer acquires certain properties, characteristic of such polarizer, and is called polarized light: $L\ell$, $RHC\ell$, $LHC\ell$, respectively.

A linear polarizer has a characteristic orientation, so sometimes it is convenient write LP (α) to indicate such orientation. A *RoCa* has the virtue of splitting the incident beam into two beams orthogonally polarized, and has what is known as Principal Plane [parallel to both the incident ray and the so called optical axis], characterized by an angle ψ .

Moreover, the light has energy, the transport of which is measured as intensity, J (energy per unit time). By measuring the intensity of light directed (perpendicularly) to a polarizer, and the intensity of light that gets through the polarizer, one can know what fraction of incident light is absorbed by the polarizer, and what fraction passes through said polarizer, acquiring the polarization induced by such a polarizer, without changing its characteristic frequencies. On the other hand the fraction of reflected light is small and often ignored, which can be significant if the beam is extremely weak.

3 Experiments with polarizers.

Experimentally we know the following results:

• Experiment 1

If natural light, of intensity $J(n\ell)$, is incident on a XP polarizer, (XP= LP, RHCP or LHCP), then the intensity of the polarized light that emerges from the polarizer is half the incident intensity:

$$J(Xp\ell) = \frac{1}{2} J(n\ell)$$

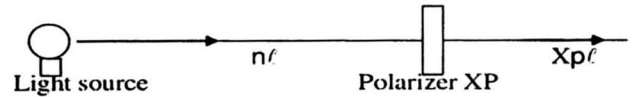


Figure 1.

• Experiment 2

If the light emerging from a polarizer XP, with intensity $J(Xp\ell)$, strikes a second polarizer $X'P$, “identical”, to XP, then the whole incident beam passes through the second polarizer, i.e.

$$J(X'p\ell) = J(Xp\ell),$$

where XP = LP, RHCP or LHCP.

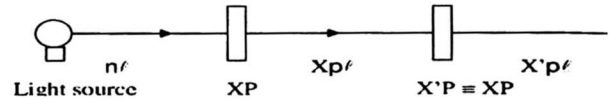


Figure 2.

• Experiment 3

Malus Law: If the light emerging from a linear polarizer LP(α) with intensity $J(p\ell(\alpha))$ is incident on a second polarized LP(β), then the intensity $J(p\ell(\beta))$ of the light emerging from the second polarizer meets :

$$J(p\ell(\beta)) = J(p\ell(\alpha)) \cos^2(\alpha - \beta)$$

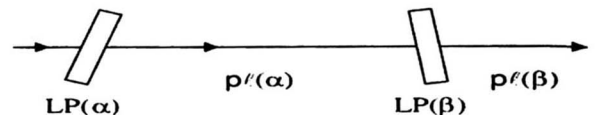


Figure 3.

• Experiment 4

If the light emerging from a linear polarizer LP is incident on a circular polarizer, CP, right-handed or left-handed, then the intensity of light emerging from the second polarizer is reduced by half,

$$J(Cp\ell) = \frac{1}{2}J(Lp\ell)$$

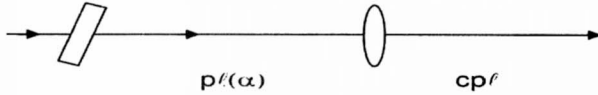


Figure 4.

• Experiment 5

If the light emerging from a circular polarizer, CP, right-handed or left-handed, hits a linear polarizer, LP, then the intensity emerging from the second polarizer is halved,

$$J(Lp\ell) = \frac{1}{2}J(Cp\ell)$$

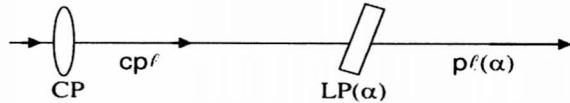


Figure 5.

• Experiment 6

If the light emerging from a circular polarizer, right-handed or left-handed, hits a left-handed or right-handed circular polarizer, respectively, then the second polarizer absorbs all light that strikes it.

$$J(Lhcpl) = 0$$

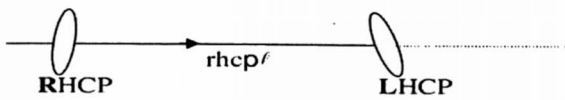


Figure 6.

• Experiment 7

If a ray of natural light falls on a rhombohedral calcite prism, then from the other end of the prism emerge two beams of linearly polarized light, with orthogonal polarizations, and each one with intensity equal to half of the incident intensity.

$$J(p\ell(\alpha)) = J(p\ell(\alpha + \frac{\pi}{2}))$$

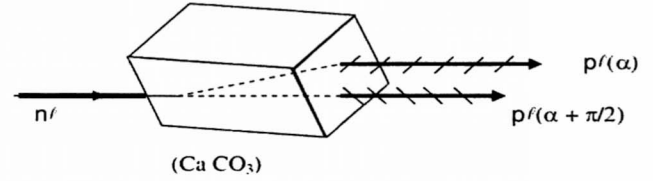


Figure 7.

• Experiment 8

If the polarized light emerging from a circular polarizer, CP, right-handed or left-handed, hits a rhombohedral calcite prism then from the other end of the prism emerge two linearly polarized beams with orthogonal polarizations, each with intensity equal to half the incident intensity.

$$J(p\ell(\alpha)) = J(p\ell(\alpha + \frac{\pi}{2}))$$

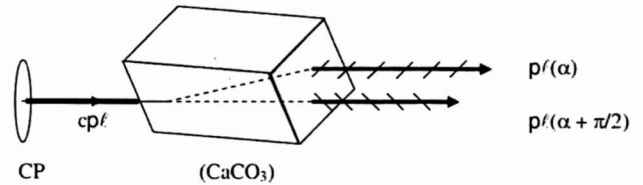


Figure 8.

• Experiment 9

If the light emerging from a linear polarizer, LP, falls on a rhombohedral calcite prism then from the other end emerge two linearly polarized beams with orthogonal polarizations, and the sum of the intensities of the emerging beams is equal to the incident intensity.

$$J(p\ell(\beta)) + J(p\ell(\beta + \frac{\pi}{2})) = J(p\ell LP)$$

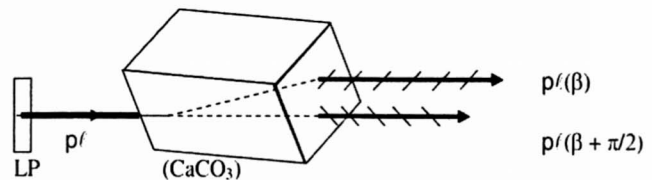


Figure 9.

• Experiment 10

If a linearly polarized beam, $p\ell(\alpha)$, is incident on a rhombohedral prism whose main plane is characterized by the angle ψ , then the two emerging rays will have the intensities J_1 and J_2 such that

$$J_1 = J(p\ell(\alpha)) \times \sin^2(\psi - \alpha)$$

$$J_2 = J(p\ell(\alpha)) \times \cos^2(\psi - \alpha)$$

[Note that (9) is a special case of case (10), and that for $\alpha = 45$ the two emerging rays have the same intensity, equal to half the incident intensity.]

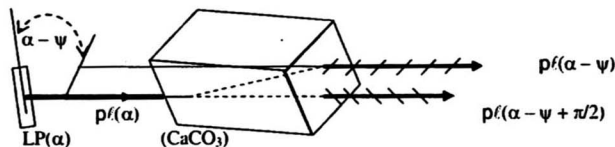


Figure 10.

We know that there exists “natural” polarized light, for example, by scattering in the atmosphere, by reflecting sunlight on the sea surface and other surfaces, by incidence of sunlight into pieces of natural crystals. Moreover, since the late nineteenth century it is known the so called Zeeman effect, whereby the light emitted by atoms placed in a magnetic field changes its polarization state. In addition, certain animals, as bees, are sensitive to polarized light and use that information for guidance, it is also believed that this is a resource of some migratory birds. Additional arguments are the rules of selection in atoms (where the spins of the photons should complement the angular momentum of the considered system) and the generation of laser light (stimulated photon decay)

If, as proposed by Einstein, light is made by special particles, called photons, all what was mentioned in the macroscopic experiments (section 3) can be explained under the assumption that photons have polarization states that can change when they interact with some objects. From Experiment 2 can be inferred that each of the photons of the light beam, once they have passed through a polarizer XP, they can go through a similar and equally oriented polarizer, without being absorbed. That is to say that experimentally a photon possesses certain polarization (eventually the polarization induced by a given polarizer). Moreover, bearing in mind Experiment 3 we can assume the existence of a quantum Malus law, under which the photons pass through the second linear polarizer with probability $\cos^2(\beta - \alpha)$. This assumption also allows us to consider that natural light photons have randomly distributed polarization directions, which, taking into account the large number of photons involved, explains the above mentioned classic effects. That is, **if we assume the validity of the quantum Malus law, then we can explain all the experimental results presented above in section 3.** In particular, this quantum law helps explain why half of the natural light photons (which have random directions of polarization) are absorbed by the linear polarizer and half of them passes (acquiring the polarization direction of the polarizer). Then, remembering that circularly polarized light is the normalized sum of two orthogonal linear polarizations, we also expect that half of the photons of natural light will be absorbed by a

circular polarizer and half will pass through the polarizer.

4 The case of few photons

To measure the polarization of a beam of light, we compared the energies of the incident beam and of the beam emerging from a polarizer. Such a thing no longer seems possible in the case of an isolated photon, or a few single photons (although equally polarized). That is a key problem in the analysis of experimental results in cases of single photons. Here we must clarify that in the case of the Taylor experiment, photons arrived at the screen properly isolated, one by one, but it was a case of a large number of single photons. On the other hand, we can consider some situations where there is only a single photon or there are a few photons present, which are not part of a larger whole that is also involved in the process. By blocking a stream of photons-Taylor-type, and unblocking it by a very short time interval, one could strike a linear polarizer (to determine the polarization of a half dozen photons) with very few photons, some of which will be absorbed by the polarizer, and the others will go through, to be counted in the detector. Here, moreover, we must remember that the number of photons that pass a polarizer is known only on average.

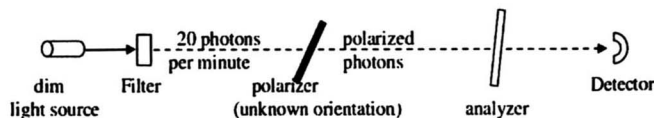


Figure 11.

That is, the photon model works satisfactorily in the case of a large number (hundreds of billions) of photons, so that the polarizers can be considered ideal, or the inaccuracies of the polarizers can be considered negligible:

- i) Ideal polarizers do not reflect a fraction of the photons striking them [what real polarizers do]
- ii) Properly oriented ideal linear polarizers allow to pass 100% of the incident photons, and those perpendicular to the above direction absorb all incident photons [real polarizers cannot be oriented exactly, because of practical limitations and , additionally, the uncertainty effect]
- iii) Ideal polarizers have a well defined polarization direction [real polarizers, as in the case of Polaroid films the degree of its polarization depends on the quality of stretching undergone by the molecules in the chains that constitute the material], and so on.

In the case of working with few photons, besides not being possible to measure the intensity (or energy) of the incident or emerged photons, it is no longer reasonable to suppose that the above mentioned inaccuracies are negligible. This assumption solves some problems with measurements of the properties of microscopic systems.

but also creates other difficulties. For example: Assuming that there are indeed virgin photons, i.e. photons that still do not have polarization:

- i) The property of being virgin photons should manifest itself also in the case of large numbers of photons, i.e., in the case of light rays. Are there virgin light rays? ,
- ii) In the case of light rays it is known that there exist 'naturally' polarized rays (i.e., its polarization has not been caused by an intentional act of measurement), for example, the rays scattered by the molecules of water atmosphere,
- iii) What would be a property that would distinguish virgin photons from polarized ones?. For example, an employee of G. I. Taylor tells us that in the next minute through for this window should appear some 5 or 6 photons, and you should determine if they are virgin photons or polarized photons.

[Note that the case would be different if the aforementioned partner tell us that this window will appear 5 or 6 photons per minute, and you have all the time necessary to make measurements]

We note that trying to measure the polarization of a single isolated photon is in itself meaningless, since a physical measurement requires that it be repeatable for the reason that any valid physical measurement is the result of a measurement process. Suppose (very fancy) there is actually a single photon (previously polarized but in a direction unknown to the operator) incident on a linear polarizer and then on a photo-multiplier placed after the polarizer, which:

- i) indicates the presence of a photon,
- ii) remains without sign of having received a photon

What can the operator conclude in each case?

- i*) He can only say that the photon passed the polarizer, i.e. its direction of polarization was not perpendicular to the direction of the polarizer.
- ii*) He can only say (assuming the photo-multiplier is 100% efficient) that the incident photon polarization was not parallel to that of the polarizer.

5 Polarized photons do exist

Note also that there is no doubt about the existence of polarized photons, as certified by the Experiment 2, in which all the photons in a beam of light, emerging from a first polarizer, passed a second polarizer "identical" to the first. There is also no doubt that if photons with polarization direction α strikes a second polarizer with direction β , then some of the incident photons will pass, while others will be absorbed, as certified by what we have called the Malus Law. That is, there is no doubt about the existence of polarized photons, what is at stake

is the existence of virgin photons.

Suppose the following situation: Imitating GI Taylor we have produced a very weak light source, so that through a window pass (according to our calculations) some 20 photons per second. Then, just a little after the window we put a discreet linear polarizer, so we can say that, on average, from the polarizer will emerge 10 photons per second, which will strike a photo-multiplier that counts them. With this arrangement we ask some experienced physicist, call him AA, to prepare the equipment to determine the polarization of the photons that arrive at the photo-multiplier in the time interval between 9:00 h and 9:00 h + 1 second. From what is currently known, the physicist AA can not measure, in the stipulated time interval, the polarization of (approximately) some 10 photons (which emerge from the discreet polarizer). Though physicist AA says that in such a short time interval (which only allows him to count a dozen photons) he can not measure the polarization of the mentioned photons, such impossibility does not mean that these photons have no polarization. In conclusion we can state the fact that the impossibility to measure a certain property of an object (or set of objects) does not necessarily imply that that object does not possess such a property. On the other hand, if the experimenter AA were allowed to prolong his experiment with a large number of photons (say, for a few hours), then he could (orienting his polarizer-analyzer to different angles and counting the number of photons passing in each orientation) determine if the photons incident on his analyzer were randomly polarized, or if they were polarized in a specific direction.

In the case of experiment 3, Malus Law, we see that a large set of photons polarized in the direction α , approximately the fraction $\cos^2(\alpha - \beta)$ of them changed their polarization to the direction β . That is, the polarizer did not create the polarization of individual photons, instead they change the direction of polarization of the passing photons. [This interpretation is debatable, but without major effects for the present discussion]

Additionally, the supposed virgin photons which composed natural light, when they go through a polarizer behave as if their polarizations were randomly oriented, which is consistent with the generation of photons by atoms that are not necessarily correlated in their excitation nor in their emission instants.

Finally, consider the following set of experiments applied to a beam of intensity J .

- i) We make such a ray impact on a polarizer of orientation α , finding that the emerging beam has intensity $\frac{J}{2}$. We changed the direction of the polarizer and the emerging intensity does not change. Apparently it is a ray of sunlight or a circularly polarized beam.
- ii) We direct the beam to a circular polarizer, right-handed or left-handed, and we find that in both

cases the emerging intensity is $\frac{J}{2}$. Apparently it is natural light.

- iii) Now we direct the same ray to a rhombohedral calcite prism and get two emerging rays of equal intensity. With this result we are almost convinced that the ray in question is a ray of natural light.

But if we had done the experiment superimposing two beams of equal intensity, linearly polarized with orthogonal polarizations (but with no coherent phases) we can verify that the results of the intensities, arising from the polarizers or rhombohedral prism, would be the same as those obtained in each of the cases mentioned above. The same result would be obtained if it were no longer the superposition of two incoherent beams orthogonal to each other, but the overlap of N-incoherent-rays, with equal intensity, with polarizations oriented parallel to the diagonals of a regular polygon of N sides. The results of the mentioned experiments basically justifies the assumption that natural light photons are polarized with random polarization directions, statistically symmetrical (as long as there is no reason to suppose a preference for certain orientations).

6 Conclusions

There exist photons with definite polarization state (equivalent to a polarization direction). This can happen naturally (by effect of a natural polarizer), as mentioned

in (05), or intentionally, using a polarizer.

As shown in (10), in the case of a single photon or few isolated single photons, but certainly polarized (in direction unknown to the experimenter), it is impossible to determine the polarization of such photons, which does not imply that these photons do not have a polarization state before falling in the analyzer (a polarizer or a rhombohedral calcite prism).

Polarizers absorb photons or they change their polarization state, but they do not "generate" the state of polarization (in the sense that photons acquire a polarization state just after they pass the polarizer). The octahedron of calcite, and other crystals also change (in two directions orthogonal to each other) the polarization states of the incident photons.

The postulate of the existence of virgin photons seems to have more the character of a metaphysical postulate than the character of a postulate of physics itself.

As challenges to those who maintain the existence of virgin photons:

- i) Build a direct experiment showing the ability to differentiate virgin photons from photons with polarization,
- ii) Explain why half of the supposed virgin photons of natural light pass through the linear or circular polarizers.

-
1. Bergman Schaeffer III-Lehrbuch der Experimentalphysik, Wellenoptik, W.Gruyter 1962
 2. J. Simons, States- Waves and Photons,
 3. M.Born- Optik, Springer, 1966
 4. R. Feynman-Lectures on Physics I, Addison Wesley
 5. E. Hecht- Optics, Addison-Wesley, 2002
 6. D. Perovich- Observation of polarization of light reflected from sea ice, Journal of Geophysical Research Oceans, 2009
 7. R. Turner, C.M.Ankenbrandt, R.C. Larsen- Polarization of cosmic-ray Muons, Phys. Rev, D4, 17 (1971)
 8. K. Able- Sky light polarization patterns at dusk influence migratory orientation in birds, Natur, 299, 1982
 9. F. Moore, J.B. Phillips-Sunset sky light polarization and the migratory orientation of yellow-rumped warblers, Animal Behaviour, Vol 26, 6, 1988.

Direct profiling of polarization fields in nitride semiconductors at nanometric scale using electron holography in the transmission electron microscope

F A Ponce

Department of Physics, Arizona State University, Tempe, AZ 85287-1504, USA

E-mail: ponce@asu.edu

Abstract: The nitride semiconductors are widely used in high efficiency light emitting devices and are currently being considered for photovoltaic applications. The reduced symmetry in the wurtzite structure compared to cubic semiconductors results in the growth of large densities of crystalline defects and in the presence of strong spontaneous and piezoelectric polarization effects. A correlation between the microstructure and the polarization fields can be achieved with electron holography in the transmission electron microscope. Electron holograms thus obtained can provide energy band profiles with sub-nanometer spatial resolution. The phase of the electron beam is sensitive to the electrostatic potential, and a direct measurement of the latter can be achieved by making the electron beam signal that traverses the specimen interfere with a reference electron beam that travels through vacuum. This technique has been quite useful in probing the fields and charges at dislocations and at interfaces in semiconductors, and it is particularly useful to determine the piezoelectric effects in group III nitride semiconductor heterostructures. A review of applications to InGaN and AlGaN based heterostructures is presented in this paper.

Key words: Electron energy band profiles, piezoelectric fields, nitride semiconductors, electron holography

1. Introduction

In the past two decades, the nitride semiconductors have acquired pronounced relevance due to their robustness and their range of band gaps¹. The bond between nitrogen and the group III elements is strong; the cohesive energies of AlN, GaN, and InN are 11.67, 9.06, and 7.97 eV, respectively.² The corresponding band gap energies of 6.28, 3.41, and 0.7 eV span the electromagnetic spectrum from the infrared into the ultraviolet, including the full visible spectrum.³ The growth of these materials is challenging due to the high temperatures required for synthesis and the absence of suitable substrates. High densities of dislocations are necessary in order to accommodate the lattice mismatch and the thermal expansion mismatch within the AlGaInN alloy system.⁴ The lattice displacement is associated with those dislocations (Burgers vectors) are large compared with the interatomic bond length.⁵ Heterojunction lattice mismatch often results in pseudomorphic thin film structures with large strain fields.⁶⁻⁹ The absence of a center of symmetry in the hexagonal wurtzite structure results in significant spontaneous polarization. In addition, piezoelectric polarization results from the strain due to dislocations and strained heterointerfaces. The electrostatic fields due to spontaneous and piezoelectric polarization are of the order of several MeV/cm,⁷ and are typically highly localized at the nanometer scale, and have a strong effect on carrier transport and carrier recombination.

Profiles with nanometer resolution of the electrostatic potential and charge distribution in solids can be obtained by electron holography in the TEM. This technique has been used to visualize the potential energy profile across p-n junctions in silicon devices used in microelectronics.¹⁰ Electron holography has been particularly useful to measure the potential and charge density variations in the

nitride semiconductors. Electron holography was first used by Cherns in 1999 to measure the piezoelectric field in InGaN quantum wells.¹¹ In this paper, a brief introduction is first given to the electron holography technique. This is followed by examples of applications to determine the nature of electrostatic potential variations in the vicinity of dislocations in GaN, at InGaN/GaN quantum wells, and at two-dimensional electron and hole gas heterostructures.

2. Electron holography in the transmission electron microscope

The use of electron holography has increased significantly in the last decade, particularly in the direct measurement of the electrostatic fields and charges in nitride semiconductor materials. In conventional TEM techniques, like those used for diffraction contrast analysis, only the intensity of the electron beam is recorded, while the phase information is lost. Electron holography is an interferometric technique which provides information about spatial distribution of the phase shift, $\Delta\theta$. In order to obtain the phase information, a highly coherent electron beam is obtained by using a field emission gun. A sample is prepared to electron transparency, but not too thin to avoid surface related effects such as strain relaxation and screening due to the surface.

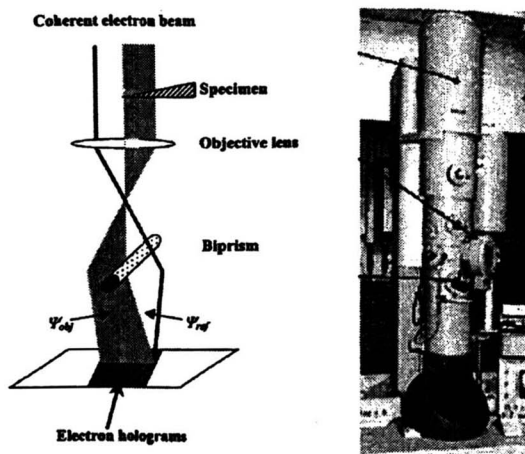


Figure 1. Schematic diagram of the electron beam path for electron holography in a transmission electron microscope.

The phase of the electron beam is modulated by spatial variations of the electrostatic potential of the specimen. Figure 1 shows the configuration used for electron holography. The sample is introduced into the electron beam path covering about half of the beam diameter. The other half of the beam travels through vacuum and it is used as a reference beam. The electrons that traverse the sample experience a change in velocity, in part due to the variation of the reference potential, and gain in phase compared to the electrons traveling through vacuum. The phase change is directly proportional to the electrostatic potential as we will see later. The phase information can be retrieved from the wave interference pattern.^{12,13}

Electron holography is an electron-interference technique that makes the recovery of the phase information possible. A conducting filament that acts as an optical biprism is used to produce an interference pattern. The biprism is aligned manually so that its image matches the boundary of the specimen at the image plane. When a positive bias is applied to the biprism, the beams traveling through vacuum (reference wave) and the sample (object wave) are made to overlap at the image plane, and their overlap is recorded. The geometric configuration around the biprism is depicted in Fig. 2(a). The biprism filament is along the y -direction, and the beam propagates along the z -direction. The biprism is at a given potential, while the two plates are grounded. In some microscopes these two plates are the edges of the selected area aperture.

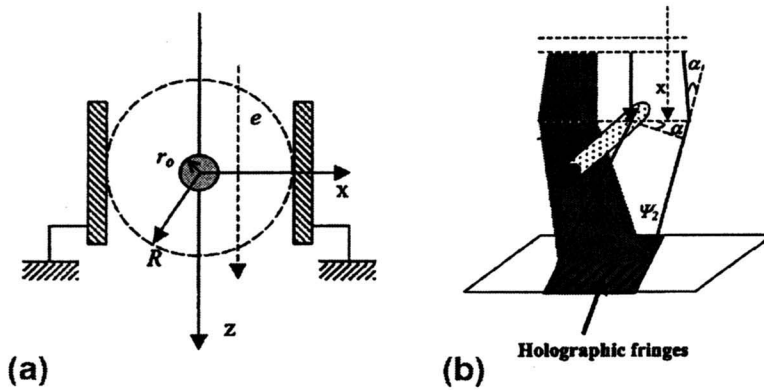


Figure 2. Holographic image formation. (a) configuration around the biprism: (b) Interference between two plane waves.¹³

As already indicated, the specimen under investigation is positioned to cover only half of the object plane, in that way one beam passes through the specimen (referred as the object wave $\psi_{obj} = A_o e^{i\theta_o}$) and the other beam passes through the vacuum (referred to as the vacuum wave $\psi_{vac} = A_v e^{i\theta_v}$). A bi-prism, which resides in the plane of the selected area aperture, deflects the two electron beams in order to cause their interference. Applying a positive bias to the filament of the bi-prism, the object wave and vacuum wave converge and yield an interference pattern in the overlapping region. The intensity distribution can be represented by:¹³

$$I = |A_v e^{i(\theta_v - \Delta)} + A_o e^{i(\theta_o + \Delta)}|^2$$

$$= A_v^2 + A_o^2 + 2A_v A_o \cos(\theta_o - \theta_v - 2\Delta)$$

where the first two terms represent the background intensity and the third terms represent the superimposed interference fringes. The extra phase shift Δ is due to the biprism deflection, which is dependent on the applied field and the geometric configuration of the biprism, but is not a property of specimen. The information that we want to obtain is the phase shift $\theta_o - \theta_v$, which is due to the modulation by the crystal potential V . The crystal electrostatic potential V is typically in the order of 10eV which is much less than the electron beam energy, so that the difference of momentum between object and vacuum waves is proportional to V . Considering relativistic effects, the relationship between phase shift $\theta_o - \theta_v$ and crystal potential V is:

$$\theta_o - \theta_v = \int \frac{p_{obj} - p_{vac}}{\hbar} dz = \frac{eV(E - E_0)}{E(E + 2E_0)} \frac{2\pi}{\lambda}$$

where E is the kinetic energy of incident electron beam, E_0 is the rest energy of electron, t is specimen thickness and λ is the wavelength of electron beam. With a fixed accelerating voltage in the TEM, this expression can also be written as:

$$\theta_o - \theta_v = C_E V t,$$

where C_E is a constant. Therefore, the crystal electrostatic potential profile can be obtained from the phase shift information that is extracted from the electron hologram.

The phase reconstruction process¹² starts by taking the Fourier transform of the intensity equation shown above. Three delta functions appear, involving convolutions represented by the symbol \otimes

$$FT(I) = \delta(q) \otimes FT(A_v^2 + A_o^2) + \delta(q - q_\Delta) \otimes FT(A_v A_o e^{i(\theta_v - \theta_o)}) + \delta(q + q_\Delta) \otimes FT(A_v A_o e^{-i(\theta_v - \theta_o)})$$

The background intensity is transferred into an autocorrelation function and appears to be in the center, while the Fourier transform of the co-sinusoidal interference fringes provides two sidebands, which are away from center by the carrier spatial frequency, q_d . The next step is to shift one of the sidebands to the center, and take the inverse Fourier transform. A complex object image with a modulus A_o and phase $\theta_o - \theta_v$ is obtained. In addition, if we remove the specimen and take the hologram, a complex image with a modulus A_v and zero phase shift can be obtained. The damping of amplitude of the electron beam across the specimen is due to the inelastic scattering process in the material and can be expressed as

$$\frac{A_o}{A_v} = e^{-t/2\lambda_i}$$

where λ_i here is the mean free path for inelastic scattering of electrons in the crystal. Therefore, the thickness profile of a specimen can be obtained from the amplitude image, while the electrostatic potential profile of the specimen can be obtained from the phase image. Furthermore, in a one-dimensional potential profile, the internal fields are proportional to the slope of the electrostatic potential. The corresponding electrostatic charges are found using Poisson's equation, and are related to the curvature of the potential profile.

3. Dislocations in GaN

As mentioned before, thin epitaxial films with optimized electronic properties have threading dislocation densities ranging from 10^8 to 10^{10} cm^{-2} .⁴ They play an important role in accommodating epitaxial growth under adverse conditions associated with lattice mismatch and thermal mismatch with the substrate and at InAlGaN compound heterojunctions. Threading dislocations have been found to be of the edge, screw, and mixed types.¹⁴ The Burgers vectors of the dislocations are large with respect to the lattice parameters, which often leads to the formation of coreless dislocations.¹⁵ Dislocations are frequently associated to non-radiative recombination centers.¹⁶ The electrostatic charges at dislocations lines were first studied using electron holography by Cherns¹⁷ and Cai.¹⁸ They found that dislocations have specific charges depending on the dislocation type.

4. Polarization fields in InGaN/GaN

In the absence of a center of symmetry, the hexagonal wurtzite structure exhibits noticeable spontaneous and piezoelectric polarization effects.⁷ Polarization fields up to a few MeV/cm in the nitride semiconductors were first determined by applying a reverse bias to the p-n junction,⁸ and later by electron holography.¹¹ The strong internal field can lead to positive and detrimental effects in light emitting diodes (LEDs). Reduced radiative recombination efficiency can result from the spatial separation of electrons and holes. Spatially separated by the Stark effect, their transition matrix elements and the radiative recombination rate can be significantly decreased. In addition, the internal fields facilitate the escape of carriers and reduce the carrier capture cross section by quantum wells.

Another detrimental effect is instability of the emission wavelength for devices operating under high injection conditions. The polarization field is screened and the emission wavelength is blue-shifted. The magnitude of this shift could be up to 100nm for large polarization cases, such as for InGaN/GaN green LEDs. Therefore, it is important to find ways to prevent polarization effects in the active region. Non-polar and semi-polar growth have been proposed and their fabrication has been successfully demonstrated. However, the in-plane polarization could still be a problem for the non-polar growth of quantum wells, as we will see later.

On the other hand, strong polarization fields could play a positive role in the conductivity of p-region and n-region in LEDs. In n-type AlGaIn/GaN heterostructures, the positive polarization sheet charges at the interface will promote the formation of two-dimensional electron gas (2-DEG) with a sheet carrier concentration as high as 10^{13} cm^{-2} . Similarly in p-type AlGaIn/GaN heterostructures, the negative polarization sheet charges at the interface could help with the formation of two-dimensional hole gas (2-DHG). By substituting the conventional n-GaN or p-GaN layer in LEDs with the proper AlGaIn/GaN multi-heterostructure, the lateral conductivity of LEDs can be enhanced, and an

improved current spreading can be expected.

In the following sections, we discuss the theoretical aspect of the polarization field in nitrides, and then the measurement of polarization effects by electron holography, including the mapping of internal electrostatic potential in the InGaN/GaN quantum wells and evidence of the formation of 2-DEG and 2-DHG in n-type and p-type AlGaIn/GaN heterostructures, respectively.

5. Electrostatic fields and charges in light emitting diodes emitting in the green region of the visible spectrum

Various aspects of polarization fields in InGaN/GaN quantum wells have been reported⁶⁻⁹ and measured.¹⁹⁻²¹ In particular, there is much interest in the polarization field and charge distribution in InGaN/GaN multi-quantum wells in the active region of green LEDs. An important feature of electron holography is that it can provide a mapping of the electrostatic potential distribution at an atomic scale, while other techniques usually give average values. An example of a green LED structure is shown in Fig. 3.²² This layer structure was grown by metal organic chemical vapor deposition on a c-plane sapphire substrate, the active region containing five InGaN/GaN quantum well periods, emitting light at ~ 2.3 eV. Cross-section samples were prepared for TEM using standard mechanical polishing and argon-ion milling techniques. Figure 3(a) is a high-resolution image of the active region. The InGaN QW and GaN barriers are uniform with a thickness of 4.6 and 9.3 nm, respectively. The interfaces are flat and atomically abrupt.

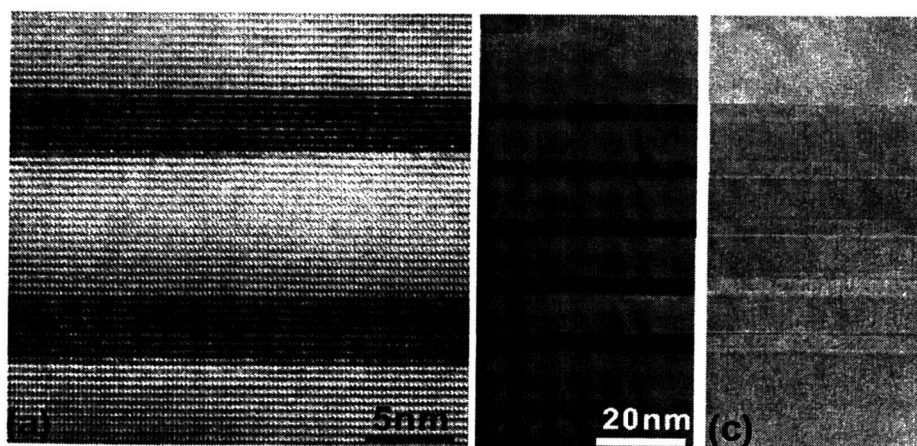


Figure 3. Active region of a green LED. (a) High resolution TEM of the quantum well region. (b) Electrostatic potential profile showing the p-n junction and quantum well region. (c) Charge density distribution due to polarization effects.²²

Electron holograms of the active region were obtained with the sample aligned to minimize diffraction conditions while keeping close to an orientation where the electron beam was parallel to the interfaces. The holograms were digitally recorded using a charge coupled device camera. The amplitude and phase values were deconvoluted from the electron hologram; the results are shown in Figs. 3(b) and 3(c). The resulting spatial profile of the electrostatic potential and charge distribution are plotted in Fig. 4. The electrostatic potential distribution due to the p-n junction is clearly observed in Fig. 4(a) with a shift of about 3 eV in the conduction band edge. Also observed is the potential profile due to the quantum wells. It is noticed that the field and the polarization charge density in the quantum wells vary with distance to the p-n junction. The sheet polarization charge densities of the first three InGaIn quantum wells have a similar value of ~ 0.018 C/m², corresponding to a 20% indium composition, for a non-screening case, and corresponding to piezoelectric fields of ~ 1.95 MV/cm. The sheet polarization charge densities in the fourth and fifth periods are estimated to be 0.007 and 0.0005 C/m², respectively. This corresponds to a significant screening of the fourth and fifth QWs.

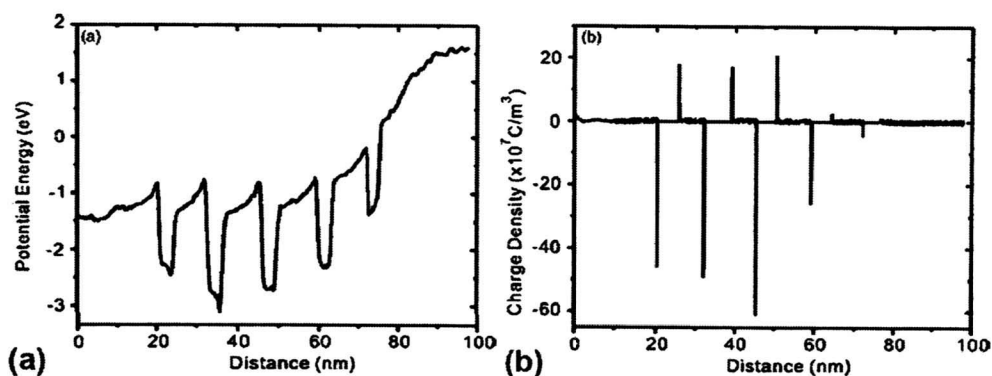


Figure 4. Active region of a green LED. (a) Electrostatic potential profile showing the p-n junction and quantum well region. (b) Charge density distribution due to polarization effects.²²

The observation of inhomogeneous polarization fields within the multi-quantum well is consistent with a depth-resolved study using cathodoluminescence,²² in which a low energy peak appears when the accelerating voltage is increased. The decay of piezoelectric fields in the fourth and fifth QWs could be due to strain relaxation. It is possible that the first three periods remain fully strained while the fourth and fifth periods undergo plastic relaxation. Another possibility involves the migration of hydrogen ions from the p-type region into the top two QWs.

6. Two-dimensional electron gas (2DEG) thin film structures

Strong polarization fields can result in two-dimensional carrier confinement, also known as two-dimensional electron gas (2DEG), in AlGaIn/AlN/GaN heterostructures. The high electron conductivity of these thin film structures can be used to provide low resistant current paths to n- and p-type layers of the LED.²³⁻²⁶ Fig. 5 shows part of a n-type AlGaIn/AlN/GaN superlattice, exhibiting 2-D electron gas at the AlN/GaN interface. The AlGaIn/AlN/GaN superlattice was epitaxially grown by MOCVD on an n-doped 10^{18} cm^{-3} , $3 \mu\text{m}$ GaN layer on a c-plane sapphire substrate. The SL consists of 11 periods; each period starts with a 0.5 nm AlN layer, followed by a 12 nm AlGaIn barrier layer with an Al concentration that is nominally graded linearly from 30% to 0%, and ends with a 28 nm GaN layer. The AlN layer is used to avoid interface alloy scattering effects and to improve the lateral mobility of the 2-DEG. Using a very thin AlN layer prevents significant reduction of the vertical conductivity. The AlGaIn layer has an n-type doping level of $2 \times 10^{19} \text{ cm}^{-3}$, and the n-type GaN in the SL is doped at $2 \times 10^{18} \text{ cm}^{-3}$. In order to avoid impurity scattering, doping of the AlGaIn layer was delayed by 4 s during growth, i.e., the first nanometer of the AlGaIn is undoped; and the doping of the GaN layer was stopped 10 nm before starting the growth of the next AlN/AlGaIn barrier. The characteristics of the full superlattice are observed in Fig. 5(a). A higher magnification TEM image showing the two top periods of the SL structure is shown in Fig. 5(b). The AlN layer appears with the brightest contrast, followed by the graded AlGaIn layer with fading brightness and the GaN layer with relatively darker contrast.

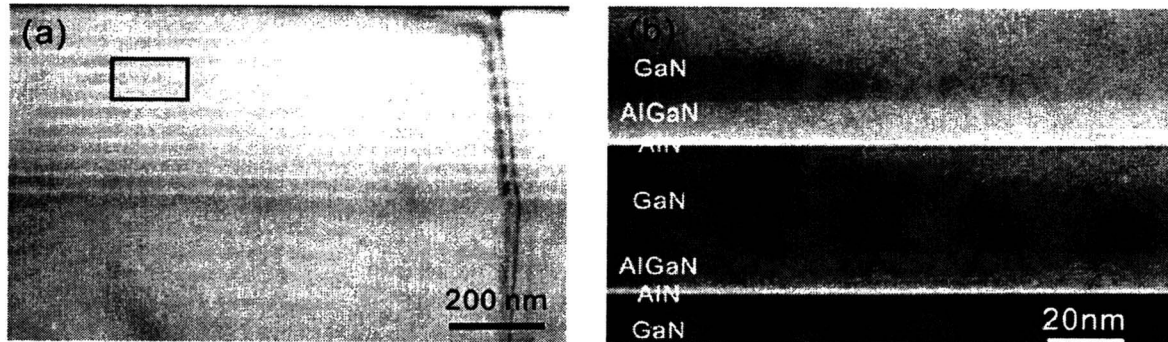


Figure 5. AlGaIn/AlN/GaN n-type superlattice. (a) Bright field TEM image showing the eleven periods. (b) Higher magnification TEM image showing the top two periods.²⁴

The electrostatic potential distribution obtained from the reconstructed phase and amplitude images are shown in Fig.6(a). A profile with higher spatial resolution is shown in Fig. 6(b). An energy dip close to the GaN/AlN interface represents the accumulation of the 2-DEG. The curvature of the potential around the 2-DEG region is negative as expected. These results are consistent with calculated energy band diagrams.²⁵

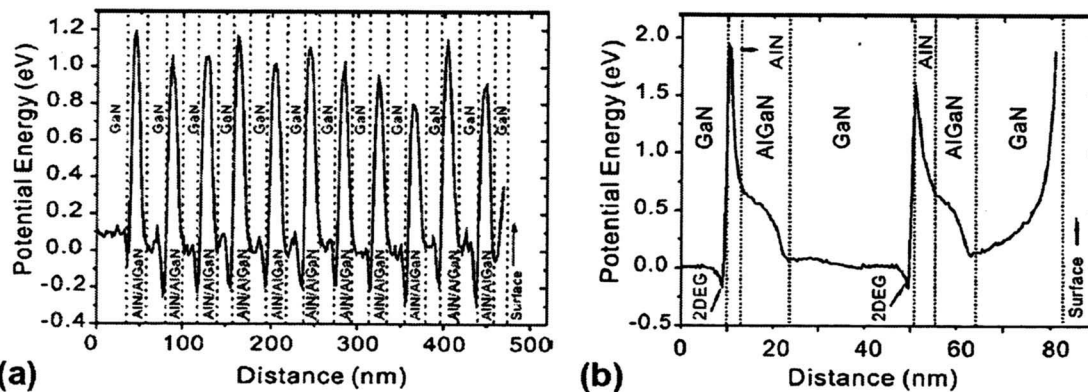


Figure 6. Potential energy profiles of AlGaIn/AlN/GaN n-type superlattice showing the features of the 2DEG in the complete superlattice (a) and in the top two periods (b).²⁴

7. Two-dimensional electron gas (2DHG) structures

If the AlGaIn/GaN multi-heterostructure is made p-type, the band diagram would be different from the n-type case due to the alignment of Fermi-level. In order to achieve the accumulation of a 2-DHG, the growth sequence of GaN/AlN/AlGaIn should be adjusted and the device structure should be optimized. We have studied a p-type AlGaIn/GaN multi-heterostructure, which is designed for the purpose of high conductivity with the accumulation of 2-DHG.²⁵⁻²⁶

The AlGaIn/AlN/GaN heterostructure in Fig. 7 was epitaxially grown by metal-organic chemical vapor deposition on a nominally un-doped 3- μm thick GaN layer on a c-plane sapphire substrate. The 2-DHG heterostructure consists of 6 periods; each period starts with a 0.5 nm AlN layer, followed by a 5 nm nominally un-doped GaN layer, and ending with a 12 nm AlGaIn barrier layer with an aluminum concentration that is nominally graded linearly from 0% to 10%. The thin AlN layer is used to avoid interface alloy scattering effects and to improve the lateral mobility of the 2-DHG. In order to optimize the vertical conductivity, a very thin AlN layer is used. The initial portion of the AlGaIn layer has a Mg doping level of $\sim 1 \times 10^{19} \text{ cm}^{-3}$, but the Mg flux is turned off for the last few nanometers. The nominally un-doped GaN shows, in secondary ion mass spectroscopy, a Mg background concentration in the range of $1\text{--}5 \times 10^{18} \text{ cm}^{-3}$, due to memory effects.

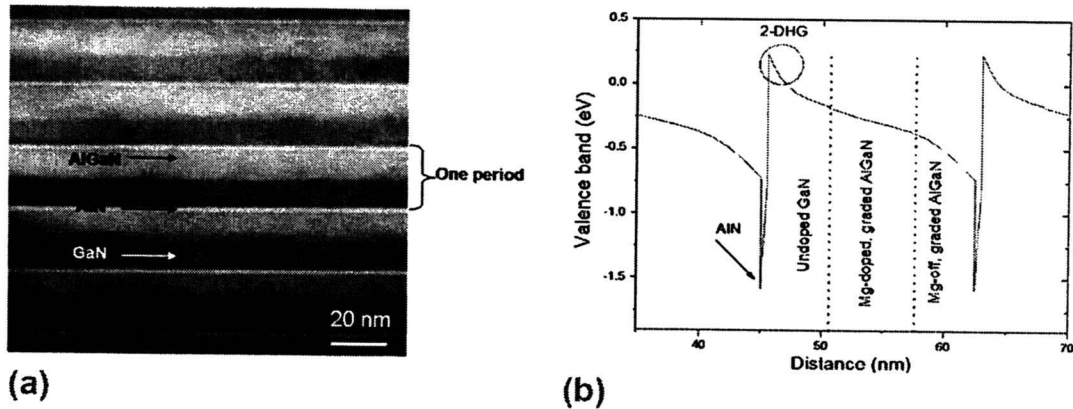


Figure 7. AlGaIn/AlN/GaN superlattice structure. (a) TEM image of the superlattice. Brighter regions indicate high aluminum content. (b) Sketch of valence band profile of one period. The 2-DHG accumulation region is circled.²⁶

Fig. 8 shows the electrostatic potential profile for 3 superlattice periods. The locations of the interface are determined by a careful match between diffraction contrast and phase image. The formation of a 2-DHG is revealed by a positive curvature in the GaN immediately to the right of the AlN layer in the potential energy profile. The two-dimensional hole gas (2-DHG) at the interface cancels the spontaneous polarization charges, and no discontinuity in the potential profile is observed. But the depletion of holes produces a positive curvature in the profile, in contrast to a negative curvature in the 2-DEG case.

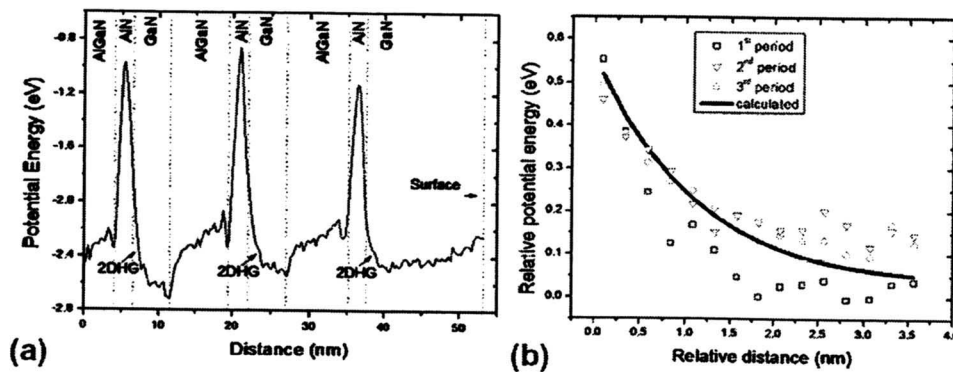


Figure 8. Potential profile of an AlGaIn/AlN/GaN superlattice doped p-type, exhibiting a 2-DHG. No discontinuity is observed at the AlN/GaN interface.²⁶

Generally, in semiconductor structures of this type, we find four major types of charges that are responsible for the variations of the electrostatic potential. (1) Fixed charges like those due to spontaneous and piezoelectric polarization, (2) ionized impurities, (3) interface fixed charges (band offset), and (4) free carriers (in our case holes) that redistribute under the effect of the other charges. For the 2-DHG case, the major contribution to the potential variation is from the holes confined on a nanoscale of length. In our case, the experimental and calculated potential distributions both show a noticeable curvature in the vicinity (~ 1 nm) of the interface, which indicate the 2DHG accumulation happens over only a few atomic monolayers of the interface.²⁶

8. Conclusion

Defects and interfaces in the nitride semiconductors are closely related to the electronic and optical properties via the induced polarization fields. A detailed understanding of the electrostatic potential properties of semiconductor structures is possible at near the atomic scale using electron holography in the TEM. A brief introduction to the technique has been presented in this paper, and some applications to the study of electrostatic fields associated with dislocations, quantum wells, and 2DEG and 2DHG heterostructures of nitride semiconductors have been discussed. This type of studies is essential in order to achieve high efficiency light emitting devices. Similar studies should also be useful for the development of photovoltaic devices based on these materials.

References

- [1] Ponce FA and Bour D P 1997 *Nature* **386**, 351-9.
- [2] Stampfl C and Van de Walle C G 1999 *Phys. Rev. B* **59**, 5521-35.
- [3] Wu J, Walukiewicz W, Yu K M, Ager III J W, Haller E E, Lu H, Schaff W J, Saito Y, and Nanishi Y 2001 *Appl. Phys. Lett.* **80** 3967-9.
- [4] Lester S D, Ponce FA, Craford M G, and Steigerwald D A 1995 *Appl. Phys. Lett.* **66**, 1249-51.
- [5] Ponce FA, Cherns D, Young W T, and Steeds J W 1996 *Appl. Phys. Lett.* **69**, 770-2.
- [6] Look D C and Sizelove J R 1999 *Phys. Rev. Lett.* **82**, 1237-40.
- [7] Bernardini F, Fiorentini V, and Vanderbilt D 1997 *Phys. Rev. B* **56**, R10024-28.
- [8] Takeuchi T, Wetzel C, Yamaguchi S, Sakai H, Amano H, Akasaki I, Kaneko Y, Nakagawa S, Yamaoka Y, and Yamada H 1998 *Appl. Phys. Lett.* **73**, 1691-3.
- [9] Wei Q Y, Li T, Wu Z H, and Ponce FA 2010 *Phys. Stat. Sol. A* **207**, 2226-32.
- [10] McCartney M R, Smith D J, Hull R, Bean J C, Voelkl E, and Frost B 1994 *Appl. Phys. Lett.* **65**, 2603-5.
- [11] Cherns D, Barnard J, and Ponce FA 1999 *Solid State Comm.* **111**, 281-5.
- [12] McCartney M R and M. Gaidardziska-Josifovska M 1994 *Ultramicroscopy* **53**, 283-8.
- [13] Cai J and Ponce FA 2002 *J. Appl. Phys.* **91**, 9856-62.
- [14] Ponce FA, Cherns D, Young W T, and Steeds J W 1996 *Appl. Phys. Lett.* **69**, 770-2.
- [15] Cherns D, Young W T, Steeds J W, Ponce FA, and Nakamura S 1998 *Phil. Mag. A* **77**, 273-86.
- [16] Ponce FA, Bour D P, Goetz W, and Wright P J 1996 *Appl. Phys. Lett.* **68**, 57-9.
- [17] Cherns D and Jiao C G 2001 *Phys. Rev. Lett.* **87**, 205504-7.
- [18] Cai J and Ponce F A 2002 *Phys. Stat. Sol. A* **192**, 407-11.
- [19] McCartney M R, Ponce FA, Cai J, Bour D P 2000 *Appl. Phys. Lett.* **76**, 3055-7.
- [20] Cai J, Ponce FA, Tanaka S, Omiya H, and Nakagawa Y 2001 *Phys. Stat. Sol. A* **188**, 833-7.
- [21] Stevens M, Bell A, McCartney M R, Ponce FA, Marui H, and Tanaka S 2004 *Appl. Phys. Lett.* **85**, 4651-3.
- [22] Wu Z H, Fischer A M, Ponce FA, Lee W, Ryou J H, Yoo D, and Dupuis R D 2007 *Appl. Phys. Lett.* **91**, 041915.
- [23] Wu Z H, Stevens M, Ponce FA, Lee W, Ryou J H, Yoo D, and Dupuis R D 2007 *Appl. Phys. Lett.* **90**, 032101.
- [24] Wu Z H, Ponce FA, Hertkorn J, and Scholz F 2007 *Appl. Phys. Lett.* **91**, 142121.
- [25] Hertkorn J, Thapa S B, Wunderer T, Scholz F, Wu Z H, Wei Q Y, Ponce FA, Moram M A, Humphreys C J, Vierheilig C, and Schwarz UT 2009 *J. Appl. Phys.* **106**, 013720.
- [26] Wei Q Y, Wu Z H, Sun K W, Ponce FA, Hertkorn J, and Scholz F 2009 *Appl. Phys. Exp.* **2**, 121001.

Semiconductor lasers

Carmen Gonzalez

France Telecom 1983-2001 / Alcatel Lucent – Thales III-VLab, 2002-2009. France.

*Summary of the short course given on the occasion of the
80th anniversary of Professor Holger Valqui – UNI 2011.*

Abstract

In this short-course, we described simply the physics of semiconductor lasers. After a few reminders of basic solid state physics on electronic states in semiconductors, how a semiconductor can emit light has been considered. The p-n junction is the basic mechanism for obtaining the population inversion required for stimulated emission gain. We presented the principles and fundamental equations governing the optical gain in these structures. The threshold conditions is then determined and the various improvements over the basic p-n junction (double heterostructure laser, quantum well laser) are set in historical perspective. Then, the variety of semiconductor lasers is illustrated by some examples: distributed feedback (DFB) lasers, distributed Bragg reflector (DBR) lasers, mode-locked lasers (MLL). Finally, an attempt to put into perspective some advanced semiconductor lasers, in particular the quantum dot lasers, has been made.

Introduction

Semiconductor lasers, invented in 1962, today have considerable importance in our societies. They are found in optical fibre telecommunications (internet, telephone, TV,...), for storing information on optical discs (CD, DVD, Blu-ray), on photocopiers, laser printers, in medical and industrial applications. They are, by economic standards and the degree of its applications, the most important of all lasers.

The main features that distinguish the semiconductor laser are: a) small physical size ($< 1\mu\text{m}$), b) direct pumping by low-power electric current (few tens of mW), c) high efficiency ($> 50\%$), d) ability to modulate its output by direct modulations of the pumping current at rates exceeding 20 GHz, e) possibility of integrating it monolithically with electronic and optical devices in III-V semiconductors to form integrated optoelectronic circuits.

1 - Basic principles of semiconductor lasers

The purpose of this section has been to provide a qualitative understanding of the physics behind the semiconductor lasers. Conditions required to operate a laser are discussed. The heart of a semiconductor laser is the p-n junction, which is why is also referred as diode laser. From this basic structure, the main characteristics of laser semiconductors have been deduced: the emission wavelength, λ , the threshold current density J_{th} , the modal optical gain G and the external differential quantum

efficiency η_D . Then, the p-i-n double-heterostructure (DH) is introduced. Semiconductor materials for DH diode lasers and epitaxial growth technologies are presented.

2 - Special semiconductor lasers structures

In this section, we intended to describe monomode lasers such as distributed feedback (DFB) lasers, distributed Bragg reflector (DBR) lasers, which utilizing a fine periodic structure, i.e. an optical Bragg grating, formed in the semiconductor structure. The Bragg grating is formed in the active section in DFB lasers, and outside the active section in DBR lasers.

Then, effects of size quantization in diode lasers are analyzed. We introduced the quantum well structures (QWL), and the phenomenological approach to the threshold current estimations through the gain parameters were studied. Also, the gain characteristics to the density of states of the active layer in connection with its dimensionality were correlated.

3 - New trends in semiconductor lasers

In this last section, latest advances in semiconductor lasers using quantum dots in the carrier recombination zone are presented (QDL). With these structures, it is expected: a) a significant reduction of the threshold current density in comparison to quantum well-based lasers and b) temperature-independent properties.

Finally, current trends on new structures such as quantum cascade lasers and monolithically integrated photonic circuits are presented.

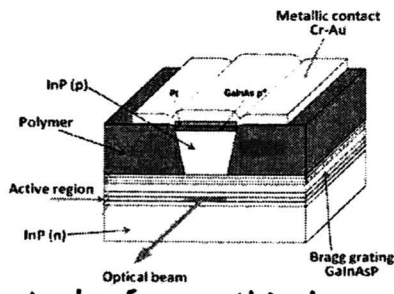
Bibliography

- 1- "Diode lasers and Photonic Integrated Circuits", L. Coldren and S. Corzine,, Ed. Wiley 1995.
- 2- "Semiconductor Lasers", G. P. Agrawal and N. K. Dutta, Ed. Kluwer 2002.
- 3- "Quantum electronics", A. Yariv, Ed. HWR-Saunders 1991.
- 4- "Quantum Dot Lasers", Victor M. Ustinov, A. Zhuvov, A. Egorov, N. Maleev, Ed. Oxford 2003

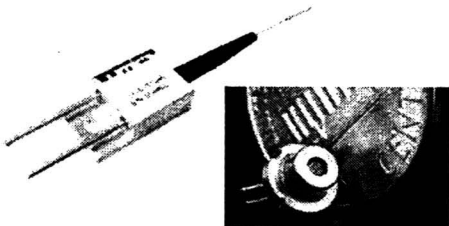
Semiconductor lasers
have invaded our daily lives

- ▶ Fiber optic telecommunication
- ▶ Optical discs: CD, DVD, Blu-ray
- ▶ Photocopier and printer lasers
- ▶ Medical and industrial applications
- ▶ Small physical size ($< 1\mu\text{m}$)
- ▶ Direct pumping by low-power electric current (few tens mA)
- ▶ High operating efficiencies ($> 50\%$)
- ▶ III-V semiconductor materials

InP-based diode laser



stack of very thin layers



Proton-fountain Electric-field-assisted Nanolithography (PEN)

A. La Rosa¹ and M. Yan²

Department of Physics¹ and Department of Chemistry², Portland State University, Portland, OR 97201, USA.

ABSTRACT

This chapter describes the implementation of Proton-Fountain Electric-field-assisted Nanolithography (PEN) as a potential tool for fabricating nanostructures by exploiting the properties of stimuli-responsive materials. The merits of PEN are demonstrated using poly(4-vinylpyridine) (P4VP) films, whose structural (swelling) response is triggered by the delivery of protons from an acidic fountain tip into the polymer substrate. Despite the probably many intervening factors affecting the fabrication process, PEN underscores the improved reliability in the pattern formation when using an external electric field (with voltage values of up to 5V applied between the probe and the sample) as well as when controlling the environmental humidity conditions. PEN thus expands the applications of P4VP as a stimuli-responsive material into the nanoscale domain, which could have technological impact on the fabrication of memory and sensing devices as well as in the fabrication of nanostructures that closely mimic natural bio-environments. The reproducibility and reversible character of the PEN fabrication process offers opportunities to also use these films as test bed for studying fundamental (thermodynamic and kinetic) physical properties of responsive materials at the nanoscale level.

Keywords Responsive materials, P4VP, nanolithography, swelling, polymer film, pH responsive, erasable patterns, PEN, biomimetic materials, DPN, hydrogels, osmotic pressure, entropy of mixing, protonation.

I. INTRODUCTION

PEN as a method for creating erasable nanostructures

The applications of tip-based nanolithography techniques that create patterns by anchoring molecules *onto* a surface of proper chemical affinity—as is the case in dip-pen nanolithography (DPN)^{1,2}—can be expanded by, alternatively, triggering the formation of nanostructures out of stimuli-responsive material substrates. Since stimuli-responsive properties may have a reversible character, the alternative nanofabrication approach could have concomitant implications in bio-technology (for creating switching gates that allow manipulating the transport, separation, and detection of bio-molecules, or for fabricating soft-material nanostructures that closely mimic natural bio-environments) as well as in emerging nano-electronics technologies (for fabricating low-cost and low-voltage operation integrated logic circuits in flexible substrates.) The potential technological implications that can be brought by harnessing the fabrication of nanostructures out of stimuli responsive materials underlines the interest for developing Proton-fountain Electric-field-assisted Nanolithography (PEN). In PEN the formation of nanostructures is triggered by the localized injection of protons into the substrate, with the charge-transfer from a sharp tip into the substrate being better controlled by the application of an external electric field. The development of PEN is thus conceived within the context of emerging developments in materials science and molecular engineering³ that pursue the design of devices that rely on the transduction of environmental signals.

PEN in the context of emerging biomimetic engineering

Inspired by the multi-functional inner working properties of living cell membranes,⁴ including the surprising sensitivity of their dynamic response to the mechanical properties of surrounding material,⁵ a current focus in biomimetic materials constitutes the development of versatile synthetic thin films that can selectively respond to a variety of signal interactions (mechanical, chemical, optical, changes in environmental conditions, etc.)⁶ In one approach, the complex synthetic hierarchy needed to eventually mimic *nature* is conceived as a combination of functional-domains separated by stimuli-responsive polymer thin films regulating the interactions between the domain compartments.⁷ In another approach, the cell is conceived not just as a chemical but also as a mechanical device,⁸ for it is found that the cell membrane is very sensitive to the mechanical properties of its surrounding matrix (affecting their growth, differentiation, migration, and, eventually, apoptosis),^{9,10} which has triggered an interest in the development of, for example, synthetic polymer scaffold for regenerative medicine.^{11,12} Both approaches, mentioned above, emphasize the need for harnessing the fabrication of synthetic thin film responsive materials.

The different approaches to biomimetic materials have resulted in the design of a variety of responsive building blocks (gels,¹³ brushes,¹⁴ hybrid systems with inorganic particles¹⁵) that respond selectively to different (pH,¹⁶ temperature,^{17,18} optical,^{15,19} and magnetic²⁰) external stimuli. Following the “bottom-up” route, functional materials have been prepared based on self-assembly of polymeric supramolecules.²¹ Progress following the alternative “top-down” approach

includes the fabrication of stimulus responsive polymer brushes,²² growth of polymers from previously DPN-patterned templates,²³ and chain polymerization of monomolecular layer by local stimulation using a STM tip^{24,25} (followed up by investigation of their working principle.²⁶) PEN falls in the top-down category approach. In the next section, we concentrate our description on hydrogels,²⁷ since the latter describes closer the experimental results obtained in current applications of PEN.

II. UNDERLYING WORKING MECHANISMS OF SWELLING IN HYDROGELS

This section provides a succinct summary of the main theoretical results underlying the working mechanisms involved in the swelling of hydrogels, where the concept of entropy plays a key role. In particular, it is worth to highlight the peculiar theoretical framework brilliantly introduced, time ago, by Paul J. Flory^{28,29} for analyzing polymer solutions; although his models have been refined, the essence of his clever approach is still used. It results quite interesting to identify (given the expected complexity of these polymer systems) the conceptual similarities between a) the analysis of a much simpler liquid-vapor system in equilibrium,^{30,31} and b) the analysis of the more complex (hydrogel) polymer solution³² (see Fig.1 below). For comparison and illustrative purposes both analyses will be presented here.

A hydrogel refers to a flexible (typically) hydrophilic cross-linked polymer network and a fluid filling the interstitial spaces of the network. The entire network holds the liquid in place thus giving the system a solid aspect. But contrary to other solid materials, these wet and soft systems are capable of undergoing very large deformation (greater than 100%). Understanding the dynamic behavior of hydrogels is worthwhile to pursue due to their widespread implications. In particular, the role of gels in living organism can not be exaggerated. As it is well put by Osada and Gong,³³ living organisms are largely made of gels (mammalian tissues are aqueous materials largely composed of protein and polysaccharide networks,) which enables them to transport ions and molecules very effectively while keeping its solidity.

What drives the swelling in a hydrogel? One of the potential mechanisms can be described in terms of the osmotic pressure (an entropic driven phenomenon), which help us understand how the additional entropy of mixing (afforded by an increase in the system's volume due to the absorption of water by the polymer network) is counteracted by a restoring force (also of entropic origin, since a larger dimension afford less polymer configurations) from the network itself. This osmotic pressure refers to the same type of phenomenon underlying the lower vapor-pressure displayed by an ideal solution (solvent + non-volatile solute) when compared to the vapor-pressure of a pure solvent. Since the latter constitutes a much simpler and familiar process, we conveniently include its (brief) description in the next paragraph. More specifically, we address the dynamics involved when a solution (water solvent + sugar solute) and pure water solvent are separated by a semi-permeable membrane, which can help us gain grasp knowledge of the osmotic pressure concept. The entropy of mixing

involved in this phenomenon is described in the framework of a lattice model (one in which solute and solvent molecules are considered to reside on the sites of a hypothetical lattice, the latter used as a resource that facilitates the calculation of the solution's entropy.) This approach will allow us to get familiar with lattice models, which are frequently used to describe polymer solutions (hydrogels). In short, we try to view the dynamics of hydrogels through the same prism used to view the equilibrium conditions of (water solvent + sugar solute) solutions.

II.A The osmotic pressure in ideal liquid solutions

The generation of a pressure difference (the osmotic pressure π_{osmotic}) across two phases as a consequence of their different vapor pressure is illustrated in Fig. 1a (see diagram on the left side). *Phase-A* (pure solvent water) and *phase-B* (solvent water + solute of sugar molecules) are separated by a membrane that allows the passage of water molecules but not the larger solute molecules.³⁰ It is an experimental fact that a dynamic equilibrium (*i.e.* equal rate, in both directions, of water molecules passing across the membrane) is reached when a hydrostatic pressure difference (the osmotic pressure π_{osmotic}) is established between the two phases. (Conceptually, the underlying mechanism at play here is that a water molecule in *phase-B* contributes greater to the total entropy than when inside the pure water *phase-A*; hence a net flow towards the former increases the total entropy.) The question to address is how to quantify this pressure difference.

Chemical potentials of an ideal solution and a pure solvent

A formal description of the osmotic pressure^{30,31} takes into account the fact that under initial conditions of equal pressure P , the chemical potential $\mu_{\text{water},P}^A$ of water in *phase-A* (pure solvent water) is greater than the chemical potential $\mu_{\text{water},P}^B$ of water in *phase-B* (as it will be justified in the next paragraph and the next section below.) Hence, when the phases are separated by a membrane permeable only to water, this constituent will not be in equilibrium and a net passage of water molecules from *phase-A* to *phase-B* is expected. As more water passes to *phase-B* the pressure increases and so does the chemical potential. Equilibrium with respect to the water constituent will then be established when μ_{water,P_2}^B (the chemical potential of water in *phase-B*, at the increasing pressure P_2) becomes equal to μ_{water,P_1}^A (the chemical potential of water in *phase-A*, at pressure P_1).³¹ The thus developed pressure difference ($P_2 - P_1$) is referred to as the osmotic pressure, π_{osmotic} .

To find a relationship between the change in chemical potential and the osmotic pressure, let's resort first to the extensive properties of the thermodynamic potentials³⁴ (namely, when the amount of matter is changed by a given factor, they change by the same factor.) A particular important relationship is obtained when this property is applied to the Gibbs free energy $G = G(T, P, N)$.

In effect, being the temperature T and pressure P intensive quantities, G has to have the form

$$G = N f(T, P),$$

where N is the number of particles of the analyzed system.

Since $dG = -S dT + V dP + \mu dN$ and $\mu = (dG/dN)_{T,P}$, the extensive property $G = N f(T, P)$ implies that μ is only a function of T and P ; that is,

$$\mu = G/N = f(T, P).$$

Accordingly, μ is the Gibbs free energy per molecule, and it is a quantity independent of N .

Thus,

$$d(G/N) = d\mu = - (S/N) dT + (V/N) dP,$$

which implies,

$$\frac{d\mu}{dP} = V/N.$$

This expression is pertinent to the quantification of the osmotic pressure. In effect, it reflects the change in chemical potential due to an increase in pressure, $\Delta\mu = (V/N) \Delta P$. Using $v \equiv (V/N)$, one obtains $\mu_{P_2} - \mu_{P_1} = v (P_2 - P_1)$, or,

$$\boxed{\mu_{P_2} - \mu_{P_1} = v (P_2 - P_1)} \quad (1)$$

(where it has been assumed that the volume does not change with pressure.)

Lattice model for calculating the entropy of mixing in ideal liquid solutions

An explicit calculation of the chemical potential difference, in terms of the number of constitutive molecules, can be derived by starting (at the most fundamental level) from a relatively simple combinatorial analysis of dissimilar solvent (water) and solute (sugar) molecules allowed to reside on the sites of a hypothetical lattice (see the diagram at the right in Fig. 1a),³² the latter introduced basically to facilitate the calculation the system's entropy of mixing ΔS_{mix} . In this lattice framework, the increase in the system entropy resulting from mixing N solvent and n solute molecules is given by $\Delta S_{\text{mix}}(N, n) = k \ln[(N+n)! / N! n!]$. For large values of N and n one can use the well known Stirling's approximation that gives $\ln n! = n \ln n - n \approx n \ln n$, or $n! \approx n^n$. Using this approximation, ΔS_M adopts the form $k \ln[(N+n)^{(N+n)} / N^N n^n]$, or

$$\Delta S_{\text{mix}}(N, n) = -k (N \ln f_N + n \ln f_n) \quad (2)$$

where $f_N \equiv N / (N+n)$ and $f_n \equiv n / (N+n)$ are the mole fractions of solvent (water) and solute (sugar) in the solution, respectively; k is the Boltzmann constant.

The Gibbs free energy has the general form $G = E + PV - TS = \mu N$. When applied to the system in *phase-B* (of N solvent and n solute molecules) at pressure P_2 and temperature T , one obtains,

$$G^{phase-B}(T, P_2, N, n) = G_{water}^{pure}(T, P_2, N) + G_{solute}^{pure}(T, P_2, n) - T \Delta S_{mix}(N, n) \quad (3)$$

where the first two terms on the right hand side are the Gibbs energy of the corresponding components in their pure state (we are assuming an ideal solution, so the components do not interact.)

The chemical potential of the water component in *phase-B* will then be given by,

$$\begin{aligned} \mu_{water}^{phase-B}(T, P_2, N, n) &= \frac{dG^{phase-B}}{dN}(T, P_2, N, n) \\ &= \frac{dG_{water}^{pure}}{dN}(T, P_2, N) - T \frac{d}{dN} \Delta S_{mix} \\ &= \frac{dG_{water}^{pure}}{dN}(T, P_2, N) + kT \frac{d}{dN} (N \ln f_N + n \ln f_n) \\ \mu_{water}^{phase-B}(T, P_2, N, n) &= \mu_{water}^{pure}(T, P_2, N) + kT \ln f_N \end{aligned} \quad (4)$$

Using (1)

$$= \mu_{water}^{pure}(T, P_1, N) + v(P_2 - P_1) + kT \ln f_N$$

At equilibrium, $\mu_{water}^{phase-B}(T, P_2, N, n) = \mu_{water}^{phase-A}(T, P_1, N)$. Therefore,

$$0 = v(P_2 - P_1) + kT \ln f_N \quad (6)$$

$$\pi_{osmotic} = v(P_2 - P_1) = -\frac{kT}{v} \ln f_N = -\frac{kT}{V/N} \ln f_N \quad (7)$$

Incidentally, since we are working with solution where $f_n \ll 1$ and thus $\ln f_N = \ln(1 - f_n) = -f_n$, useful equivalent formulas can be obtained,

$$\pi_{osmotic} = \frac{kT}{V/N} f_n \quad (\text{for a dilute solution}) \quad (8)$$

or, by expanding further f_n , one gets $\pi_{osmotic} = \frac{kT}{V/N} \frac{n}{n+N} \approx kT \frac{n}{V}$.

$$\pi_{osmotic} \approx kT \frac{n}{V} = RT \frac{\text{moles of solute}}{V} = \frac{RT}{MW} \frac{\text{mass of solute}}{V} \quad (9)$$

where n is the number of solute molecules in a volume V , and MW is the solute's molecular weight.

II.B Lattice model for describing ideal polymer solutions

The basic results displayed by expressions (2) and (7) constitute a proper starting step for describing more complicated (and apparently unrelated) cases like, for example, a cross-linked polymer network interacting with a pool of water, *i.e.* a hydrogel. The elegant twist in the coming description lies in considering the polymer network as the solute.^{32,35,36} That is, the trend of analysis using a hypothetical lattice conveniently remains the same; hence the observed similarity between the lattice displayed in Fig.1a (used to analyze a liquid solution system) and the lattice in Fig.1b (used to analyze a polymer solution.)^{37,38} In the latter, **one polymer molecule** is considered to be **a chain of x segments**, each segment (arbitrarily) considered equal in size to a solvent molecule. The objective becomes calculating the entropy of the polymer solution resulting from the different configurations that can be arranged with n_1 molecules of solvent and n_2 **polymer molecules** in a lattice containing $(n_1 + x n_2)$ cells.

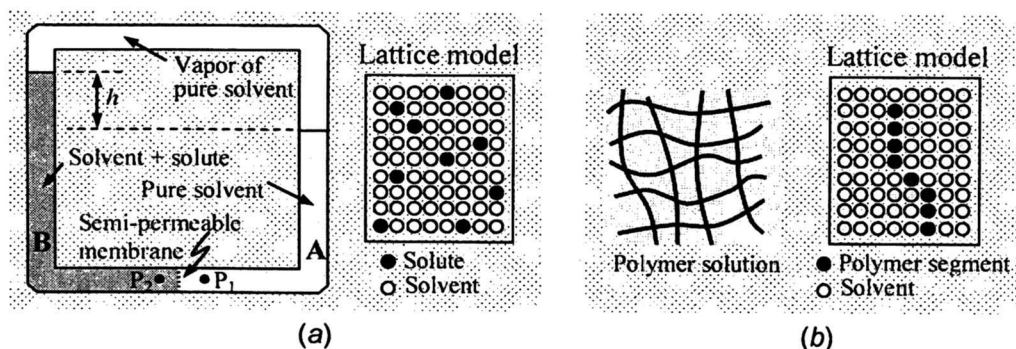


Figure 1. Two different thermodynamic systems studied using similar lattice model analysis. **a)** Left: Because of their different vapor pressures, a solution (water solvent + sugar solute molecules) and a pure water solvent (separated by a membrane permeable only to solvent molecules) generate an osmotic pressure ρgh (ρ is the density of the solution). Right: Solvent and solute molecules considered to reside in a hypothetical lattice, for entropy calculation purposes. The diagram on the left has been reproduced from page 47 of Ref. 30 [K. Huang, "Statistical Mechanics," John Wiley & Sons, 2nd Ed. (1987)] and reprinted with permission of John Wiley & Sons, Inc. **b)** Left: Polymer gel system. Right: Polymer chain considered as solute immersed in solvent, where all molecules

are considered to reside in a hypothetical liquid lattice. Figures adapted from Ref. 32 [P. J. Flory, "Principles of Polymer Chemistry," Cornell University Press (1969)], Copyright @ 1953 Cornell University and Copyright @ 1981 Paul J. Flory; used by permission of the publisher, Cornell University Press.

Lattice model for calculating the entropy of mixing

For the case of a polymer solution the calculation of the entropy can be conceived by counting first the different permutations associated with a given configuration of the polymer molecules (assuming no solvent molecules were present), and then adding the configurations resulting from their mixing with the solvent molecules (polymer segments and solvent molecules replacing one another in the liquid lattice.) While the former is expected to contribute more effectively in a process of polymer fusion, here the interest focuses mainly on the entropy of mixing. The latter takes, quite surprisingly again, a very simple form,^{29,32}

$$\Delta S_{\text{mixing}} = -k (n_1 \ln v_1 + n_2 \ln v_2) \quad (10)$$

where v_1 and v_2 are the volume fractions of solvent and solute respectively,

$$v_1 = n_1 / (n_1 + x n_2)$$

$$v_2 = x n_2 / (n_1 + x n_2)$$

Notice the similarity between expressions (2) and (10), except that volume fractions appear in the latter formula (mixing of molecules of different size) instead of mole fraction in the former (mixing of molecules of the same size).

The heat energy of mixing ΔE_M , the excluded volume effect, and the Helmholtz free energy ΔF_M

Given the fact that the dynamics of a polymer solution depends not only on the entropy but also on the energy of the system (the Helmholtz free energy $F=E-TS$ has to be minimum) this latter aspect is addressed in this section. In fact, the interactions between the water and the polymer molecules in a hydrogel make the polymer network a highly non-ideal thermodynamic system, where the cross-linked network structure plays an important role in determining the equilibrium aspects of the gel. In principle, any realistic model then has to take into account the intermolecular interactions due to the close proximity of the molecules, although some approximation can be applied depending on the temperature range being considered. On one hand, at relatively low temperatures a net attractive interaction between the monomers prevails, resulting in a net negative energy of the polymer system. On the other hand, at higher temperatures the repulsive interaction between the monomers when they are at very short distance (implicitly reflecting the fact that a monomer can not supplant the space already occupied by another monomer, a phenomenon better known in the jargon of polymer science as "excluded volume effect" or "excluded volume interaction")³⁹

will lead to a net positive energy of the polymer system. Below we provide some expressions that quantify this energy contribution.

In the lattice model, where each cell is able to accommodate either a solvent molecule or a polymer segment, the heat of mixing results from the replacement of some of the contacts between like-species (1-1 or 2-2) with unlike-constituents (1-2). If Δw_{12} represents the change in energy when one of these replacements occurs, then the heat associated with the formation of a particular configuration having p_{12} unlike-neighbors will be equal to $\Delta E_M = p_{12} \Delta w_{12}$.

For the calculation of p_{12} , it is plausible to assume that the probability a particular site adjacent to a polymer segment is occupied by a solvent molecule to be proportional to the volume fraction v_1 of the solvent. On the other hand, if the number of cells which are first neighbor to a given cell is z (here z is expected to be on the order of 6 to 12,) then zx is the number of contacts per polymer molecule, and $zx n_2$ would be proportional to the total number of contacts. p_{12} turns out to be then proportional to $(zx n_2) v_1$. On the other hand, using the definition $v_2 = x n_2 / (n_1 + x n_2)$, one obtains $n_1 v_2 = x n_2 n_1 / (n_1 + x n_2) = x n_2 v_1$. Hence, $\Delta E_M = p_{12} \Delta w_{12} \sim (zx n_2) v_1 \Delta w_{12} = (z v_2 n_1) \Delta w_{12} = (z \Delta w_{12}) n_1 v_2$. This result is typically expressed as,

$$\Delta E_M = kT \chi_1 n_1 v_2 \quad (11)$$

where the quantity $kT \chi_1 \equiv z \Delta w_{12}$ characterizes (for a given solute) the interaction energy per solvent molecule.

As mentioned above, at relatively low temperatures the net contribution from this term has a negative value. However, at relatively high temperatures, it was pointed out early on by Flory that the interpretation of χ_1 should be expanded as to include also other potential interactions between neighboring components that could have concomitant (positive value) contribution to the Helmholtz free energy. The new interpretation relates χ_1 to the number of pair-molecules collisions,³⁹ which should be proportional also to the number of pair contacts developed in the solution, just as for the heat exchange.

Using (10) and (11), the Helmholtz free energy of mixing is given by,³²

$$\begin{aligned} \Delta F_{\text{mix}} &= \Delta E_{\text{mix}} - T \Delta S_{\text{mix}} \\ &= kT (\chi_1 n_1 v_2 + n_1 \ln v_1 + n_2 \ln v_2) \end{aligned} \quad (12)$$

II.C Swelling of neutral polymer networks

Notice that the above formulation basically describes the mixing of two liquids; the view of a polymer network held up by its cross-links has not appeared yet. The calculation given in (12), however, is a good initial step towards calculating the structural entropy of network formation. The latter was originally done in a very clever way by Paul Flory and John Rehner in their seminal papers^{28,29} (latter improved by Flory),³⁷ who modeled the cross-linked network as a system composed of ν polymer chains of the same contour length (a chain

counted as a polymer thread between two cross-link intersections), with the location of the cross-link points positions defined, on average, on the vertex of a regular tetrahedron.²⁸ Any external distortion of the network (swelling, for example, due to the mixing of the polymers with a solvent) would be monitored by the distortion of this average tetrahedron cell.

In the Flory's description, the first step consists of calculating the different configuration that results from the dilution of ν (short) polymer chains (prior to the cross-linking) with n solvent molecules. This is given by expression (10), with ν playing the role of n_2 ,

$$\Delta S_D = -k \left(n_1 \ln \frac{n_1}{n_1 + x \nu} + \nu \ln \frac{\nu}{n_1 + x \nu} \right) \quad (13)$$

This is followed by a more elaborated calculation of the additional entropy corresponding to the different configurations that lead to the formation of a network of tetrahedron cells in a sea of solvent molecules.

$$\Delta S_{V'} = \Delta S_D + \Delta S_{\text{network of tetrahedron cells}} \quad (14)$$

Here the sub index V' stands for the final total volume of the swollen polymer network due to the mixing of the polymer with the solvent molecules. The objective here, however, is to calculate the net change in entropy $\Delta S_{\text{swelling}}$ due just to the swelling; that is,

$$\Delta S_{\text{swelling}} = \Delta S_{V'} - \Delta S_V \quad (15)$$

where ΔS_V stands for the entropy of the network when no solvent molecules are present. The result, in terms of the volume fraction $v_2 = x \nu / (n_1 + x \nu)$, is given by,³⁷

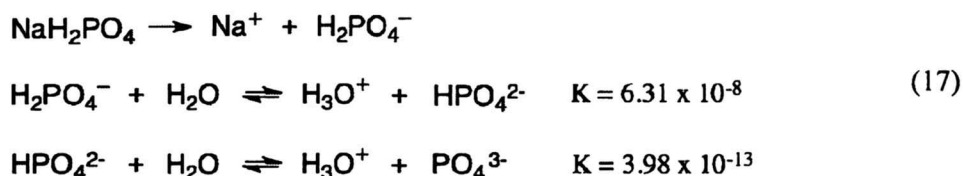
$$\Delta S_{\text{swelling}} = \underbrace{-k n_1 \ln (1 - v_2)}_{\text{Entropy of mixing polymers and solvent}} - \underbrace{\frac{3}{2} k \nu \left[(1/v_2)^{2/3} - 1 \right] - \frac{1}{2} k \nu \ln v_2}_{\text{elastic entropy arising from the deformation of the network}} \quad (16)$$

This expression reflects the contribution to the entropy from two different sources. Before dilution with the solvent molecules, the configuration of the network corresponds to one of maximum entropy, hence any deformation (due to swelling) would lead to a configuration of comparatively lower entropy. The latter then competes against the tendency for an entropy increase caused by the addition of solvent (and volume) that favors the creation of new configurations. Expression (15) embodies then the physical mechanism underlying the swelling process in an uncharged hydrogel. Using (11) and (15), the change in free energy is given by, $\Delta F = \Delta E_M - T \Delta S_{\text{swelling}}$. Equilibrium is governed by the condition $\Delta F = 0$.

II.D Swelling of Ionic Polymer Networks

Another potential channel for causing a polymer network to swell is the existence of charge centers, or molecular groups, which can strongly interact

among themselves and with other ions contained in the solvent. That is the case in a poly(4-vinylpyridine) (P4VP), whose pyridyl groups can react with hydronium ions H_3O^+ thus forming positively charged nitrogen centers (N^+). The situation is depicted in Fig. 2. If the fixed pyridinium cations were the only ions present there would be an exceedingly large electrostatic repulsion, but such interaction is partially screened by the presence of counterions resulting from the dissociation of the phosphate salts in water,



where K stands for the corresponding dissociation constants.

Notice in Fig. 2 that the equilibrium between the swollen ionic polymer network and its surroundings resembles the situation depicted in Fig. 1a where a membrane prevents the solute sugar molecules from entering the pure solvent region. In this case, the polymer acts as a membrane, preventing the charged ions from freely diffusing into the outer solution, establishing a higher concentration of mobile ions inside the network than in the outside (mainly because of the attraction of the fixed ions.) There thus exists an associated osmotic pressure arising from the difference in mobile ion concentration. Consequently, in addition to the swelling caused by the entropic mixing of polymer and water solvent, the fixed charges and counterions produce an additional driving force for the network to swell.

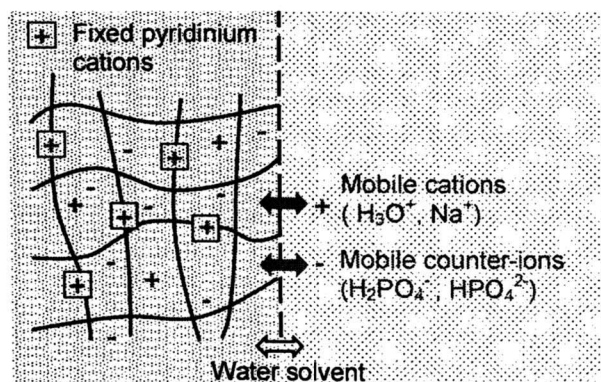


Figure 2. Schematic description of an exchange of ions and solvent between a P4VP polymer network and its surrounding electrolyte. Adapted from Ref. 32 [P. J. Flory, "Principles of Polymer Chemistry," Cornell University Press (1969)], Copyright © 1953 Cornell University and Copyright © 1981 Paul J. Flory; used by permission of the publisher, Cornell University Press.

III. FABRICATION PROCEDURE

III.A Preparation of the P4VP responsive material.

In a typical procedure, a solution of P4VP (molecular weight ca. 160,000) in n-butanol (10 mg/ml) is prepared with the reagents used as received. Silicon wafers with a native oxide layer are cut into square pieces $\sim 1\text{cm} \times 1\text{cm}$, and subsequently cleaned either by sonication in isopropyl alcohol for 15 minutes, or, alternatively, by immersion into piranha solution for 60 min at 80 °C followed by thorough cleaning in hot water. (Caution: the piranha solution reacts violently with many organic solvents.) Subsequently, the P4VP solution is spin-coated onto the wafers at 2000 rpm for 60 seconds. For crosslinking purposes, the sample is irradiated with a 450-Watts medium-pressure Mercury lamp (measured intensity of 5 mW/cm^2) for about 5 minutes. The irradiated films are then soaked in n-butanol for 24 hours to remove the unbound polymer. One way to estimate the thickness of the resulting film is to use an ellipsometer system. In that case, a value of 1.54 for the refractive index of the P4VP is used in the calculation.⁴⁰ This overall procedure gives film thickness in the 60 nm to 100 nm range.

III.B Preparation of the acidic fountain tip.

A source of hydronium ions H_3O^+ (or, if desired, hydroxide ions OH^- as well) is prepared out of phosphate buffered solutions, which have the remarkable property that can be diluted and still keep the same concentration of H_3O^+ . Different pH values can be obtained by dissolving corresponding quantity ratios of sodium dihydrogen phosphate (NaH_2PO_4) and sodium hydrogen phosphate (Na_2HPO_4) in distilled water. For example, mixing 13.8 g/l and 0.036 g/l of the two salts, respectively, gives 0.1 M buffer solution of pH equal to 4.0.⁴¹ This acidic solution serves as the source of hydronium ions which, upon penetrating a P4VP film, protonate the P4VP's pyridyl groups, as suggested in Fig.3.⁴² To achieve the protonation in very localized and targeted regions, however, PEN currently uses a sharp atomic force microscope (AFM) tip as an ion delivery vehicle in a similar fashion to dip-pen nanolithography.² For that purpose, as outlined in Fig. 4, an AFM tip of relatively high spring constant ($k = 40\text{ N/m}$) is coated by simply soaking the probe into the buffer solution for about 1 minute (Fig. 4a) and then allowing it to dry in air for 10 minutes or, alternatively, by blowing it with nitrogen (Fig. 4b.) Such fountain tip constitutes the probe for delivering hydronium ions very locally over targeted sites on a responsive material substrate.

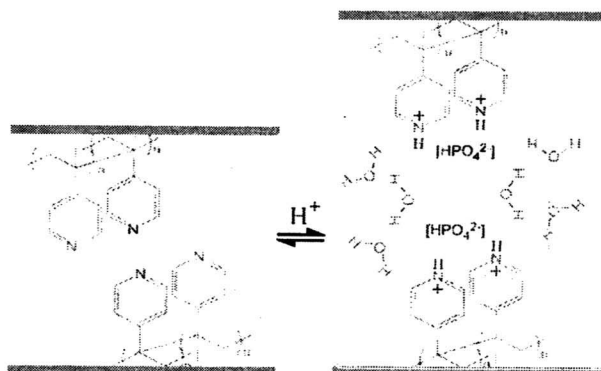


Figure 3. Schematic illustration of the reversible swelling mechanism in P4VP, suggested here to be a consequence of the corresponding electrostatic interaction upon the protonation of the pyridyl groups. Illustration adapted from Ref. 42 [C. Maedler, H. Graaf, S. Chada, M. Yan, and A. La Rosa, "Nano-structure Formation Driven by Local Protonation of Polymer Thin Films", *Proc. SPIE*, **7364**, 736409-1 (2009)] and reproduced with permission of the SPIE.

III.C Procedure for the local protonation.

The fountain-tip is mounted on the head-stage of an AFM system,⁴³ whose electronic station comprises lithography software for controlling the lateral scanning of the tip along pre-determined paths, and the capability to apply a bias voltage to the probe. The accumulated experimental results suggest that, upon bringing the probe into contact with the polymer film (contact force of $\sim 1 \mu\text{N}$), both *i*) the surrounding water meniscus that naturally forms between the tip and polymer film, and *ii*) the buffer concentration gradient, facilitate the transport of hydronium ions from the tip into the polymer (Fig. 4c). The subsequent protonation of the P4VP's pyridyl groups (as suggested in Fig. 3) gives rise to a net electrostatic repulsion that causes the corresponding polymer region to swell (Fig. 4d.) However initially successful (see Fig. 5a), this simple contact method unfortunately did not guarantee the pattern formation reproducibly. Subsequently we discovered that pattern formation occurred more consistently when applying an electric field between the probe and the substrate.

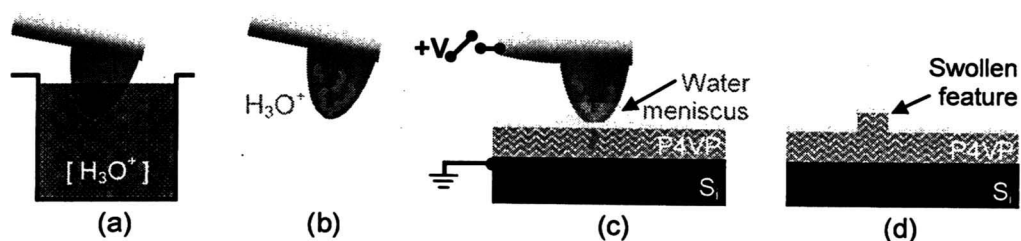


Figure 4. Schematic procedure for attaining spatially-localized protonation of a P4VP film resting on a silicon substrate. After dipping an AFM tip into an acidic buffer solution, the probe is pressed against a P4VP film. The transfer of

hydronium ions from the tip to the film is aided by the presence of a naturally-formed water meniscus (experiment performed at relative humidities above 40%.) The protonation of the P4VP's pyridyl groups triggers the swelling response of the film. The bias voltage helps the process to be more reliable and allows controlling the height of the features.

III.D Pattern Formation

First, with the "ink" loaded tip, an atomic force microscope (AFM) is set in "tapping imaging-mode," where the probe cantilever undergoes oscillations perpendicular to the sample's surface. To obtain an initial knowledge of the "blank paper" (a UV cross-linked P4VP polymer film), an image is taken at a relatively fast lateral scanning rate (5 $\mu\text{m/s}$). The use of high rates prevents transferring the buffer molecules into the substrate. Subsequently, the microscope is switched to "contact imaging-mode" for pattern formation under physical parameters controlled by the operator, namely contact forces of the order of 1 μN , 'writing' speeds up to 400 nm/sec, and fixed bias voltages up to 5V. Features of different planar morphologies can be generated with pre-programmed software designs, which guide the voltage-controlled XY lateral scanning of the tip while an electronic feedback-control keeps the probe-sample contact force constant. In our case, the sample rests on a XY piezo scanner stage that is equipped with strain-gauge sensors for overcoming piezoelectric hysteresis via another internal feedback control⁴³. The tip is held by an independent piezoelectric z-stage, thus conveniently decoupling the sample's horizontal XY scanning motion from the probe's vertical z-displacements. Finally, the microscope is switched back to the tapping mode for topography imaging in order to verify whether the patterns have been formed.

Figure 5a shows two sequentially acquired images of the first polymer structures created in our labs by exploiting the responsive characteristics of P4VP. The cross-linked P4VP film swelled only in the areas where the phosphate "ink" was delivered, forming two narrow-line terraces. These two images demonstrate the ability for sequentially creating an initial pattern, then imaging the resulting sample topography with the same ink-loaded tip, and subsequently creating additional patterns. The line profile in Fig. 5a reveals that both features are approximately 100 nm in height and 500 nm in width. The relatively large dimensions of these features (compared to the finer patterns we have recently created in our laboratories) may be attributed to the relatively large amount of ink initially attached to the tip (prior to the writing process.) That is, after dipping the tip into the buffer, no nitrogen was blown out in front of the "pen" to let it dry; the latter became afterwards a standard practice in our laboratory as a way to evaluate the reproducibility of the fabrication method under similar conditions as possible. The line profile also reveals the hydrogel characteristics of the patterns since the 100 nm constitutes a substantial swelling compared to the initial 80 nm thickness of the polymer film. Unfortunately, the reproducibility of these initial experiments was very poor from day to day, or month to month. At the time when the structures in Fig. 5a were created, no electric field was used in the experimental setting.

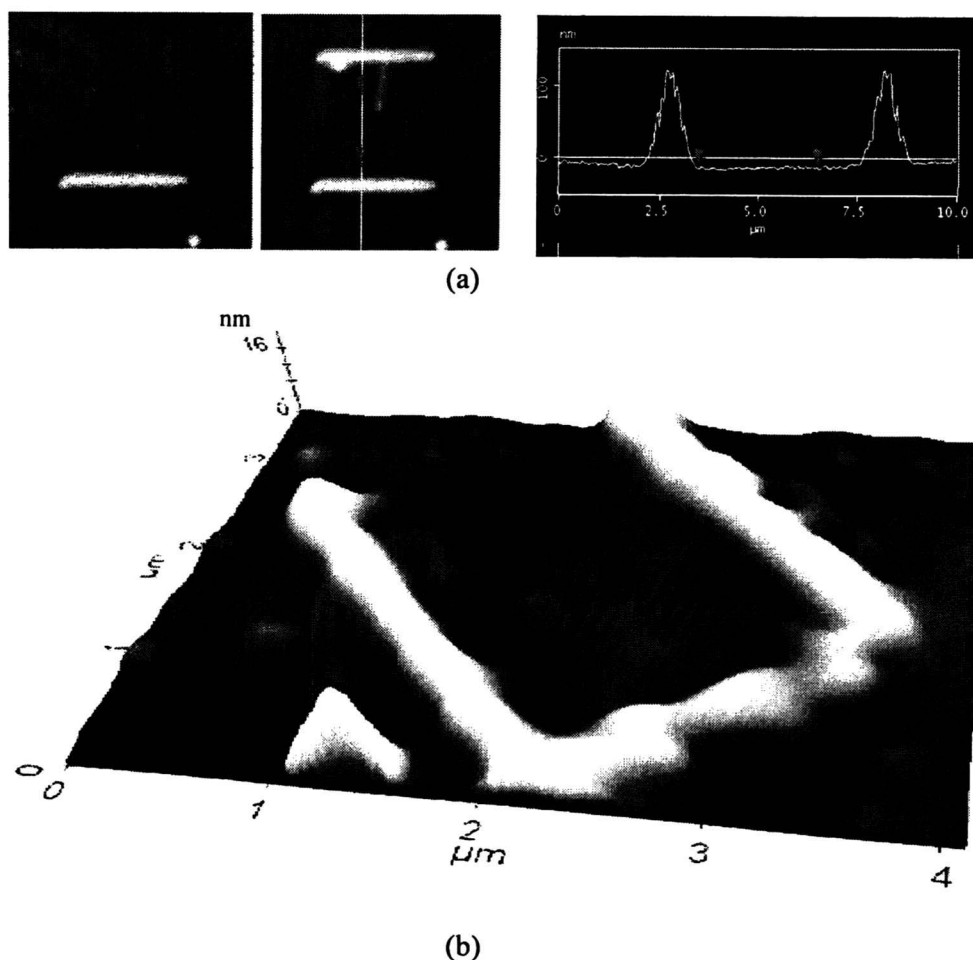


Figure 5 a) Top-view images ($10\mu\text{m}\times 10\mu\text{m}$) of two patterns formed sequentially by local protonation of a P4VP film (no electric field was applied.) The line profile displays one cross section of the second image. b) 3D-view of a "U" shape pattern fabricated on a $4\mu\text{m}\times 4\mu\text{m}$ region of another P4VP film (Contact force = $1\mu\text{N}$ and 5V tip-sample bias voltage.)

The reproducibility of the fabrication process greatly improves when an electric field is applied between the silicon tip and the silicon substrate (see setting in Fig. 4c). Typical bias voltages are up to 5 V, which when applied through a 50 nm film sets a strong electric field $\sim 10^6$ V/cm (still leaving the film apparently undamaged.) Provided that the humidity is above 40%, the patterns are consistently fabricated with this PEN method. Fig. 5b shows an 8 nm height

"U" shape structure fabricated under the new procedure, by applying a 5 V bias voltage and under 1 μN contact force. The smaller height of the structures (compared to the ones in Fig 5a) could be attributed to the minimal coating ink on the probe (nitrogen gas is blown on the tip right after dipping it into the buffer solution) thus a lower number of hydronium ions diffused through the polymer network. The lateral dimension of the line-features in both cases presented in Figs 5a and 5b is limited by the diffusion of hydronium ions into the polymer network. Still, under smaller contact forces and control of the dwelling time, line features as thin as 60 nm can be created, as shown in Fig. 6.^{44,45}

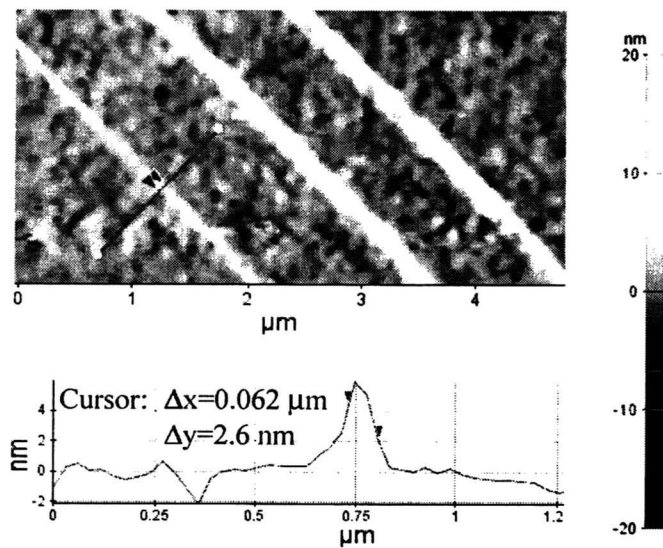


Figure 6 Three line features fabricated at 45% humidity using (from left to right) contact forces of 0.6 μN , 0.9 μN , 1.0 μN and bias voltages of 5V, 5V, and 4V respectively. The line profile across the line structure at the left displays a line feature ~ 6 nm tall and ~ 60 nm wide.⁴⁴

The effect of applying an electric field appears straightforward in PEN, helping drive the positive hydronium ions into the polymer network. To provide some context, here we mention other more sophisticated situations where the use of an electric field turns out to be also valuable. For example, an electric field is used to trigger specific conformational transitions switching between straight (hydrophobic) and bent (hydrophobic) states of low-density self-assembled monolayer (SAM) of 16-mercaptohexadecanoic acid on gold, which changes the wettability of the surface while maintaining the system's environment.⁴⁶ Also, nanometer-scale hydrogels, produced by surface grafting of a polymerizable monomer onto Au, undergo reproducible changes in thickness when a potential is applied across the film.⁴⁷ More strikingly, an infinitesimal change in electric potential across a polyelectrolyte gel produces volume collapse in polyelectrolyte gels.⁴⁸ A broader account on the effects of electric fields on gels has been given by Osada and Gong.³³

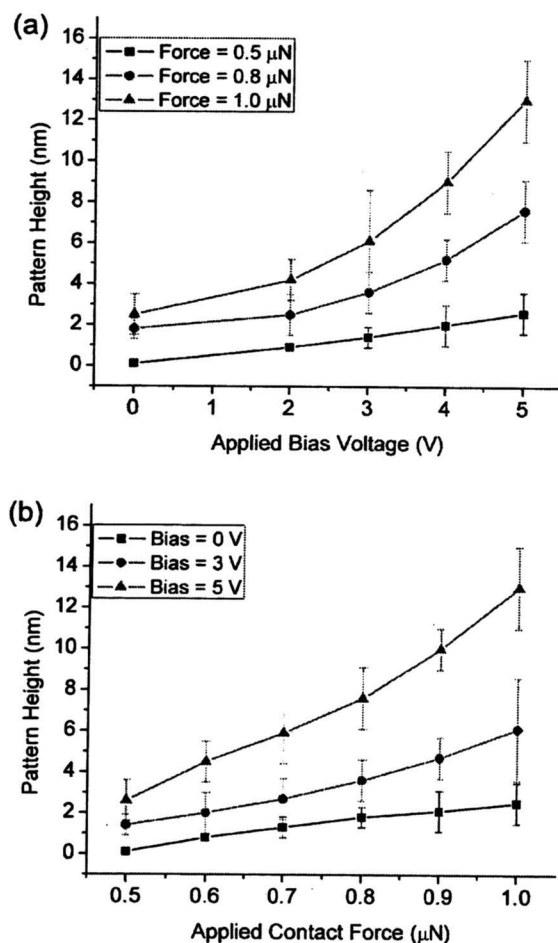


Figure 7. (a) Pattern height (nm) vs. applied bias voltage (V) at various fixed contact forces. (b) The pattern height (nm) vs. applied contact forces (μN) under different constant bias voltages. Reprinted, with permission, from Ref. 45 [Xiaohua Wang, Xin Wang, R. Fernandez, L. Ocola, M. Yan, and A. La Rosa; "Electric Field-Assisted Dip-Pen Nanolithography on Poly(4-vinyl Pyridine) Films," *ACS Appl. Mater. Interface* **2**, 2904–2909 (2010)], Copyright @ 2010, American Chemical Society.

The ability of PEN to create patterns having a variety of vertical dimensions in the nanometer range can be capitalized to study molecular interactions in these hydrogel systems with very much detail, including, for example, phase transitions. Polymer gels are known to exist in two phases (swollen and collapsed) where the volume transitions between the phases may occur either continuously or discontinuously.^{49,50} These transitions have been studied by monitoring the swelling as a function of different parameters (solvent quality, temperature,⁵¹ pH,⁵² visible light radiation,^{53,54} which occur as a result of a

competition between intermolecular forces (repulsive forces are usually electrostatic whereas attraction is mediated by hydrogen bonding⁵⁵, van der Waals interactions, or radiation pressure⁵⁶) that act to expand or shrink the polymer. However, these investigations do not disclose the microscopic view of the structure of gels. While neutron scattering has been used to elucidate these interactions at mesoscopic scales,⁵⁷ here we have an opportunity to complement these studies with nanoscale-sized individual gels fabricated via PEN.

IV. COMPARISON BETWEEN PEN AND OTHER DPN TECHNIQUES

A characteristic of the PEN technique is the dependence of the pattern dimensions on the positive contact force used to press the tip against the polymer film. The greater the applied external force, the larger the pattern size. This differs greatly from the DPN technique in which the line-width of the pattern is independent of the contact force.⁵⁸ DPN works even under gentle negative external forces (where the probe pulls away from the film, but the adhesion forces keep the probe attached to the sample) simply because its underlying mechanism implies just the deposition of molecules on the surface. In PEN, however, the buffer molecules (initially coating the AFM-tip) have to, in addition, penetrate into the polymer substrate. Experimental results indicate that the latter occurs less efficiently with weaker contact forces.⁵⁹ In fact, no apparent features are formed for contact forces smaller than 0.5 μN (at a writing speed of 80 nm/s,) even when applying bias voltages of up to 5V.⁴⁵ Certainly, the longer the dwelling time per pixel (i.e. slower writing speeds), the more hydronium ions diffuse into the polymer thus giving rise to larger patterns⁵⁹ which would also allow reducing the contact force. However, extremely long dwell times would make the technique less attractive, thus, a trade off exists between the writing speed and contact force for applying PEN efficiently. (Implementation of PEN in a parallel modality, in order to increase the throughput efficiency, is suggested in Section V below.)

Confirmation of a contrast difference between DPN and PEN is revealed from test experiments aimed at providing further evidence that the pattern formation in PEN results from the mechanical response of the polymer film and not just the deposition of buffer molecules:

First, when scanning a non-coated tip on a P4VP film under 3 μN contact force, no elevated features were obtained, as shown in Fig 8a.⁵⁹ Only scratches of 2 nm deep resulted from this operation, as evidenced by the line profile; the cross section indicates that the observed small protuberances are the result removed material from the scratches. This result suggests that a buffer solution is needed for the creation of PEN patterns.

Second, an argument could be made about whether or not, when using a coated tip, the patterns could result from the physical deposition of buffer molecules. To refute this argument, this time the test experiment was carried out using a buffer-coated tip scanned at low speed over a non-responsive polymer, polystyrene. The result presented in Fig. 8b shows no pattern formation, thus providing favorable evidence that features created were not just the deposition of buffer

molecules on a polymer surface, but the swelling response of the substrate itself when using responsive materials.

Third, the results given above also favor the hypothesis of protonation. But would the hydronium ions be the only ones diffusing into the polymer network? To test this hypothesis C. Maedler *et al* investigated the swollen structures.⁶⁰ Armed with Kelvin Probe Force Microscopy as a tool,⁶¹ and following up studies of deposition of charges in silicon,⁶⁰ they corroborated to have the sensitivity for monitoring the presence of a net charge in the polymer structures, if any. They found none. This implies that not only the hydronium ions penetrate the polymer, but they bring with them their corresponding counterions (see expression (17) above,) thus keeping the charge neutrality of the sample.

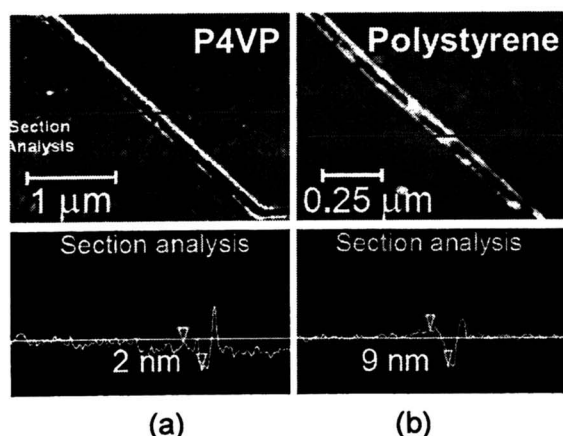


Figure 8 **a)** Scanning along a diagonal path with a contact force of 3 μN on a P4VP film, using an uncoated tip, produces only scratches, as revealed by the cross section line profile. **b)** Image of the resulting topography produced by a line scan on a polystyrene film (a non responsive material) using a contact force of 2 μN and a coated tip. A cross section line profile reveals only a scratch indent in the substrate. Reprinted, with permission, from Ref. 59 [C. Maedler, S. Chada, X. Cui, M. Taylor, M. Yan, and A. La Rosa, "Creation of nanopatterns by local protonation of P4VP via dip pen nanolithography." *J. Appl. Phys.* **104**, 014311 (2008)], copyright 2008, American Institute of Physics.

DPN and PEN agree on the role played by the humidity. PEN pattern formation in P4VP films under different humidity levels has been well documented by C. Maedler^{42,59} and X. Wang.⁴⁵ No features formation occurs below 40% of relative humidity, suggesting that a water meniscus⁶² naturally created between the probe and the polymer film plays a key role. In this regard, there is a coincidence of arguments with the DPN community that supports the hypothesis that a water meniscus is a fundamental component for the technique to work.⁶³ There is evidence that a liquid bridge exists even at zero humidity conditions.⁶⁴ (However, some authors have concluded that a meniscus is not involved in DPN experiments with certain molecules like 1-octadecanethiol.⁶⁵)

Finally, the reversibility of PEN contrasts with the, in general, non-reversible character of DPN. An exception to this notion is observed, however, in one DPN application where a voltage biased tip is used to create positive and negative charged nanopatterns on 1-hexadecanethiol (HDT) self-assembled monolayers (SAMs) on Au. The positive nanopatterns are gold oxide, which can be reduced back to gold by ethanol ink via DPN.⁶⁶ This development followed the “molecular eraser”⁶⁷ where a tip with a negative bias voltage (relative to the electrically grounded substrate that contain a monolayer of MHA 16-mercaptohexadecanoic acid) causes the molecules to desorb. Thus, regions of the alkanethiol can be selectively removed, which constitutes the pattern. The recessed areas can be refilled via DPN thus erasing the pattern. On the other hand, PEN instead is an intrinsically reversible process, using a base ink to erase the pillars previously formed by the acidic ink. The reversibility of the process was elegantly implemented by X. Wang *et al.*⁴⁵ using the same tip but inked with a basic phosphate buffer solution of pH 8.3, this time without applying a bias voltage. The removal of a pre-fabricated pillar structure was performed by keeping the tip in contact and stationary on top of the pillar. Results showed the progressive attenuation of the height of the structure. The basic buffer neutralizes the pyridinium converting it to the neutral pyridyl group and causing the film to “deswell”.

On the other hand, PEN differs from Electrochemical AFM Dip-Pen Nanolithography,^{68,69} where the tiny condensed water between the tip and the substrate is used as a nanometer-sized electrochemical cell in which metal salts (coating the probe) are dissolved, reduced into metals electrochemically, and deposited onto the surface. This method was used first to deposit Pt on a silicon substrate. The process involves coating an AFM cantilever tip with H_2PtCl_6 and applying a positive voltage to the tip relative to the electrically grounded substrate (the latter constituted by silicon with its native oxide). Despite the oxide, the conductivity is sufficient for the reduction (electrons gain) of the precursor ions PtCl_6^{2-} . During the process, H_2PtCl_6 dissolved in the water meniscus is electrochemically reduced from Pt(IV) to Pt(0) metal at the (cathode) silicon surface and deposits as Pt according to the reduction reaction $\text{PtCl}_6^{2-} + 4\text{e}^- \rightarrow \text{Pt} + 6\text{Cl}^-$. (The DC voltages are kept in the 1-4 V; higher voltages would tend to oxidize the silicon substrate and form SiO_2 instead of Pt nanostructures.) Electrochemical AFM Dip-Pen Nanolithography is therefore a technique where the deposited materials constitute the nanostructure, similar to DPN but different than PEN.

PEN is also different from Constructive Nanolithography (CNL)^{70,71} and its similar Electro Pen Nanolithography (EPN).⁷² Contrasting Electrochemical AFM Dip-Pen Nanolithography (described above), whose working principle is based on *reducing* (gaining of electron) a precursor ion, CNL and EPN exploit an *oxidation* (loss of electron) process instead. In one application of CNL, a silicon substrate is first coated with a 2.5 nm thick monolayer of octadecyltrichlorosilane

(OTS, an organic molecule with a methyl-terminated, 18-carbon alkyl chain,) which serves as the initial blanket surface to be oxidized on specific regions by the apex of a conductive AFM tip. The oxidation process is implemented by applying a positive bias voltage to the substrate relative to the tip (the latter kept at electrical ground.) As a result, the oxidized areas (initially hydrophobic, neutral, and inert) become hydrophilic, negative-charged, and chemically active, forming the multifunctional templates for post, or in-situ, surface generation of organic (insulator), metal, or semiconductor nanocomponent features, including the implementation of a wetting-driven patterning.^{73,74} The post pattern generation process of CNL is elegantly alleviated in EPN where the electrically biased conductive AFM tip is also coated with ink (trialkoxysilane and quaternary ammonium salts), which are in situ transferred during the scanning process as the local oxidation occurs. The virtue of CNL and EPN is that the transfer of molecules occurs only on the regions that have been oxidized, which can be as small as 25nm, thus bypassing the diffusion effect that limit the lateral resolution of DPN. CNL and EPN therefore have characteristics similar to DPN, hence different than PEN, for their patterns are constituted by anchoring molecules on the substrate, although a chemically modified substrate.

PEN is more alike to Chemical Lithography (ChemLith).⁷⁵ This technique exploits the fact that photoresist materials change their solubility upon an acid-catalyzed chemical reaction. As an alternative to the diffraction limited photolithography method (where the photoacid generator is mixed in the resist formula and the acid is generated by photon-initiated reactions), ChemLith delivers the catalyzing acid proton source to the desired position on a negative resist film via either a nanoimprinting method or by using a sharp stylus. In a post bake step, the resist molecules are catalyzed to cross-link with each other, and thus become insoluble in the final development step. Thus ChemLith nanopatterns result from a local protonation process (similar to PEN) and a post baking step (the latter renders the process irreversible, unlike PEN).

As mentioned at the beginning of this Chapter, PEN joins other efforts for developing responsive materials (including polymer nano-composites,⁷⁶ soft hydrogels^{3,7} and chemically functionalized metal nanoparticles,^{15,77} which may also be driven by proton sources as well as the use of bias voltages), but putting emphasis into the nanoscale size regime. This PEN development occurs in parallel to other electrochemistry SPM-based efforts for fabricating energy storage devices,⁷⁸ and, more generally, for attaining reliable surface modification using SPM.^{79,80}

V. PERSPECTIVES

We have introduced Proton-fountain Electric-field-assisted Nanolithography (PEN) as a technique for fabricating erasable nanostructures that closely mimic natural bio-environments. Its distinct feature, contrasting other DPN-based techniques, resides in the fact that the pillar patterns are made of the substrate material itself (a polymer film). The pattern formation results from the delivery of hydronium ions through an acidic-fountain tip in contact with the substrate,

causing the polymer film to swell at targeted locations. In addition, the patterns can be erased with a basic-ink loaded tip. Such a reversible character of the fabrication process could be relevant to a wide range of potential applications, ranging from microelectronics (the swollen and no-swollen states could represent the ones and zeros in a memory device) to biotechnologies (variable size gates that open and close compartment in micro- and nanofluidic devices used for separation of molecules or chemical reagents.) Further, its capability for controlling the dimension of the patterns with nanometer precision, via an electric field, offers an opportunity to use these films as testbed for studying fundamental (thermodynamic and kinetic) physical properties of responsive materials at the nanoscale level. This is concomitant to the far reaching goal towards developing thin film-based responsive materials that can selectively respond to a variety of external stimuli (mechanical, chemical, optical, and changes in environmental conditions.)

Despite its current progress, PEN would benefit from further development in order to place the technique in more solid grounds. For example, it would be convenient to study with more detail the effects of humidity (at different levels) in conjunction with the application of a range of bias voltages; until now their effects have been reported only separately. The objective would be to understand, optimize, and to know better, the dynamics involved in the transportation of hydronium-ions from the tip to inside the polymer network. This study would also help to elucidate why relatively strong forces ($\sim\mu\text{N}$) are needed for a rapid pattern formation to occur in PEN. Apparently, the first monolayers of water,^{81,82} found naturally adsorbed on the polymer film, is the first barrier that the ions from the tip face before rapidly diffusing into the polymer film. Functionalizing the responsive polymer film with, in turn, hydrophobic or hydrophilic layers would help to contrast this hypothesis. In passing, the latter proposed experiment could also lead to the formation of the thinnest line feature under DPN, for by writing on an hydrophobic-functionalized responsive-material film, only the region in contact with the tip would create a richer water content bridge for the hydronium ions to penetrate into the polymer network. Certainly the thinnest line width would be determined by the diffusion of the ions while inside the polymer; still, the degree of cross-link of the starting material polymer film could be used to influence this diffusion, which offers another variable to control the ultimate resolution in PEN.

To gain versatility on the PEN implementation, and with a perspective on application for building memory devices (the reversibility of the PEN allowing the implementation of memory level changes,) it would be worth to explore the use of solid state sources of protons.^{75,83,84} Such an application could benefit from current efforts for developing solid acid materials as electrolytes in fuel cells.^{85,86} In that direction of technological applications, it would be worth to also explore the implementation of PEN in a parallel format; that is, to develop capability for fabricating many patterns at once. For this application, PEN can capitalize on the fact that relatively strong forces are needed to fabricate the patterns, which provides some leverage to implement a stamping type fabrication modality. For example, metallic (master) features can be fabricated on a flat (glass) substrate, all of them interconnected as to be able to apply a bias voltage. After spin-

coating a layer of buffer solution on the metallic master features, the resulting wet stamp would be pressed against a P4VP film. By applying a bias voltage, the protonation would be more effective on the regions defined by the metallic features, hence the patterns from the master stamp will be replicated onto the polymer substrate by its swelling reaction. Further, this methodology could become very versatile, since different metallic regions on the master stamp could be electrically addressed at will; thus a given mask would be used for fabricating different patterns according to a programmable voltage pattern.

Finally, it is worth to point out that the intrinsic hydrogel nature of the P4VP patterns fabricated with PEN constitutes an advantage for replicating structures that closely mimic natural bio-environments. This capability is particularly relevant to a recent trend in research that conceives the cell not only as a chemical factory but also as a sensitive mechanical device. Incidentally, it has recently been reported that stem cells do not regenerate efficiently in vitro environment unless the surrounding medium is made out of flexible gels.⁸⁷ Hence, further developments on PEN should benefit the implementation of these attractive bio-engineering applications.

VI. REFERENCES

- ¹ K. Salaita, Y. Wang and C. A. Mirkin, "Applications of dip-pen nanolithography," *Nature Nanotechnology* **2**, 145 (2007).
- ² R. D. Piner, J. Zhu, F. Xu, S. Hong, and C. A. Mirkin, "Dip-Pen Nanolithography," *Science* **283**, 661 (1999).
- ³ I. Tokarev and S. Minko, "Stimuli-responsive hydrogel thin films," *Soft Matter* **5**, 511 (2009).
- ⁴ L. Anson, "Membrane protein biophysics," *Nature* **459**, 343 (2009).
- ⁵ C. Ainsworth, "Stretching the imagination," *Nature* **456**, 696 (2008).
- ⁶ B. Bhushan, "Biomimetics: lessons from nature-an overview," *Phil. Trans. R. Soc. A* **367**, 1445 (2009).
- ⁷ I. Tokarev, M. Motornov, and S. Minko; "Molecular-engineered stimuli-responsive thin polymer film: a platform for the development of integrated multifunctional intelligent materials," *J. Mater. Chem.* **19**, 6932 (2009).
- ⁸ C. Cofield, "Cell is mechanical device," *The American Physical Society, APS news*, Series II, **19**, 4 (June 2010).
- ⁹ C. Wu, Y. Li, J. H. Haga, R. Kaunas, J. Chiu, F. Su, S. Usami, and S. Chien, "Directional shear flow and Rho activation prevent the endothelial cell apoptosis induced by micro patterned anisotropic geometry," *PNAS* **104**, 1254 (2007).
- ¹⁰ C. S. Chen, M. Mrksich, S. Huang, G. M. Whitesides, and D. E. Ingber, "Geometric control of cell life and death," *Science* **276**, 1425 (1997).
- ¹¹ M. A. Greenfield, J. R. Hoffman, M. Olvera de la Cruz, and S. I. Stupp, "Tunable mechanics of peptide nanofiber gels," *Langmuir* **26**, 3641 (2010).
- ¹² M. M. Stevens, J. H. George, "Exploring and engineering the cell surface interface," *Science* **310**, 1135 (2005).
- ¹³ S. Maeda, Y. Hara, T. Sakai, R. Yoshida, and S. Hashimoto, "Self-walking gel," *Adv. Mater.* **19**, 3480 (2007).

- ¹⁴ T. K. Tam, M. Ornatska, M. Pita, S. Minko, and E. Katz, "Polymer brush-modified electrode with switchable and tunable redox activity for bioelectronic applications," *J. Phys. Chem. C* **112**, 8438 (2008).
- ¹⁵ D. Wang, I. Lagzi, P. J. Wesson, and B. A. Grzybowski, "Rewritable and pH-sensitive micropatterns based on nanoparticle 'Inks'," *Small* **6**, 2114 (2010).
- ¹⁶ J. Ruhe, M. Ballauff, M. Biesalski, P. Dziezok, F. Grohn, D. Johannsmann, N. Houbenov, N. Hugenberg, R. Konradi, S. Minko, M. Motornov, R. R. Netz, M. Schmidt, C. Seidel, M. Stamm, T. Stephan, D. Usov and H. Zhang, "Polyelectrolyte brushes" in: *Polyelectrolytes with defined molecular architecture*, M. Schmidt Ed. *Adv. Polym. Sci.* **165**, 79 (2004).
- ¹⁷ C. Liu, H. Qin, P. T. Mather, "Review of progress in shape-memory polymers," *J. Mater. Chem.* **17**, 1543 (2007).
- ¹⁸ R. Yoshida, K. Uchida, Y. Kaneko, K. Sakai, A. Kikuchi, Y. Sakurai and T. Okano, "Comb-type grafted hydrogels with rapid deswelling response to temperature changes," *Nature* **374**, 240 (1995).
- ¹⁹ T. Suzuki, S. Shinkai, K. Sada, "Supramolecular crosslinked linear poly(trimethylene iminium trifluorosulfonimide) polymer gels sensitive to light and thermal stimuli," *Adv. Mater.* **18**, 1043 (2006).
- ²⁰ B. A. Evans, A. R. Shields, R. Lloyd Carroll, S. Washburn, M. R. Falvo, and R. Superfine, "Magnetically actuated nanorod arrays as biomimetic cilia," *Nano Lett.* **7**, 1428 (2007).
- ²¹ O. Ikkala and G. Brinke, "Functional materials based on self assembly of polymeric supramolecules," *Science* **295**, 2407 (2002).
- ²² M. Kaholek, W. Kyung Lee, B. LaMattina, K. C. Caster, and S. Zauscher, "Fabrication of stimulus-responsive nanopatterned polymer brushes by scanning-probe lithography," *Nano Lett.* **4**, 373 (2004).
- ²³ X. Liu, S. Guo, and C. A. Mirkin, "Surface and site-specific ring-opening metathesis polymerization initiated by dip-pen nanolithography," *Angew. Chem. Int. Ed.* **42**, 4785 (2003).
- ²⁴ Y. Okawa, M. Aono, "Nanoscale control of chain polymerization," *Nature* **409**, 683 (2001).
- ²⁵ S. P. Sullivan, A. Schnieders, S. K. Mbugua, and T. P. Beebe Jr, "Controlled polymerization of substituted diacetylene self-organized monolayers confined in molecule corrals," *Langmuir* **21**, 1322 (2005).
- ²⁶ Y. Okawa, D. Takajo, S. Tsukamoto, T. Hasegawa and M. Aono, "Atomic force microscopy and theoretical investigation of the lifted-up conformation of polydiacetylene on a graphite substrate," *Soft Matter* **4**, 1041 (2008).
- ²⁷ P. Calvert, "Hydrogels for soft machines," *Adv Mater.* **21**, 743 (2009).
- ²⁸ P. J. Flory and J. Rehner, "Statistical mechanics of cross-linked polymer networks I. Rubberlike elasticity," *J. Chem. Phys.* **11**, 512 (1943).
- ²⁹ P. J. Flory and J. Rehner, "Statistical mechanics of cross-linked polymer networks II. Swelling," *J. Chem. Phys.* **11**, 521 (1943).
- ³⁰ K. Huang, "Statistical Mechanics," John Wiley & Sons, 2nd Ed. (1987).
- ³¹ F. P. Chinard and T. Enns, "Osmotic pressure," *Science* **124**, 472 (1956.)
- ³² P. J. Flory, "Principles of Polymer Chemistry," Cornell University Press (1969).
- ³³ Y. Osada and J. Gong, "Soft and wet materials: Polymer gels," *Adv. Mater.* **10**, 827 (1998).
- ³⁴ L. D. Landau and E. M. Lifshitz, "Statistical Physics" 3rd Edition Part 1, Elsevier (2006). See pages 72 and 267.
- ³⁵ Lattice models of polymer solutions are widely used for their simplicity and computational convenience. Their use for predicting solution properties of polymers solutions dates back to the 1940s.
- ³⁶ P. J. Flory, "Thermodynamics of high polymer solutions," *J. Chem. Phys.* **10**, 51 (1942).

- 37 P. J. Flory, "Statistical mechanics of swelling of network structures," *J. Chem. Phys.* **18**, 108 (1950).
- 38 W. Hu and D. Frenkel, "Lattice-model study of the thermodynamic interplay of polymer crystallization and liquid-liquid demixing," *J. Chem. Phys.* **118**, 10343 (2003).
- 39 A. Yu. Grosberg and A. R. Khokhlov, "Giant molecules. Here, there and everywhere," Academic Press (1997).
- 40 D. Woo, "Spectroscopic Ellipsometry Studies of Polymers on Silicon Wafer," Thesis for the Master degree in Physics, Portland State University (2009).
- 41 Since all phosphate salts are used in hydrated condition, the molecular weight (MW) should include the corresponding portion of water. For NaH_2PO_4 we should include one molecules of water, hence the MW is 137.99. On the other hand, for the Na_2HPO_4 we should consider 7 molecules of water (heptahydrate,) which gives a MW of 268.07. Hence, if 13.8 g and 0.036 g of NaH_2PO_4 and Na_2HPO_4 are used respectively, then we can quote the concentration of NaH_2PO_4 (the buffer strength) to be practically equal to 0.1M.
- 42 C. Maedler, H. Graaf, S. Chada, M. Yan, and A. La Rosa, "Nano-structure formation driven by local protonation of polymer thin films", *Proc. SPIE*, **7364**, 736409-1 (2009).
- 43 AFM XE-120 from Park Systems Inc.
- 44 Xiaohua Wang, "Characterization of mesoscopic fluid-like films with the novel shear-force/acoustic microscopy," Thesis for the Master degree in Physics, Portland State University, Portland, Oregon (2010).
- 45 Xiaohua Wang, Xin Wang, R. Fernandez, L. Ocola, M. Yan, and A. La Rosa; "Electric field-assisted dip-pen nanolithography on poly(4-vinyl pyridine) films," *ACS Appl. Mater. Interface* **2**, 2904–2909 (2010).
- 46 J. Lahann, S. Mitragotri, T. Tran, H. Kaido, J. Sundaram, I. S. Choi, S. Hoffer, G. A. Somorjai, R. Langer, "A reversibly switching surface," *Science* **299**, 371 (2003).
- 47 I. S. Lokuge and P. W. Bohn, "Voltage-tunable volume transitions in nanoscale films of poly(hydroxyethyl methacrylate) surfaces grafted onto gold," *Langmuir* **21**, 1979 (2005).
- 48 T. Tanaka, I. Nishio, S. Sun, S. Ueno-Nishio, "Collapse of gels in an electric field," *Science*, **218**, 467 (1982).
- 49 M. Annaka and T. Tanaka, "Multiple phases of polymer gels," *Nature* **355**, 430 (1992).
- 50 A. Matsuyama, "Volume phase transitions of smectic gels," *Phys. Rev. E* **79**, 051704 (2009).
- 51 W. Xue, I. W. Hamley, "Thermoreversible swelling behaviour of hydrogels based on *N*-isopropylacrylamide with a hydrophobic comonomer," *Polymer* **43**, 3069 (2002).
- 52 A. Richter, G. Paschew, S. Klatt, J. Lienig, K. Arndt, and H. P. Adler, "Review on hydrogel-based pH sensors and microsensors," *Sensors* **8**, 561(2008).
- 53 T. Tanaka, D. Fillmore, Shao-Tang Sun, I. Nishio, G. Swislow, and A. Shah. "Phase transitions in ionic gels," *Phys. Rev. Lett.* **45**, 1636 (1980).
- 54 A. Suzuki, and T. Tanaka, "Phase transition in polymer gels induced by visible light," *Nature* **346**, 345 (1990).
- 55 F. Ilmain, T. Tanaka and E. Kokufuta, "Volume transition in a gel driven by hydrogen bonding," *Nature* **349**, 400 (1991).
- 56 S. Juodkazis, N. Mukai, R. Wakaki, A. Yamaguchi, S. Matsuo and H. Misawa, "Reversible phase transitions in polymer gels induced by radiation forces," *Nature* **408**, 178 (2000).
- 57 M. Shibayama and T. Tanaka, "Small-angle neutron scattering study on weakly charged temperature sensitive polymer gels," *J. Chem. Phys.* **97**, 6842 (1992).
- 58 S. Hong and C. A. Mirkin, "A nanoplotter with both parallel and serial writing capabilities," *Science* **288**, 1808 (2000).

- ⁵⁹ C. Maedler, S. Chada, X. Cui, M. Taylor, M. Yan, and A. La Rosa, "Creation of nanopatterns by local protonation of P4VP via dip pen nanolithography," *J. Appl. Phys.* **104**, 014311 (2008).
- ⁶⁰ C. Maedler, "Applying different modes of atomic force microscopy for the manipulation and characterization of spatially localized structures and charges," Diploma Thesis for the academic degree of Diplom Physiker, Faculty of Natural Sciences Institute of Physics, Chemnitz University of Technology (2009).
- ⁶¹ M. Nonnenmacher, M. P. O'Boyle, and H. K. Wickramasinghe, "Kelvin probe force microscopy," *Appl. Phys. Lett.* **58**, 2921 (1991).
- ⁶² M. Schenk, M. Futing, R. Reichelt, "Direct visualization of the dynamic behavior of water meniscus by scanning electron microscopy," *J. Appl. Phys.* **84**, 4880 (1998).
- ⁶³ L. M. Demers, D. S. Ginger, Z. Li, S. -J. Park, S. -W Chung, C. A. Mirkin, "Direct patterning of modified oligonucleotides on metals and insulators by dip-pen nanolithography," *Science* **296**, 1836 (2002).
- ⁶⁴ S. Rozhok, P. Sun, R. Piner, M. Lieberman, and C. A. Mirkin, "AFM study of water meniscus formation between an AFM tip and NaCl substrate," *J. Phys. Chem. B* **108**, 7814 (2004).
- ⁶⁵ P. E. Sheehan, L. J. Whitman, "Thiol diffusion and the role of humidity in 'dip pen nanolithography'," *Phys. Rev. Lett.* **88**, 156104 (2002).
- ⁶⁶ Z. Zheng, M. Yang, and B. Zhang, "Reversible nanopatterning on self-assembled monolayers on gold," *J. Phys. Chem. C* **112**, 6597 (2008).
- ⁶⁷ J.-W. Jang, D. Maspoch, T. Fujigaya, and C. A. Mirkin, "A 'molecular eraser' for dip-pen nanolithography," *Small* **3**, 600 (2007).
- ⁶⁸ Y. Li, B. W. Maynor, and J. Liu, "Electrochemical AFM 'Dip-pen' nanolithography," *J. Am. Chem. Soc.* **123**, 2105 (2001).
- ⁶⁹ F. C. Simeone, C. Albonetti, and M. Cavallini, "Progress in micro-and nanopatterning via electrochemical lithography," *J. Phys. Chem. C* **113**, 18987 (2009).
- ⁷⁰ R. Maoz, E. Frydman, S. R. Cohen, and J. Sagiv, "Constructive nanolithography: Inert monolayers as patternable templates for in-situ nanofabrication of metal-semiconductor-organic surface structures-a generic approach," *Adv. Mater.* **12**, 725 (2000).
- ⁷¹ Zhikun Zheng, Menglong Yang, and Bailin Zhang, "Constructive nanolithography by chemically modified tips: Nanoelectrochemical patterning on SAMs/Au," *J. Phys. Chem. C* **114**, 19220 (2010).
- ⁷² Y. Cai and B. M. Ocko, "Electro pen nanolithography," *J. Am. Chem. Soc.* **127**, 16287 (2005).
- ⁷³ D. Chowdhury, R. Maoz, and J. Sagiv, "Wetting driven self-assembly as a new approach to template-guided fabrication of metal nanopatterns," *Nano Lett.* **7**, 1770, 2007
- ⁷⁴ A. Zeira, D. Chowdhury, S. Hoepfner, S. T. Liu, J. Berson, S. R. Cohen, R. Maoz, J. Sagiv, "Patterned organosilane monolayers as lyophobic-lyophilic guiding templates in surface self-assembly: Monolayer self-assembly versus wetting-driven self-assembly," *J. Langmuir* **25**, 13984 (2009).
- ⁷⁵ P. Yao, G. J. Schneider, J. Murakowski, and D. W. Prather, "Chemical lithography," *J. Vac. Sci. Technol. B* **24**, 2553 (2006).
- ⁷⁶ K. Shanmuganathan, J. R. Capadona, S. J. Rowan, and C. Weder, "Stimuli-responsive mechanically adaptive polymer nanocomposites," *ACS Appl. Mater. Interface* **2**, 165 (2010).
- ⁷⁷ R. Klajn, P. J. Wesson, K. J. M. Bishop, and B. A. Grzybowski, "Writing self-erasing images using metastable nanoparticle 'Inks'" *Angew. Chem. Int. Ed.* **48**, 7035 (2009).
- ⁷⁸ Sergei V. Kalinin, Nina Balk, "Local Electrochemical functionality in energy storage materials and devices by scanning probe microscopies: Status and perspectives," *Adv. Mater.* **22**, E193-E209 (2010).

-
- ⁷⁹ A. A. Tseng, A. Notargiacomo, T. P. Chen, "Nanofabrication by scanning probe microscope lithography: A review," *J. Vac. Sci. Technol. B* **23**, 877 (2005).
- ⁸⁰ A. A. Tseng, A. Notargiacomo, T. P. Chen, Y. Liu, Yuchan, "Profile uniformity of overlapped oxide dots induced by atomic force microscopy," *J. Nanosci. Nanotechnol.* **10**, 4390 (2010).
- ⁸¹ K. B. Jinesh and J.W. M. Frenken, "Capillary condensation in atomic scale friction: How water acts like a glue," *PRL* **96**, 166103 (2006)
- ⁸² P. J. Feibelman, "The first wetting layer on a solid," *Physics Today* **63**, 34 (2010).
- ⁸³ E. Kapetanakis, A.M. Douvas, D. Velessiotis, E. Makarona, P. Argitis, N. Glezos, and P. Normand, "Hybrid organic-inorganic materials for molecular proton memory devices," *Organic Electronics* **10**, 711 (2009).
- ⁸⁴ E. Kapetanakis, A. M. Douvas, D. Velessiotis, E. Makarona, P. Argitis, N. Glezos, and P. Normand, "Molecular storage elements for proton memory devices," *Adv. Mater.* **20**, 4568 (2008).
- ⁸⁵ S. M. Haile, Dane A. Boysen, C. R. Chisholm, and R. B. Merle, "Solid acids as fuel cell electrolytes," *Nature* **410**, 910 (2001).
- ⁸⁶ T. Norby, "The promise of protonics," *Nature* **410**, 877 (2001).
- ⁸⁷ P. M. Gilbert, K. L. Havenstrite, K. E. G. Magnusson, A. Sacco, N. A. Leonardi, P. Kraft, N. K. Nguyen, S. Thrun, M. P. Lutolf, and H. M. Blau, "Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture," *Science* **329**, 1078 (2010).

Perspectives in Industrial Nano-materials

Jose A Alarco

Discipline of Nano-materials
School of Chemistry, Physics and Mechanical Engineering
Science and Engineering Faculty
Queensland University of Technology
Queensland, Australia

Abstract

Extensive debate has recently risen as a result of regulation on the manufacture and use of nano-materials. Historical examples, some of them centuries old, are given that show that nano-materials are not really new. These examples are followed by a more detailed account of a case study out of Dr Alarco's professional life doing R&D in industrial nano-materials.

1. Introduction

This article describes the contents of the presentation, made remotely using Skype by Dr Jose A Alarco, for the XV Meeting of Physics, Universidad Nacional de Ingenieria (UNI), Lima, Peru held between 01-03 August, 2011. The occasion also coincided with the celebration of Professor Holger Valqui's 80th birthday.

Dr. Alarco has been a former student of Professor Valqui at the Department of Physics, Faculty of Sciences, UNI between 1982-1988. Dr Alarco learned from Professor Valqui not only General Physics, Mathematical Physics and Quantum Mechanics, but also good approaches to learning inside and outside the classroom. These teachings have accompanied Dr Alarco all through his career, a significant part of which has been focused on nano-materials within an industrial environment.

After his Bachelor in Sciences, Major in Physics, at UNI, Dr Alarco did a PhD in Materials Science at the Department of Physics, Chalmers University of Technology, Sweden. After his PhD, Dr Alarco joined The Centre for Microscopy and Microanalysis and Advanced Ceramics Development at the University of Queensland (UQ), Australia in 1994. This position exposed Dr. Alarco to a series of industrial projects spanning tailor made nano-materials for high performance applications, catalysis and wastewater remediation among others and also the development of chemical processes and automated, scalable prototype equipment for the manufacture of the relevant materials.

A few years later in 1999, Dr Alarco teamed up with Dr Peter Talbot (both of whom left UQ) to co-found Scienceworks Consultants (SWC), a highly specialized scientific consulting company. SWC subsequently teamed up with a group of local Australian investors to form Very Small Particle Company (VSPC) [1], a

small company founded in late 1999, focused on the development and manufacture of nano-scale complex metal oxides (described in more detail below).

SWC has had the role of core senior research and development (R&D) team of VSPC and has had the task of inventing all the new materials and technologies that have given VSPC its competitive edge between 2000-2011. Through their involvement in SWC/VSPC, Dr. Alarco and colleagues have undertaken collaborative development work in nano-materials towards commercialization with some of the worlds major companies and have included projects such as transparent conducting oxides (Degussa, Germany), various industrial catalysts (BASF and Sud-Chemie, Germany) and vehicle exhaust catalysts (Engelhard Corp., USA, which later became BASF Catalysts LLC, USA). A schematic of some of the areas of application of nano-scale complex metal oxides is shown in Figure 1 below.

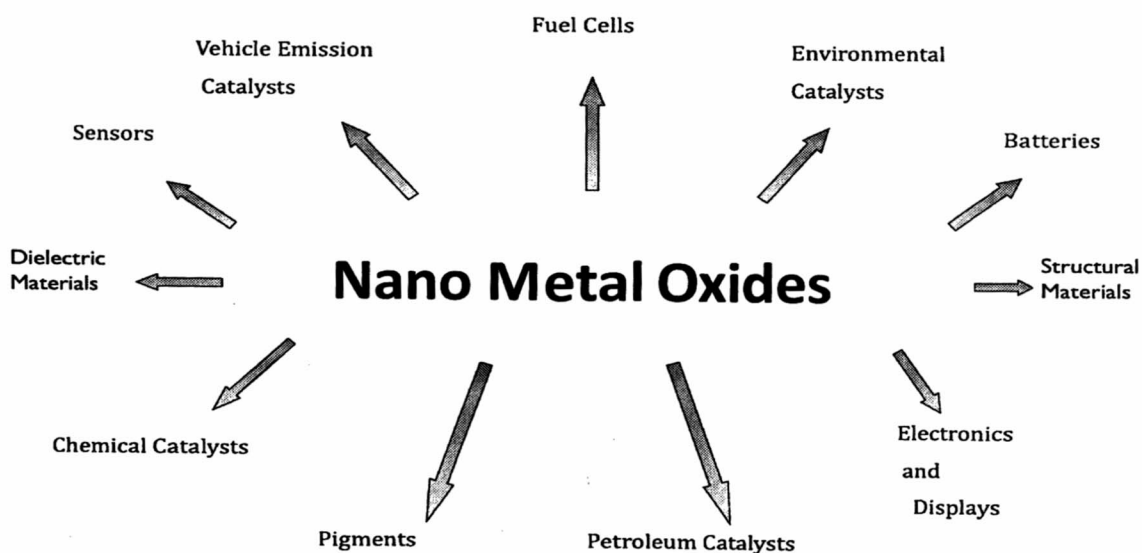


Figure 1. Schematic of various areas of industrial application of nano-scale complex metal oxides developed and explored to various extents by VSPC.

In more recent years, the focus of the commercial collaborations of VSPC has been on high-rate, high-capacity, safe battery materials for electric vehicles (in particular LiFePO_4 materials) and this has also involved interactions with major chemical companies and battery manufacturers worldwide (more information is given in the Case Study in section 4).

Nowadays, Dr Alarco is based at the School of Chemistry, Physics and Mechanical Engineering, Science and Engineering Faculty, Queensland University of Technology (QUT). There, he continues to develop new processes and applications of nano-materials for industrial applications. He is working on new materials synthesis, wastewater treatment and also developing theoretical ideas

to explain the relationship between the crystal chemistry and properties of complex metal oxides, which happen to be governed by interactions at the nano-scale [2-3].

2. Selection criteria for industrial nano-materials

Selection criteria for industrial nano-materials are essentially no different to selection criteria for any coarser versions of industrial material. Several factors have to be considered simultaneously, while the final decision on viability will be a matter of the economics (and even politics) involved. A list of some of the aspects to be considered when embarking on industrial nano-materials R&D is given below:

- Physical properties of interest (typically enhanced for nano-materials)
- Life time in application
- Toxicity
- Environmental friendliness
- Compatibility with further processing
- Simplicity
- Availability
- Cost

As mentioned on the list above, physical properties of interest are typically enhanced with nano-materials. Even when a new material can be developed that ticks all the criteria boxes, final adoption of the nano-material or process by industry has to overcome hurdles of established industry monopolies, established manufacturing practices and economies of scale. This is a complex problem per se and requires a broad perspective, open-minded approach. Industry, like any other business, is based on competition. Nano-materials have the potential to be “disruptive” for the established businesses and manufacturing practices. Therefore, most large companies will attempt, in general, in one way or another, to own or control any new technology that could destabilize some of their multi-billion dollar businesses.

On the other hand, as part of the industrial activity in new materials, Government institutions worldwide such as Environmental Protection Agencies (EPA's), Occupational Health and Safety (OH&S) Authorities, etc. are charged with monitoring and creating regulations meant to protect the workers, the environment and the consumers. Given that the history of regulations of nano-materials is not very long, the “safe” approach to the risks with new industrial nano-materials seems to be classifying them as “too dangerous until proven otherwise”. This has created currently extensive debate by many companies which have been producing nano-materials for some time now, and that argue that there are many examples of nano-materials, which are not really “new”.

The following section will provide some historical perspective of the commercial uses of nano-materials through examples (most illustrations have been obtained

from Google images). The historical examples are then followed by a more detailed case study from VSPC.

3. Historical examples of Nano-materials usage

The terminology “nano-materials” is rather new, but the materials themselves are not necessarily so in comparative terms. A few examples divided into centuries old, decades old and more recent are given below.

3.1 Centuries old examples

Centuries old examples of nano-materials certainly made use of the special material effects (provided by the nano-scale) although most likely their introducers were unaware of the actual nano-scale.

A large group of examples are connected to the uses of clay minerals, which can be considered to be a natural nano-material [4]. Bleaching fleece and clothes and removal of grease and stains are among uses of clays dating back to > 5000 BC in Cyprus and Greece. There are also records that the Romans enhanced the clays as a cleaning product with soil and decaying urine (for its ammonia content). This was apparently a well-organized commercial operation involving collection of urine from a multitude of public urinals.

A second example of a centuries old usage of clay nano-materials is in Chinese porcelain dating back to the 6th or 7th century AD. In fact, the word Kaolinite (one of the major clay minerals) is derived from the Chinese word Kao – ling meaning mountain. More recent references dating back to the 9th century AD make references to the fine ceramics made with clays, with wall thickness as small as < 0.4 mm. Some of these ceramics are as strong as modern technological ceramics and advances in knowledge and understanding of ceramics has established that the clay powder particles would have to be < 1000 nm in diameter to achieve such outstanding results.

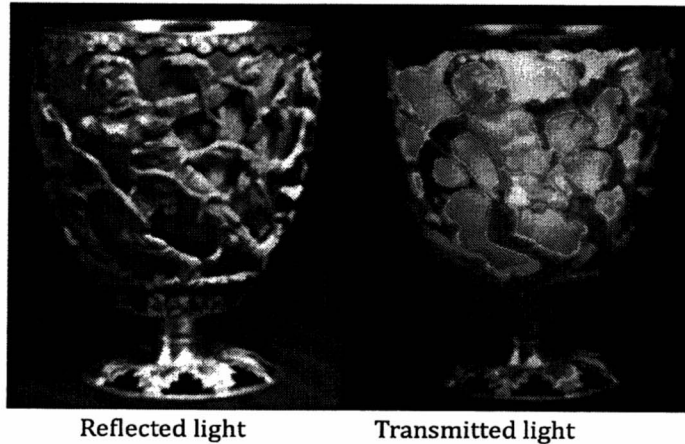
Another centuries-old use of clay minerals has been the use of pigments by various indigenous people around the world. One example is the use of red ochre in Egypt, which is a pigment made of clay naturally tinted with hydrated Fe³⁺.

A somewhat different example of centuries old usage of nano-materials has been in glass manufacturing by the Romans, a very famous cup called the Lycurgus cup dates back to the 4th century AD and it is well known for its visual effect when illuminated with either reflected or transmitted light as shown in Figure 2.

Glass

Imregnated with Gold Nanoparticles

- 4th Century AD
- Roman
- Lycurgus Cup



Reflected light

Transmitted light

Figure 2. Images of the famous Lycurgus cup as seen with reflected (left) and transmitted (right) light (source: Google images).

3.2 Decades old examples

Decades old examples of nano-materials display good awareness the nano-scale and early knowledge of the nano-size effects. It is in this category where the main industries in the middle of the debate are, which may be potentially affected by new regulations.

The first example of decades old usage of nano-materials is as reinforcing fillers for the rubber tyre industry [5], where nano-particles of amorphous silica and carbon black have been extensively used for wear and abrasion resistance in particular for tread compounds (see Figure 3, which shows the schematic of a modern tyre construction).

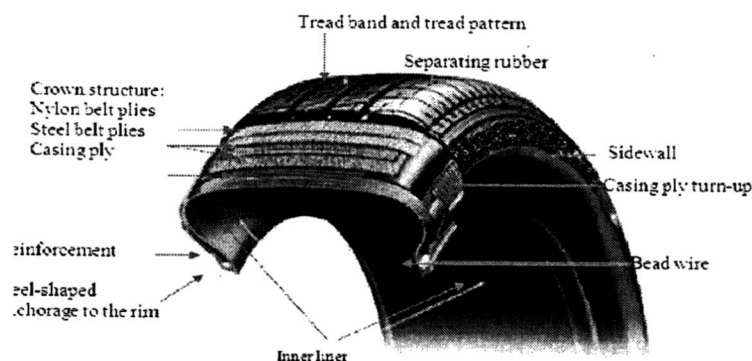


Figure 1. An example of a tyre construction

Figure 3. Schematic of tyre construction (source: Google images).

Other examples of decades old usage of nano-materials have been as colloidal nano-particles of silver and gold. The patent literature shows the first records from a manufacturer of a nano silver product in 1954. Many of the initial uses were as algacides and in water filter manufacture. Colloidal silver has also been used and continues to be sold as a diet supplement, with antiviral properties. Not helping the debate, a few cases of argyria (see Figure 4) where a concentration of silver on the skin turns it blue/gray in colour. The response of the supplement suppliers is that there are too few cases of argyria given the extensive market of the colloidal silver as a supplement, and that the argyria problems could be attributable to some people using the wrong silver formulation or form (probably produced by themselves in-house).

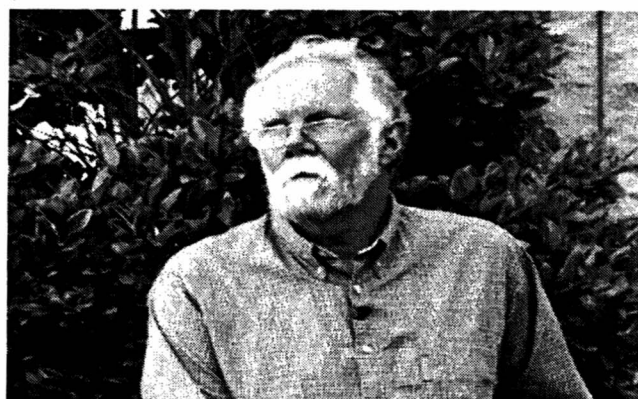


Figure 4. Picture of a person displaying argyria (source: Google images).

In a more positive light, the potential uses of colloidal gold in diagnostics and therapeutics has been described in documents dating back to their original manufacture in 1857. Extensive use of colloidal gold has taken place in cancer treatment since the 1940s. The interest in colloidal gold is related to the ability to create functionalized surfaces capable of overcoming biological barriers allowing for a very targeted drug delivery.

3.3 More recent examples (about 1 decade old or less)

The more recent examples of usage of nano-scale materials display very good awareness of the nano-scale dimensions and a very rapidly growing knowledge of nano-size effects. Recent examples of nano-materials often have higher structural and/or compositional complexity. They often constitute the basis of an “enabling technology”, where the application would not be possible without the nano-scale effects of the nano-structured material.

One example of the recent progress in nano-materials is Quantum dots [6-7]. Quantum dots emit different colours of light due to different quantum confinement effect, controlled through the characteristics of the colloidal nano-

particles. In Figure 1.2 of reference [6] (see note [8]), pictures of dispersions of hollow gold nano-particles having different diameters and wall thicknesses, responsible for the different colours, can be seen. Further explanation of the electromagnetic theory for the colour effects of nano-particles can be found in reference [9].

Another recent industrial application of nano-materials that has attracted large amounts of funding and investment is ink-printing techniques for the production of solar panels using pastes composed of nano-particles of solar material [10]. These techniques have become the most cost-competitive of the solar technologies. In a near future, with some additional enhancements costs of the panels could be brought to values that may compete with the grid.

4. Case study: Battery nano-materials at VSPC

4.1 The VSPC process

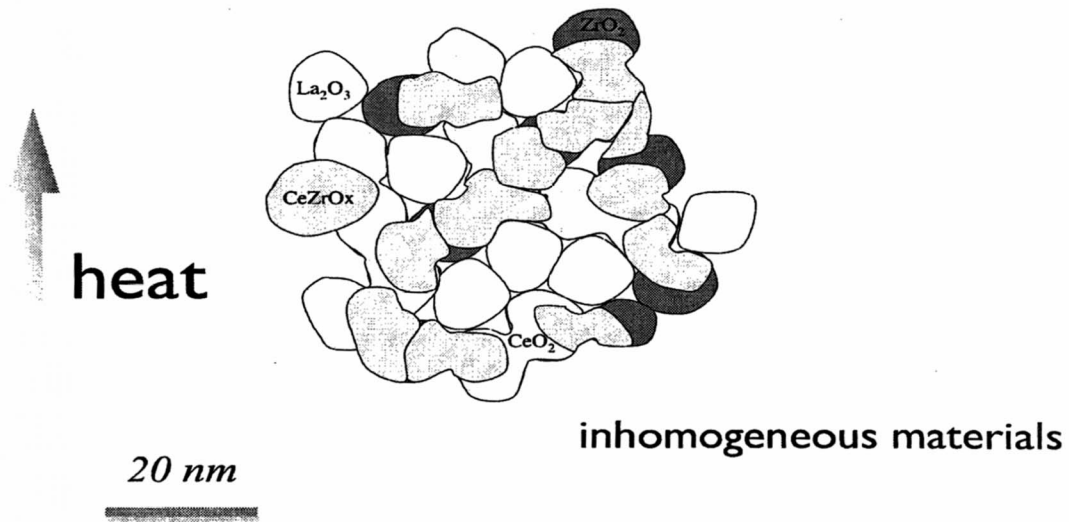
The VSPC process has been subject to worldwide patent applications [10]. Other inventions deal with extensions and refinements of the VSPC process and are also covered by worldwide patent applications (searchable in the patent literature available in the public domain (see for instance Patent Lens [11])).

VSPC invented a process for production of **nano-scale “complex” metal oxides**. By “complex”, it is understood in general more than two metal cations ending homogeneously distributed in the final single crystalline phase material (unless mixed crystalline phases are targeted for other specific purposes). Thus, by complex (in the context used by VSPC), it is meant that the nano-particles are **homogeneous in chemical and phase composition** at the end of the process.

VSPC's use of complex is not to be confused with a complex mixture of metal oxides, where mixed particles of different composition and phase can be mixed together (see schematic below) (although one may attempt to produce such a composite for other deliberate reasons).

The effects of temperature on agglomerates of nano-particles composed of different types of “complex” metal oxides are illustrated in Figure 5 below.

Complexity



Complexity

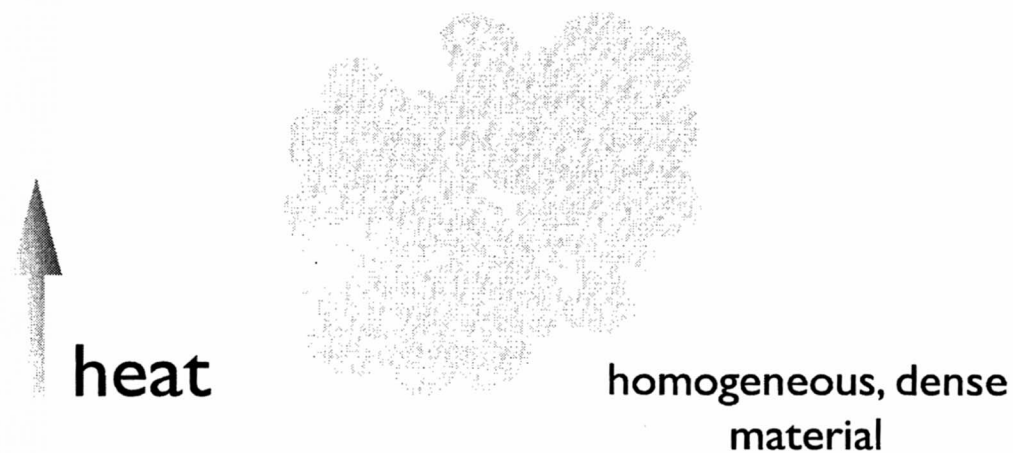


Figure 5. Schematic of the effects of heating agglomerates of nano-particles composed of different types of “complex” metal oxides (depending on the definition of “complex” being used)

An additional requirement for a practical process for the production of nano-scale complex metal oxides is the **control of the target chemical composition**. Such control has to be achieved simultaneously with the **control of**

homogeneity. In other words, the process needs to be able to define a target composition and segregation has to be avoided at a nano-scale level during the various stages of the process.

The VSPC invention is based on the OAM method, where OAM stands for Optimal Atomic-scale Mixing of cations in solution. Cations in solution are mixed with surfactant molecules, in concentration levels so that micelles and a micellar liquid are formed [12-14]. Surfactants are selected to assist in maintaining the homogeneity of the solution mix, without introducing adverse effects. The VSPC invention then attempts to maintain this homogeneity as much as possible through controlled heating and reactions (including control of the interactions with the surfactant micelles). A schematic of the process concept is shown in Figure 6 below.

VSPC's Process Concept

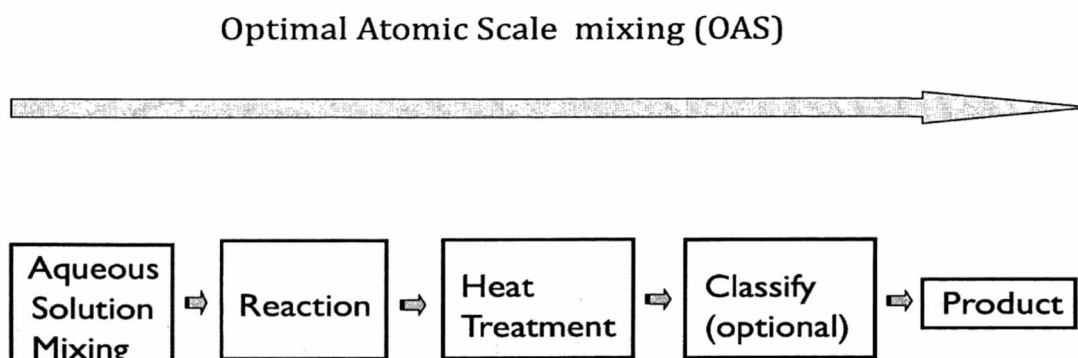


Figure 6. Schematic of the VSPC's process concept.

4.2 Case Study: Battery Materials at VSPC

A number of electric vehicles are becoming available in the market. However, wide spread adoption of electric vehicles will be slow until prices (particularly of the batteries) drop significantly and consumers develop more confidence on long battery life (or suppliers implement a warranty system that replaces them when performance has significantly dropped). Figure 7 shows a picture of the Chevy Volt electric car.



Figure 7. Picture of Chevy Volt electric car (source: Google images).

Two major factors have contributed to the upsurge of electric vehicles popularity between car manufacturers:

- i) The increased safety of new battery chemistries, and
- ii) The ability to charge and discharge batteries at a very fast rate

This combination of characteristics means that very fast discharge and recharge times are possible. “Refueling” then becomes feasible within a few minutes, provided that the fueling station can carry the currents for the large battery packs. It also means that very high power draws are possible in situations of high acceleration. All these advantages have to be available preferably with materials with high capacity. In this way, the battery packs are limited in size and weight for a decent range between recharges (generally around 200-400 Kms although more limited range is also an alternative for more specific uses). Engineered nano-materials for the electrodes have made it possible to meet these requirements of the batteries.

The battery electrochemistry requires both electronic and ionic transport. The movement of the ions inside the crystals by diffusion is the slowest of all the transport properties (see Figure 8). Ionic transport is much faster in the liquid or gel electrolyte. By limiting the distance that the ions need to migrate inside the nano-crystals, the overall rate of transport can be made to remain high. This is the basis for recent improvements in battery technology, when combined with the higher stability of the LiFePO_4 chemistry that allows the internal heating of the battery electrodes to remain within safe limits. Because electronic connection to the electrode collectors is also required and most battery materials are not very good electron conductors, this is often achieved with electron conducting coatings, which have to be thin and/or nano-porous) so as

not to block the ion paths. Therefore, the battery material is in fact a composite structure, that has to simultaneously meet a series of requirements.

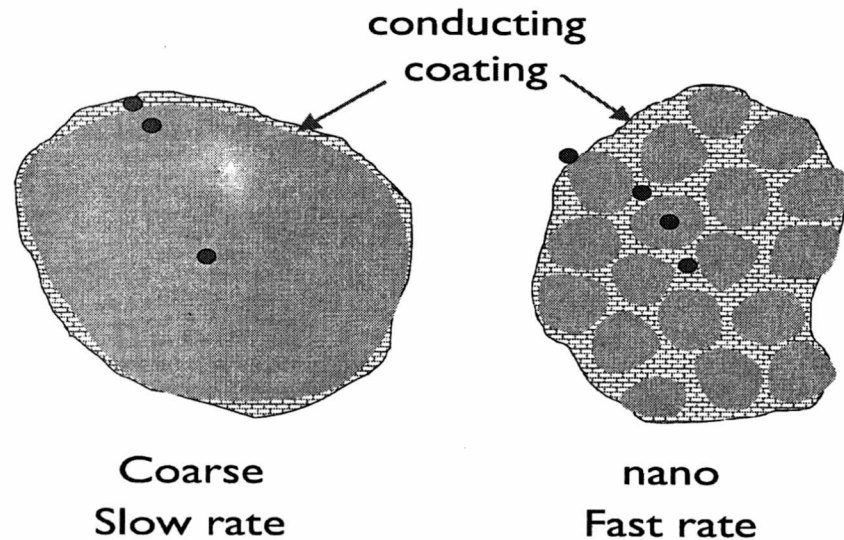


Figure 8. Schematic of the advantages of nano-materials for the charge transport processes.

A video of a technical presentation made by Dr Jose Alarco and Dr Shelley Brown-Malker (former colleague of Dr Alarco at VSPC), covering many aspects related to those in this article, can be found at the website of Engineers Australia [15].

5. Concluding remarks

Nano-materials have the potential to enable new technologies and as such they can be disruptive for established industries. The industrial environment brings new dimensions to the R&D in new nano-materials that are generally ignored or very lightly touched upon in the academic world. Nano-materials for industrial applications have to meet simultaneously a combination of selection criteria including physical properties, environmental and health regulations, scalability and economic factors among others. Not meeting some of these criteria satisfactorily can mean that the application of the nano-material never becomes viable.

A balance needs to be found between regulation and the pursuit of new knowledge through R&D in industrial nano-materials. The subject has the potential to be scientifically as well as commercially rewarding. More recent work by Dr Alarco and colleagues on the relationship between the crystal chemistry and properties of complex metal oxides [2-3], will further emphasize the relevance in these materials of interactions at the nano-scale.

References

- [1] www.vspc.com
- [2] J. A. Alarco and P. C. Talbot, "A phenomenological model for the structure-composition relationship of the high T_c cuprates based on simple chemical principles", *Physica C* 476 (2012) pp. 32-47.
- [3] J. A. Alarco and P. C. Talbot, "Lattice vibrations of quasi-periodic superstructures and superconductivity of the high T_c cuprates", in preparation.
- [4] Giora Rytwo, "Clay Minerals as an Ancient Nanotechnology: Historical Uses of Clay Organic Interactions, and Future Possible Perspectives", *macla* no 9, septiembre '08, revista de la sociedad española de mineralogía, pp. 15-17 (www.ehu.es/sem/macla_pdf/macla9/macla9_15.pdf)
- [5] G. S. Crutchley, "Nanoparticles and their possible use in rubber", www.kumhotech.co.uk
- [6] Jin Zhong Zhang, "Optical Properties and Spectroscopy of Nanomaterials", World Scientific Publishing Co., 2009.
- [7] Guozhong Cao and Ying Wang, "Nanostructures and Nanomaterials- Synthesis, Properties and Applications, 2nd edition, World Scientific Publishing Co., 2011.
Equivalent to patent US 6752979 B1 filed Nov. 21, 2000 and awarded Jun. 22, 2004.
- [8] Chapter 1 of reference [6] is downloadable for free from www.worldscibooks.com/etextbook/7093/7093_chap01.pdf
- [9] P. Mulvaney, "Not all that's gold does glitter", *MRS Bulletin* 26 (2001) pp. 1009-1014.
- [10] www.nanosolar.com
- [11] www.patentlens.net/daisy/patentlens/patentlens.html
- [12] Bo Jonsson, Bjorn Lindman, Krister Holmberg and Bengt Kronberg, "Surfactants and Polymers in Aqueous Solution", John Wiley and Sons, Inc., 1998.
- [13] Milton J. Rosen, "Surfactants and Interfacial Phenomena – Third Edition", John Wiley and Sons, Inc., 2004.
- [14] K. Robert Lange, "Surfactants – A Practical Handbook", Hanser Verlag, Munich, 1999.
- [15] <http://www.engineersaustralia.org.au/nano-engineering/publications>

Transmission of electric energy without using wires

Ener Salinas

ABB Corporate Research (Sweden)

Talk in celebration of the 80th Birthday of professor Holger Valqui, teacher and friend

In this presentation, some properties of the wireless power transfer (WPT) problem are elaborated in terms of physical barriers. There are interesting facts in this approach and its related applications. For example, it is neither necessarily true that the power must be transmitted by electromagnetic fields nor entirely correct that the transmitting medium should be vacuum or air. The inadequacy of the direct application of Faraday's induction law to provide efficient power transfer will be illustrated. Furthermore, what is necessary to add in order to produce high transfer efficiency will also be shown. The scheme will be classified according to the length of the physical barrier, namely, short (*mm*), medium (*dm*), large (*m*) and ultra large (>100 m or km). Some examples of these technologies will be presented.

1 INTRODUCTION

The demands of modern technology require more and more that some devices of practical and frequent use be powered in such a way that they become free from galvanic contact and current carrying cables. We have first to remember that the problem of transmitting very little power (*mW* or μW) to practically any distance was resolved a century ago. Furthermore, it is possible to include a signal in this type of transmission and bring it to the other side of the planet or even outside earth, to a satellite or space probe. However, the efficiency of this transmission is very poor, most of the emitted power is wasted, very little arrives and such signal would probably need to be amplified, which causes spending even more energy. Yet, this operation serves its valuable purpose i.e. transmission of information. On the other hand, transfer of a larger amount of power, such as *W* or *kW*, and recovering it with high efficiency is still a present-day challenge. This latter research area was initiated by Nikola Tesla [1], coincidentally also about a century ago. Unfortunately he left no clear indication of a final WPT device for large power, although his initial thoughts were to provide electrical energy to the world wirelessly with only one system [2].

2 THE CONCEPT OF PHYSICAL BARRIER

In order to clearly specify WPT phenomena and develop related technologies, we define the concept of *physical barrier* as follows: the concept of electric power transmission from one region of space to another without the use of conductors means that electrons should not pass between the two media (*physical barrier*). Therefore other means of transmitting such power should be used ¹ (e.g. mechanical torque, magnetic fields, vibration, light, to mention just a few). Furthermore, it is highly desirable that the maximum amount of transmitted energy is recovered after passing such barrier.

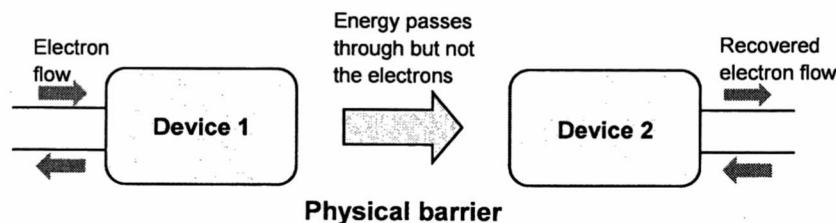


Figure 1 Wireless power transmission (WPT) through a physical barrier.

¹ The strong form of this condition rules out the use of semiconductor-based media as physical barrier since it will still be a channel for electrons. However, depending on the specific application (e.g. low voltage), a flexible condition may consider this possible.

3 CLASSIFICATION OF WPT

We can classify WPT according to power size, frequency, transmitting distance, type of technology or alternatively by its uses. When electromagnetism is directly involved in the transmission there is a heuristic correlation between distance and frequency, namely the longer the distance we require the energy to be transferred, the higher the frequency needed for the operation. Yet, this is not a rule because often in practice there are other reasons for keeping frequencies high even at very short distances. Let's first give an example of energy transfer that is not directly related to electromagnetics. Then the case will be built for the classification according to distance, considering what is necessary for efficient power transfer.

3.1 Energy transfer via an insulating shaft

Wireless power transfer does not necessarily have to involve fields in the physical barrier. An example is the energy transmission via a rotating shaft, which is a motor-generator combination. Electricity is the input and also the output type of energy, yet there are no moving electrons along the physical barrier, which in this case is the insulating shaft itself.

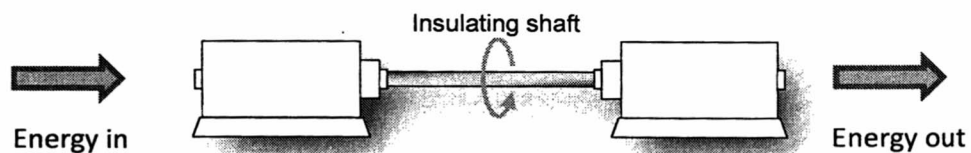


Figure 2 Energy transfer through a rotating shaft.

The advantages of this technology are: practically unlimited power, short time overload capability, high energy efficiency, conventional components and basically low cost. The disadvantages are: mechanical wear, electrical discharges in the bearings, requirement of maintenance and the positions of source and receiver are fixed.

3.2 Insufficiency of the direct application of Faraday law

One can apply directly the Faraday induction law and even design useful devices such as a wireless and battery-less system of light generation for bicycles as shown in the demonstrator in Figure 3.

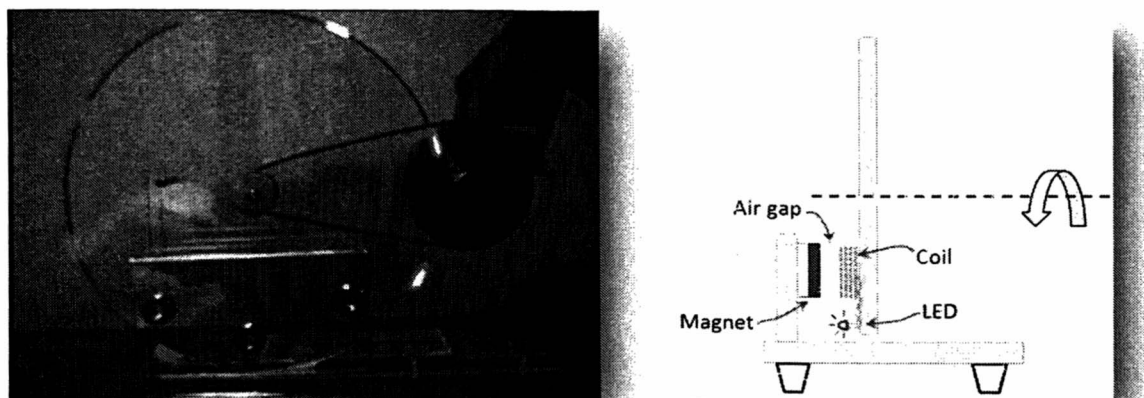


Figure 3 A demonstrator of WPT for a system of lateral lights for bicycles using the Faraday law of induction; this system is rather inefficient but it is simple and serves its purpose.

Analysing this design more carefully we realize that the efficiency of the transfer is rather low because a considerable amount of magnetic flow does not penetrate the coil surface, thus not contributing to the induction process. Having said that, it can also be added that this device serves its purpose well and its operation can be considered satisfactory, taking into account its simplicity.

Yet, in most WPT large (kW) systems, efficiency is very important. The reason is that these systems will be competing against direct plug-in power transfer which represents practically 100% efficiency.

Let us consider a system composed of two coils, a transmitting coil (TX) and a receptor coil (RX). A power source feeds the TX coil and the aim is to transfer power as efficiently as possible to the coil RX, where a load R_L is connected. Knowing already that a direct application of Faraday's law would not be enough, as shown in Figure 4 (a), there are various ways of enhancing the magnetic coupling between two coils. One way is to place a concentrating ferromagnetic medium between them as in transformers. Then, provided losses in the ferromagnetic core and windings are controlled, this operation becomes indeed very efficient, but it makes the system inflexible and not movable. Another way is to place the concentrating medium strategically around the coils but leaving an air gap. It is worth noticing a similarity between the latter and the way how a stator transfers energy to the rotor in an induction machine.

Yet there are other means of enhancing WPT by exploiting the concept of resonance. In this way, the impedance of the system can be minimized and as a result there is a moderate improvement in the magnetic coupling as shown in Figure 4 (b). However, introducing two loops each one resonating at f_0 with a maximal quality factor Q_{max} gives strong coupling by induction and magnetic resonance combined (SCMR) [3],[4], as in Figure 4 (c). The comparison of the three cases is given in Table 1.

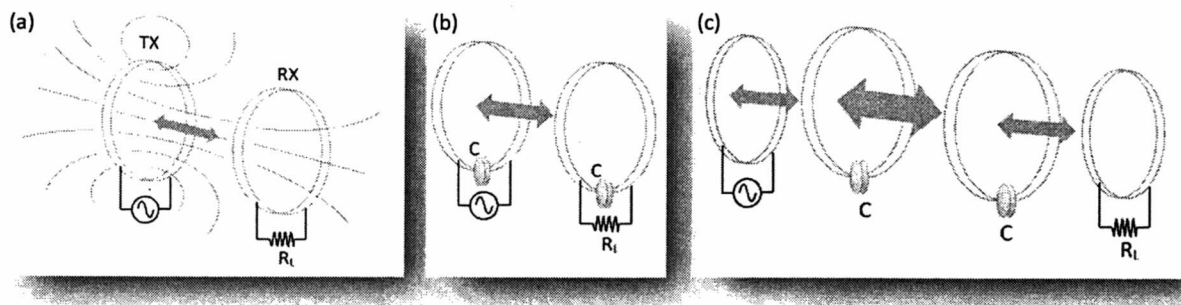


Figure 4 Improving WPT of inductive coupling (a) between two coils; making it inductive resonant coupled (b) or strongly coupled magnetic resonance SCMR (c).

Table 1 Improvement of inductive coupling by resonance and strongly coupled magnetic resonance.

Coupling type	Inductive coupling (IC)	Inductive resonant coupling (IRC)	Strongly coupled magnetic resonance (SCMR)
Frequency of operation	f_0	$f_0 = f_r = 1/2\pi\sqrt{LC}$	$f_0 = f(Q_{max}) = f_r = 1/2\pi\sqrt{LC}$
Efficiency	Low	Modest	High
Comment	Pure Faraday law, no resonance involved	Capacitors are introduced and RLC resonance is attained. Additional improvements are possible	Two RLC loops each one resonating at f_0 with a maximal Q -factor gives strong coupling by induction and magnetic resonance

4 WPT FOR VERY SHORT DISTANCES (MILLIMETRE RANGE)

This is also called “electric toothbrush” technology, because earlier designs of inductive coupling technologies were successfully applied for recharging electric toothbrushes. In the example of Figure 5, the energy is transferred by the simple inductive coupling of Figure 4 (a), or alternatively by

resonant type, as in Figure 4 (b), through an insulating plastic protection. The initial power of the socket is converted to low voltage/high current and high frequency (around 20 kHz). Sometimes it is convenient to add a ferromagnetic coupler to enhance the transfer as shown in Figure 5-left.

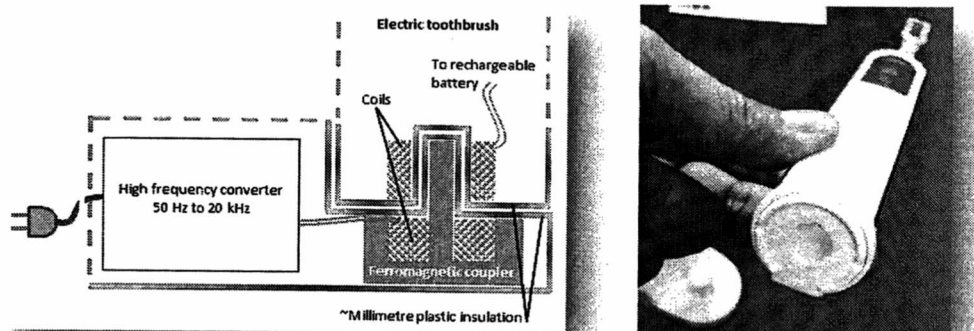


Figure 5 Design of an electric toothbrush charger (left) that uses WPT by resonant inductive coupling to transmit the required power through the waterproof plastic cover (right).

Another millimetre range WPT system is the base station charger for multiple electronic appliances such as mobile phones, digital cameras and pads. The electronic design here becomes much more sophisticated as it involves energy control and a manifold of coils interconnected to compensate misalignments which tend to drastically deter transfer efficiency. Some modifications are required in each device that has to be recharged. Fortunately, the added components can be flattened and miniaturized producing no major impact on the shape and weight of the portable device [5-7]. This technology is soon reaching global standards so customers can be guaranteed device compatibility anywhere in the world. This is being achieved by the group Qi (wireless power consortium) [8].

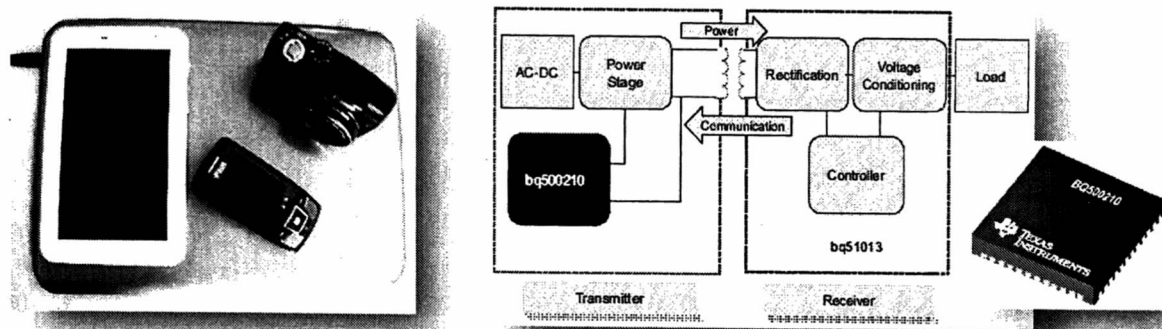


Figure 6 A base station platform for charging multiple devices using WPT (left) and a commercial WPT controller TI-bq500210 (right).

5 WPT FOR DECIMETER RANGE

The most common use for decimeter range distance WPT is for battery charging of electric vehicles. The idea here is to transfer the necessary amount of power from a transmitter coil in the ground to a receiver on the bottom of an electric or hybrid vehicle. The separation is usually between 10-30 cm depending on the vehicle type and model, while the transmitted power ranges from 3kW (small cars) to 60 kW (buses). It is called inductive power transfer (IPT) and uses the second type of coupling in Table 1. Although, at first glance, it seems that IPT choice is not as efficient as the SCMR, source designs for the second method (resonant IC) require much lower frequency than SCMR and this is relevant when high power devices are involved, for it is not easy to get or design large power (several kW) sources at MHz frequency range. Thus to improve efficiency some actions are needed:

- Trap the leaking magnetic flux by using ferromagnetic concentrators
- Use conductive shielding plates to further “bounce” the field lines to the resonant system
- Use Litz wire (dedicated twisted bundle of thin insulated parallel wires) to reduce losses due to skin effect

The most common frequency in these designs is 20 kHz, which coincidentally is also the frequency used in induction cooking stoves. Other designs use 40 kHz and up to 100 kHz especially when larger air gaps are involved [9-11]. Figure 7 illustrates various stages involved in IPT technology for a small electric vehicle.

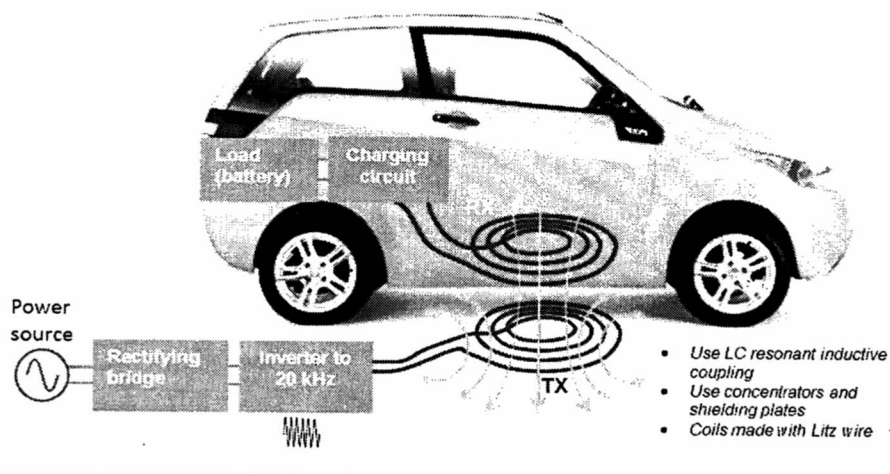


Figure 7 Inductive power transfer (IPT) of 3kW, for a small electric vehicle, some actions are listed to achieve high efficiency; over 95% coil to coil and over 90% total (source to load).

Figure 8 shows an implementation diagram in our lab for a primary IPT system imbedded in carbon fiber reinforced concrete. The 20 kHz, 3kW high frequency power source will be located in the rectangular space at the left of the figure. The screws are made of fiber glass.

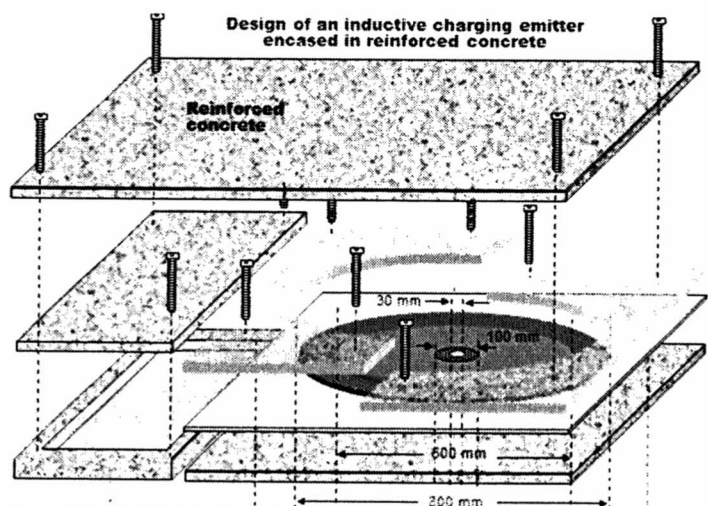


Figure 8 The primary coil (in orange) and the source of the 20 kHz, 3kW IPT (to be placed on the left) are imbedded in reinforced concrete. A ferrite concentrator and aluminum shield are part of the system.

One of the major issues concerning the implementation of IPT is the possibility of adverse effects of this type of magnetic field on human health. Therefore it is important to know what is permissible or advised by the expert organizations. According to the International Commission on Non-Ionizing Radiation Protection (ICNIRP) guidelines of 2010 [13], the emissions in the range of 3 kHz to 10 MHz should not exceed 27 μT for general public and 100 μT for occupational exposure (e.g. electric workers). In Figure 8 the range of magnetic fields according to the guidelines is shown up to 100 kHz. In the same graphic are also displayed permissible values at lower frequencies, such as the ones emitted by most common power components at 50 Hz. In order to obtain values in Figure 9, the commission of experts (among them physicists, biologists, engineers and medical doctors) gathers regularly and examines the latest results of experiments and numerical simulations and accordingly agrees on the most suitable values [14-16]. Thus, when designing the IPT system it is important to ensure that the area outside the space between coils follows the ICNIRP recommendations.

A second issue concerns pets (e.g. cats or small dogs) that attracted by the lukewarm surroundings between the two active coils may go under the car and try to rest or sleep there. To deal with this issue, it is possible to implement a weight detector that uses strain gages and switches off the system when the detected weight exceeds a pre-defined limit.

Another issue is the possibility of dropped metallic objects, such as coins and keys, since they would be affected by inductive heating due to the high frequency of the magnetic field incident on the metal. If someone tries to pick up these objects from the floor they could be very hot and may produce injuries. There are two considerations here, the first being that the lower the frequency, the lower the induced heating; which means that it is more convenient to apply an IPT design with 20 kHz rather than 40 kHz or 100 kHz. The second consideration is, especially for larger metallic objects, that the presence of a metal on top of a coil changes the self-inductance of the coil as well as the mutual inductance of the system; so with dedicated electronic controls it should be possible to determine this change and temporarily shut down the system until the object is removed.

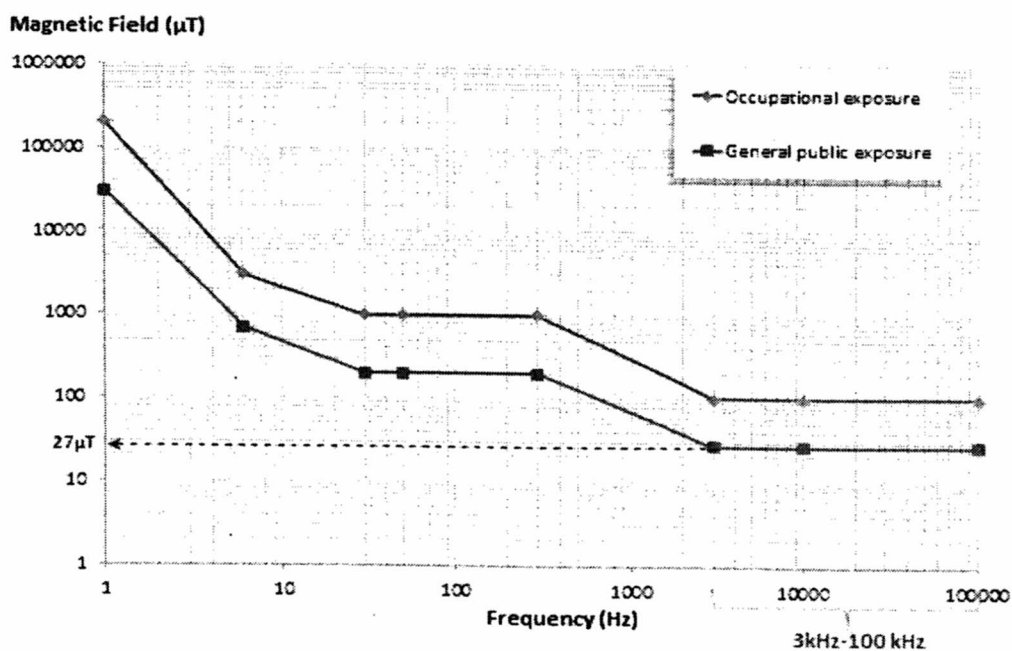


Figure 9 Permissible magnetic field values along various frequency ranges according to the 2010 ICNIRP guidelines.

Other applications of IPT at decimeter range are busses, trains and heavy vehicles. Recently, the Korean Advanced Institute of Science and Technology (KAIST) has introduced a battery powered bus called *On-Line Electric Vehicle* (OLEV) [17], it also works by improved resonant inductive coupling (Table 1 second case). The frequency is 20 kHz and the transfer power is 60 kW, indeed much larger than for an electric personal vehicle. But maybe the major difference when compared to the previously described technology for small vehicles is that the power is transfer along the road where there is a 200A flowing through power cables, just below the pavement. The field from these cables is concentrated by ferromagnetic material under the cables and then picked up by a resonant coil under the bus. The current development of this technology gives an efficiency of 80%. An analogous development is currently made by Bombardier and is called PRIMOVE. It will run in the city of Braunschweig, Germany. Bombardier has also installed a pilot tram in the Augsburg tram network in Germany [18].

6 WPT FOR METER RANGE

This is a very useful yet difficult transmission range because the efficiency is expected to be much lower than for shorter distances. Other aspect is related to in-house devices, as the use of this technology will not be easy to shield or confine as the field “beams” may be on the way of walking humans. Yet, recent developments by the MIT group [19] show that, using strongly coupled magnetic resonance (third method in Table 1), a middle range transfer distance of about 2m was achieved at 9.9 MHz and the transfer efficiency was about 40%, obtaining 60 W at the load consisting in a light bulb. In fact the actual total efficiency wall-to-load was much less (15%) because the initial power for the source was 400W. One relevant fact in this experiment is that although the frequency was considerably high, the operation used nonradiative energy transfer [4]. Furthermore, the type of magnetic field existing between the two coils TX and RX, shown in Figure 4 (c), comply in this case with the permissible values of the ICNIRP guidelines [13]. Other technologies can be used for trying to transfer energy in this distance range but they would involve radiative transfer.

7 WPT FOR EVEN LARGER DISTANCES (KM)

Microwave technology suggests a possible way of transmitting power at even greater distances. In 1975, W. Brown performed the largest microwave power transfer demonstration at the Venus Site of JPL Goldstone Facility. He used a parabolic antenna of disk diameter $D = 26\text{ m}$ at 2.4 GHz. The initial microwave power from the klystron source was 450 kW. The beam was captured by a rectenna (i.e. an array of antennas and rectifiers) obtaining 30 kW rectified power at 82.5% rectifying efficiency and the transmitting distance was 1.4 km. Figure 10 (left) shows this experiment.

An even larger microwave WPT system is planned by Japan. It is called solar power satellite (SPS) which is an array of satellites and solar panels in space that converts solar energy to microwave power and is sent to ground to large rectennas [20-22]. A smaller demonstrator of this concept has been built in the lab at Kyoto University in Japan. The size of the device is 2 m x 2.3m x 2.8 m; the array of halogen lamps consists of 133 lamps of 75W each one. The solar cells output is >166W. The solid-state microwave transmitter has a frequency of 5.77 GHz and 25W output and the Active Phased Array Antenna has 10 x 10 elements, the demonstrator is shown in Figure 10 (right).

Laser is also an option for ultra large distance transmission of power. Although in the 80s the development of high power lasers was classified information as part of the US Strategic Defense Initiative, new research at NASA has tested laser technology for WPT. Dryden Flight Research Center and the University of Alabama in Huntsville have recently demonstrated a small-scale aircraft that flies just by means of propulsive power from an invisible, ground-based infrared laser beam. This technology is flexible and the power is transmitted along a thin laser beam [23-24]. It is an emergent PT technology and seems to work very well for low power. Although not much is known about its characteristics at higher power it is also considered a potential area (along with microwaves) for applications in space travel and planetary exploration.

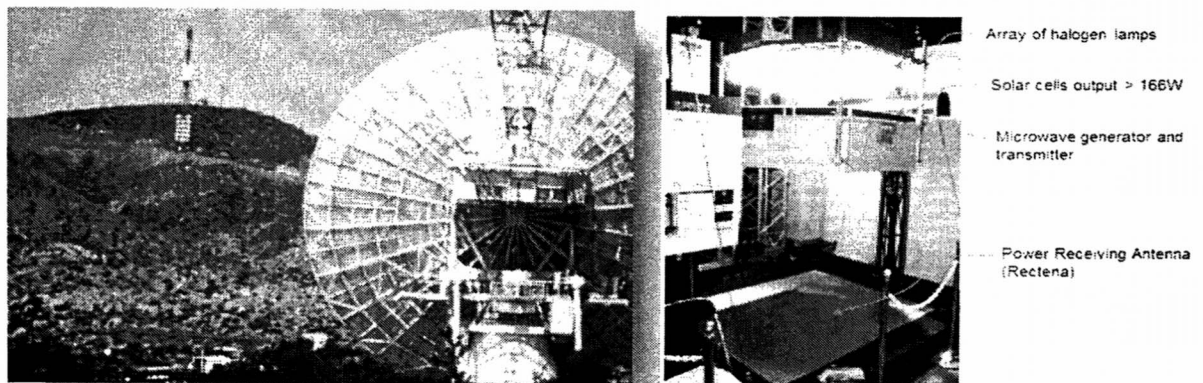


Figure 10 Power transfer using microwaves. Left: 26 m disk and a rectenna receiver at 1.4 km distance [21]; Right: smaller version of the project SPS [22].

8 CONCLUSIONS

This article has presented a view of wireless power transfer (WPT) according to transfer distance. From this approach it is clear that the number of applications of WPT is vast. Yet, there are other applications not treated here which are currently subject of intense investigation, for example: sensors, WPT for powering robots, WPT for keeping continuous operation of tiny medical devices implanted in the human body, medicament release micro systems and moving inspection devices, to mention just a few. As mobile/portable devices are becoming indispensable in practically every moment of our lives; in a near future we will see more improvements of these technologies in relation to WPT. Finally, it is also important to consider the possibility of adverse human health effects due to leaking fields from these devices as well as provide mitigating solutions such that they become integrated into the final prototypes.

9 REFERENCES

- [1] N. Tesla, "*Apparatus for transmitting electrical energy*", US patent number 1,119,732. 1914.
- [2] N. Tesla, "*The Wireless Tesla*" a collection of 13 of his essays by Wilder Publications, 2007, also in Google books.
- [3] Dionigi, M. and Mongiardo, M. "*A novel resonator for simultaneous Wireless Power Transfer and Near Field Magnetic Communications*", Microwave Symposium Digest (MTT), IEEE MTT-S International, 2012.
- [4] André Kurs, Aristeidis Karalis, Robert Moffatt, J. D. Joannopoulos, Peter Fisher, Marin Soljacic, '*Wireless Power Transfer via Strongly Coupled Magnetic Resonances*' SCIENCE, Volume 317, 6 July 2007.
- [5] S. Y. R. Hui, and Wing. W. C. Ho, "*A New generation of Universal Contactless Battery Charging Platform for Portable Consumer Electronic Equipment*," IEEE Transactions on Power Electronics, Vol. 20, No. 3, pp. 620-627, May 2005.
- [6] X. Liu, and S. Y. R. Hui, "*Optimal design of a hybrid winding structure for planar contactless battery charging platform*", in Proc. IEEE IAS'06, Oct. 2006, pp. 2568-2575.
- [7] S. C. Tang, S. Y. Hui, and Henry Shu-Hung Chung, "*Coreless Planar Printed-Circuit-Board (PCB) Transformers-A Fundamental Concept for Signal and Energy Transfer*," IEEE Transactions on Power Electronics, Vol. 15, No. 5, pp. 931-941, September 2000.

- [8] <http://www.wirelesspowerconsortium.com/what-we-do/qi/>
- [9] Mecke, R. et al. "*Analysis of inductive energy transmission systems with large air gap at high frequencies*". European Conference on Power Electronics and Applications, Toulouse 2003, Proceedings.
- [10] R. Mecke, and C. Rathge, "*High Frequency inverter for contactless energy transmission over large air gap*" 35th Annual IEEE Power Electronics Specialists Conference, pp. 1737-1743, 2004.
- [11] Grant A. Covic, John T. Boys, Michael L. G. and Howard G. Lu "A Three-Phase Inductive Power Transfer System for Roadway-Powered Vehicles" IEEE Transactions on Industrial Electronics, Vol. 54, NO. 6, December 2007.
- [12] H. Takanashi, Y. Sato, Y. Kaneko, S. Abe, and T. Yasuda, "*A large air gap 3 kW wireless power transfer system for electric vehicles*", in proceedings of IEEE ECCE 2012, 2012, pp. 269-274.
- [13] International Commission on Non-Ionizing Radiation Protection (ICNIRP), "*Guidelines for limiting exposure to time varying electric, magnetic, and electromagnetic fields*", 2010.
- [14] B. V. Dawson, I. G. C. Robertson, W. R. Wilson, L. J. Zwi, J. T. Boys, A. W. Green, "*Evaluation of potential health effects of 10 kHz magnetic fields: A rodent reproductive study*". Bioelectromagnetics Volume 19, Issue 3 , Pages 162 – 171, 1998 Wiley-Liss, Inc.
- [15] Robertson IG, Wilson WR, Dawson BV, Zwi LJ, Green AW, Boys JT, "*Evaluation of potential health effects of 10 kHz magnetic fields: a short-term mouse toxicology study*". Bioelectromagnetics. 1996; 17(2):111-22.
- [16] Boorman GA, Rafferty CN, Ward JM, Sills RC. "*Leukemia and lymphoma incidence in rodents exposed to low-frequency magnetic fields*". Radiat Res. 2000 May;153(5 Pt 2):627-36.
- [17] Seungyoung Ahn and Joungho Kim "*Magnetic Field Design for High Efficient and Low EMF Wireless Power Transfer in On-Line Electric Vehicle*" EuCAP 2011 convened papers.
- [18] www.primovecity.bombardier.com
- [19] Aristeidis Karalis, J.D. Joannopoulos, Marin Soljacic "*Efficient wireless non-radiative mid-range energy transfer*", Annals of Physics 323 (2008), pp. 34–48.
- [20] Glaser, Peter E. (22 November 1968). "*Power from the Sun: Its Future*" (PDF). Science Magazine 162 (3856): 857–861.
- [21] Brown., W. C. (September 1984). "*The History of Power Transmission by Radio Waves*". IEEE Transactions on Microwave Theory and Techniques 32 (Volume: 32, Issue: 9, pp.1230 – 1242.
- [22] H. Matsumoto, "*Research on solar power station and microwave power transmission in Japan, Review and perspectives*" IEEE Microwave Mag., pp. 36-45, Dec 2002.
- [23] Dickinson, R M. and Grey J. "*Lasers for Wireless Power Transmission*" Jet Propulsion Laboratory and NASA Report, Jan 29, 1999.
- [24] Leopold Summerer and Oisin Purcell, "*Concepts for wireless energy transmission via laser*", ESA Advanced Concept Team, Report, 2009.

Fission-track dating applied to Peruvian volcanic glasses

C.A. Tello^{a,1}, E.A.C. Curvo^b, P.J. Iunes^{c,2}, C.J. Soares^d, I. Alencar^c, S. Guedes^c,
W.M. Nakasuga^a, R.S. Resende^a, A.M. Osório^a, G. Bigazzi^e, A. Zúñiga^f, P.A. Figueroa^g, J.C.
Hadler^c

^a Departamento de Física, Química e Biologia, UNESP, CEP 19060-900, Presidente Prudente-SP, Brazil

^b Instituto de Física, UFMT, CEP 78060-900, Cuiabá-MT, Brazil

^c Instituto de Física "Gleb Wataghin", UNICAMP, CEP 13083-970, Campinas-SP, Brazil

^d Instituto de Geociências e Ciências Exatas, UNESP, CEP 13506-900, Rio Claro-SP, Brazil

^e Istituto di Geoscienze e Georisorse, CNR, 56100, Pisa, Italy

^f Instituto Peruano de Energía Nuclear, Lima, Peru

^g Universidad Nacional San Agustín, Arequipa, Peru

Abstract

Fission-Track Dating (FTD) has been applied to six Peruvian volcanic glasses. These samples presented different ages and optical characteristics. One of them (Sara Sara) had inclusions, which could be mistaken for tracks, presenting also a poor surface quality after polishing. For that reasons, this sample could not be dated. For the others samples, two independent FTD determinations were performed. One of these determinations was carried out irradiating unknown age samples together with the age standard Jankov Moldavite. The apparent ages were corrected using the size-correction method. The other determination was based on the absolute dosimetry. This time, apparent ages were corrected using the plateau-age technique. In the first case the dated samples were Macusanite, Purupurini, Batan Orq'ó, Rajchi and Firura. In the second one they were Batan Orq'ós, Rajchi and Firura. The ages determined for Macusanite, Rajchi and Firura were: (4.98 ± 0.57) Ma; (1.186 ± 0.080) Ma; and (1.065 ± 0.070) Ma, respectively. The values showed above for Rajchi and Firura are the mean values obtained from both determinations. In the case of the sample Purupurini only the apparent age, (0.0275 ± 0.0094) Ma, was obtained due to the low density of fossil tracks. Concerning the sample Batan Orq'ó, a statistical procedure was developed in order to carry out FTD even if spontaneous tracks were missing. This statistical procedure was also compared with the method proposed by Galbraith [1]. Accordingly, an upper limit of 3430 years was found for this sample. This relatively small age's upper limit, added to the fact that the sample had been collected in an important archaeological site, suggests that the sample Batan Orq'ó could have been part of an artefact that experienced a heating during its use, which would have completely erased the pre-existing tracks. The comparison between the ages of the samples Macusanite, Purupurini, Rajchi and Firura, obtained both through the size-correction method plus the Jankov Moldavite and the plateau-age technique plus absolute dosimetry, shows that the responses provided by both methods are compatible.

¹Corresponding author. E-mail address: tello@fct.unesp.br

²*In memoriam.*

1 Introduction

One of the most remarkable characteristics of the Andean Cordillera is its past and its recent volcanism [2]. Some of the by-products of this volcanic activity are the vitreous materials or volcanic glasses that can present themselves as obsidian blocks or as small grains in volcanic ashes. Transparency and the existence of inclusions are the features which will assert the viability for them to be dated via the Fission-Track Dating (FTD) method. This work aimed at applying the FTD to six Peruvian glasses samples: Sara Sara, Purupurini, Rajchi, Firura, Macusanite and Batan Orq'ó. Table 1 presents characteristics such as granulometry, transparency and localization of the samples.

For the samples Batan Orq'ó, Rajchi and Firura the dating was performed through two independent methods: comparison with the age standard Jankov Moldavite along with the size-correction method, and the absolute dosimetry method along with the plateau-age correction technique [3]. This makes possible also the comparison of both methodologies.

In the most part of the cases, little is known about the studied volcanic glasses. The exceptions are the samples of Macusanite and Purupurini. For the sample Macusanite, ages are known through different dating techniques. In the case of the FTD via plateau-age technique, the ages range from 4.3 to 7.9 Ma [4–9]. The K-Ar dating performed by Barnes *et al.* [10] provided an age of (4.2 ± 1.5) Ma, while the one performed by Poupeau *et al.* [7] resulted in (5.59 ± 0.09) Ma. The Rb-Sr dating performed by Pichavant *et al.* [11] resulted in an age of 4.7 Ma. For the sample Purupurini, the age obtained via FTD with the plateau-age correction technique was (0.053 ± 0.008) Ma [9]. For the other glasses addressed in this work, no dating data was found in the literature.

FTD is based on the accumulation of tracks originated through the passage of fission fragments from the spontaneous fission of ^{238}U . By measurements of surface etched fossil track densities in the samples and by the knowledge of its uranium content, it is possible to obtain the fission-track ages [12–17]. Fission-tracks are also sensitive to thermal treatments [18,19], which makes the track registration a reversible process. Such process is known as annealing. Since the annealing process decreases the latent track size, it affects track revelation/counting efficiency. Therefore, the revelation/counting efficiency of the fossil tracks is smaller than the induced tracks one. The difference between the efficiencies (fossil against induced tracks) leads to an (apparent) age smaller than the age when the tracks started to be registered (corrected age). To correct ages, two methodologies

Table 1: *Physical characteristics and localization data for the six Peruvian Glasses studied.*

Sample	Granulometry	Transparency	Latitude Longitude	Site
Sara Sara	small grain	bad	15°20' S 73°30' W	Nevado Sara Sara
Purupurini	small grain	good	16°00' S 71°20' W	Puru Puru city
Rajchi	block	good	14°00' S 71°20' W	Rajchi archaeological complex
Firura	block	good	15°10' S 72°40' W	Nevado Firura
Macusanite	block	very good	14°00' S 70°30' W	Macusani volcanic field
Batan Orq'ós	block	very good	13°40' S 71°40' W	Archaeological site of Batan Orq'ós

were developed: the size-correction method [20]; and the plateau-age correction technique [21]. These methodologies can be described as follow:

- Size-correction method: The correction through the size of the tracks is achieved by means of a curve which relates the density reduction of etched tracks (ρ/ρ_0) and their complementary diameter reduction (D/D_0), where ρ (ρ_0) is the density of annealed (non-annealed) tracks and D (D_0) is the annealed (non-annealed) track diameter. The (ρ/ρ_0 versus D/D_0) curve is obtained through laboratory heating experiments of samples bearing neutron induced tracks. It allows the correction of the fossil track density and consequently the age of the sample.
- Plateau-age technique: Tracks with a previous annealing history are more resilient to annealing performed at the laboratory than tracks which have not experienced it (induced tracks). The age correction via the plateau method exploits this feature. In order to obtain the plateau corrected age, it is necessary to divide the sample in two aliquots, one containing the fossil tracks (ρ_S) and the other containing only the induced tracks (ρ_I). Both fractions are then simultaneously submitted to thermal treatments. When the densities ratio (ρ_S/ρ_I) reaches unity (also $D_S/D_I \approx 1$), it means that fossil and induced tracks have the same annealing degree. At this point, the revelation/counting efficiencies of both aliquots are equal.

2 Experimental Procedure

Aiming at obtaining the ages through the Jankov Moldavite age standard (a tektite from southern Bohemia, Czech Republic [22]), samples (Macusanite, Purupurini, Batan Orq'ós, Rajchi and Firura) were separated in two aliquots. One of them was used for the fossil track density measurement, while the other was irradiated with thermal neutrons together with the age standard (Jankov Moldavite). As the neutron irradiation parameters are the same for the unknown age samples and the standard, it is not necessary to determine the neutron fluence. In this case, the unknown ages depend only on the standard age, (14.34 ± 0.08) Ma [23], and on the track densities (spontaneous and induced) for both unknown age samples and standard, according to:

$$T_{Ap} = T_{Std} \frac{(\rho_S/\rho_I)}{(\rho_S/\rho_I)_{Std}} \quad (2.1)$$

Both aliquots (spontaneous and induced) were mounted in epoxy resin and sequentially polished with diamond pastes of 6, 3 and $1\mu\text{m}$. The correspondent aliquots were simultaneously etched with fluoride acid (HF 24% wt) at 15°C during 4 minutes. Measurements of track densities and diameters were performed with a Dialux 20 EB Leica optical microscope at $500\times$ nominal magnification. The age correction curve was obtained using the experimental data from Sandhu and Westgate [24], where a relation ρ/ρ_0 versus D/D_0 is presented.

With the purpose of determining the ages through the absolute dosimetry approach, the samples Batan Orq'ó, Rajchi and Firura were also separated into two aliquots. One of them was pre-annealed, in order to erase all its spontaneous tracks, and then was irradiated with thermal neutrons. The procedures of mounting, polishing, etching and counting were the same as described previously. The absolute thermal neutron fluence determination was performed through IRMM-540 standard glasses calibrated against thin films of natural uranium [3, 25]. Usually the correction through the plateau technique is performed by using three heating points, in order to indicate that the plateau region was reached. However, this could be done with only one heat of 220°C for 4 hours as presented by Bigazzi *et al.* [9, 26].

The constant's values employed for age determination were: $\lambda = 1.55125 \times 10^{-10} \text{ a}^{-1}$ [27]; $\lambda_F = 8.46 \times 10^{-17} \text{ a}^{-1}$ [28], value compatible with Holden and Hoffman [29] and Guedes *et al.* [30–32]; $\sigma = 5.86 \times 10^{-22} \text{ cm}^2$ [33]; and $\eta = {}^{238}\text{U}/{}^{235}\text{U} = 137.88$ [34].

3 Results and Discussion

Table 2 shows the values of apparent age, T_{Ap} , and the corrected ages, T_{corr} , obtained via the size correction method plus the Jankov Moldavite as the standard age. An experimental error of 10% was assumed in the correction curve. This table also presents the counted number of fields (n_S, n_{S+I}) and tracks (N_S, N_{S+I}), the track densities (ρ_S, ρ_{S+I}) and the spontaneous to induced track diameter ratio D_S/D_I . The indexes S and S+I refer to samples with spontaneous tracks (non-irradiated) and samples with spontaneous and induced tracks (irradiated), respectively. The samples were irradiated at the position 14B prat.5 (Cd ratio 4.5 for Au) from reactor facility of IPEN/CNEN in São Paulo-SP, Brazil at a nominal fluence of 1×10^{15} neutrons/cm².

The Sara Sara sample could not be dated due to inclusions which could be mistaken for tracks. These inclusions were intrinsic to the sample, so they could not be removed by polishing no matter how carefully it was performed.

The corrected age for the Macusanite sample presented in Table 2, $(4.98 \pm 0.57) \text{ Ma}$, is in agreement with the one presented by Bigazzi *et al.* [9], obtained via the plateau-age technique plus absolute dosimetry, whose mean value is $(4.64 \pm 0.17) \text{ Ma}$ (determined at Campinas) and $(5.20 \pm 0.11) \text{ Ma}$ (determined at Pisa).

In the case of the Purupurini sample the low density of fossil tracks made it difficult to obtain the ratio D_S/D_I , what forbade the obtainment of the sample's corrected age. Bigazzi *et al.* [9] carried out a dating of this sample through the plateau-age technique plus absolute dosimetry. In his work the sample's obtained apparent age was $(0.050 \pm 0.011) \text{ Ma}$, in agreement with the age obtained in this work $(0.03 \pm 0.01) \text{ Ma}$. The weighted mean value between Bigazzi's result and the result shown in Table 2 is $(0.037 \pm 0.007) \text{ Ma}$.

For the sample Batan Orq'ó, where no fossil tracks were found, two statistical procedures were applied in order to achieve an upper limit for the age. One of the procedures is that proposed by Galbraith (p. 50 in Reference [1]) for small (<5) or zero counts. Galbraith's method calculates upper and lower confidence limits for ρ_S/ρ_I by adapting standard confidence intervals for a binomial parameter (θ) through the equation: $\rho_S/\rho_I = \theta/(1 - \theta)$. By using the Galbraith proposed statistics and the data from Table 2, the age upper limit, at a 95% confidence level, was ≈ 6000 years. Another possibility to assess the age upper limit is based on the fact that the decay process obeys the Poisson statistics. The Poissonian probability of finding "zero tracks" is $e^{-\mu}$, where μ is the average of the distribution. This probability becomes less than 5% when the average becomes higher than 3. In other words, when more than 3 tracks had been found, the fossil density of tracks (ρ_S) would be $> 3/0.131 \text{ cm}^2$ (number of tracks divided by observed area), or $\approx 23 \text{ cm}^{-2}$. By applying Equation 2.1, an age of ≈ 9000 years is obtained. So there is a 95% chance that the sample is less than 9000 years in age. Both estimates for the Batan Orq'ó age upper limit are presented in Table 2.

In Table 3 the results obtained via the plateau-age technique plus absolute dosimetry are shown for the samples Rajchi, Firura and Batan Orq'ó. Samples were irradiated in the Lazy Susan (Cd ratio 6.5 for Au and 48 for Co) facility of the Triga Mark II reactor of LENA, University of Pavia, Italy. The neutron fluence was determined by using the IRMM-540 standard glass [35] calibrated against thin films of natural uranium [3].

In the case of the sample Batan Orq'ó, an upper limit for the apparent age was obtained through both statistical procedures discussed above. The analysed area was 0.202 cm^2 (Table 3),

Table 2: Fission-track ages of Peruvian volcanic glasses obtained via the size correction method plus the Jankov Moldavite as the standard age. n_S (n_{S+I}): fields counted in spontaneous (spontaneous+induced) measurements (field size is $1.05 \times 10^{-4} \text{ cm}^{-2}$ unless stated differently); N_S (N_{S+I}): total number of tracks counted in spontaneous (spontaneous+induced) measurements; ρ_S (ρ_{S+I}): spontaneous (spontaneous + induced) track density; T_{Ap} (T_{corr}): apparent (corrected) age; D_S/D_I : spontaneous to induced track diameter ratio.

Sample	n_S	N_S	ρ_S (10^3 cm^{-2})	n_{S+I}	N_{S+I}	ρ_{S+I} (10^5 cm^{-2})	T_{Ap} (Ma)	D_S/D_I	T_{corr} (Ma)
Moldavite	391	404	9.84 ± 0.49	256	2025	0.75 ± 0.02	-	0.99 ± 0.02	-
Macusanite	289	908	29.9 ± 1.0	195 ^a	2164	6.60 ± 0.14	4.53 ± 0.22	0.95 ± 0.03	4.98 ± 0.57
Purupurini	6946 ^a	9	0.08 ± 0.03	237 ^a	1081	2.71 ± 0.08	0.03 ± 0.01	-	-
Batan Orq'ó	1243	0	-	55	2164	2.43 ± 0.05	-	-	$<0.006^b$ $<0.009^c$
Rajchi	1164	185	1.51 ± 0.11	97	1096	1.08 ± 0.03	1.36 ± 0.12	0.99 ± 0.03	1.38 ± 0.19
Firura	1804	227	1.20 ± 0.08	136	1546	1.08 ± 0.03	1.07 ± 0.08	0.88 ± 0.03	1.35 ± 0.17

Errors represent standard deviation of the mean (1σ).

^a Field size is $1.68 \times 10^{-5} \text{ cm}^{-2}$.

^b Galbraith [1] determination based on small counts.

^c Determination based on Poissonian probability of finding zero tracks.

which implies an age upper limit of ≈ 5400 years either by using Galbraith's or the $e^{-\mu}$ procedures. As both estimates provide the same result, further support is lent to the approach presented in this work. The areas analysed during the two dating procedures (Tables 2 and 3) can be summed up because they belong to different portions of the sample and are therefore independent. Thus, the total examined area is 0.333 cm^2 . This leads to a maximum age upper limit of ≈ 3430 years, by using the Poissonian probability of finding zero tracks.

It is worth mentioning that the Batan Orq'ó is a surface sample collected in an important archaeological site in Cuzco. This site is situated on a hill overlooking the Huaro Valley, located 10 km south-east of Pikillacta. Pikillacta is one of the largest archaeological sites dating to the Middle Horizon (A.D. 540-900) in Peru. The Middle Horizon is associated with the widespread expansion of an important pre-Inca civilization, the Wari. The Wari occupied much of Cuzco, building the large architectural complex of Pikillacta and a large settlement in the Huaro Valley, including the elite cemetery of Batan Orq'ó. Thus, the upper limit of age presented in this work could be associated with the possibility that this sample was part of an artefact that experienced a heating that occurred during its use and that have completely erased the pre-existing tracks.

Concerning the Rajchi sample the presented age determinations (1.38 ± 0.19) Ma (Table 2) and (1.14 ± 0.09) Ma (Table 3) are concordant in 2σ . The weighted mean value from both determinations is (1.19 ± 0.08) Ma. It should be noted that the Rajchi sample is also a surface sample and had been collected in the Rajchi Archaeological Complex. This Complex is related with the Inca civilization and was a heavily-populated area of vast dimensions, spreading out across 264 hectares. However, in this case, the obtained age probably is associated with a volcanic event.

Regarding the Firura sample, the age determinations showed, (1.35 ± 0.17) Ma (Table 2) and (1.01 ± 0.08) Ma (Table 3), are also concordant in the 2σ limit. The weighted mean value from both

Table 3: Fission-track ages of Peruvian volcanic glasses obtained via the plateau-age technique plus absolute dosimetry. Sample heating: thermal treatment imposed for plateau age determination; n_S (n_I): fields counted in spontaneous (induced) measurements (field size is $1.05 \times 10^{-4} \text{ cm}^{-2}$); N_S (N_I): total number of tracks counted in spontaneous (induced) measurements; ρ_S (ρ_I): spontaneous (induced) track density; D_S/D_I : spontaneous to induced track diameter ratio; T : determined age.

Sample heating	n_S	N_S	ρ_S (10^3 cm^{-2})	n_I	N_I	ρ_I (10^5 cm^{-2})	D_S/D_I	T (Ma)
<i>Batan Orq'ós</i>								
none	1924	0	-	41	1109	2.56 ± 0.08		$<0.0054^a$
<i>Rajchi</i>								
none	1495	215	1.37 ± 0.09	112	1430	1.22 ± 0.03	0.91	1.03 ± 0.08
4 h at 220°C	1547	190	1.17 ± 0.08	130	1285	0.94 ± 0.03	≈ 1	1.14 ± 0.09
<i>Firura</i>								
none	2441	264	1.03 ± 0.06	111	1485	1.27 ± 0.03	0.86	0.75 ± 0.05
4 h at 220°C	2009	192	0.91 ± 0.07	161	1423	0.84 ± 0.02	≈ 1	1.01 ± 0.08

Errors represent standard deviation of the mean (1σ).

Absolute dosimetry yields a value of $(1.85 \pm 0.09) \times 10^{15}$ neutrons/cm² for the thermal neutron fluence.

^a Determined through $e^{-\mu}$ and Galbraith's [1] upper limit for age.

determinations was (1.07 ± 0.07) Ma. The mean value of the Firura's age probably is associated to a volcanic event.

4 Conclusions

A new method to predict the upper age limit, even in the absence of spontaneous tracks, was presented and compared with the one proposed by Galbraith [1]. This method allowed to situate Batan Orq'ó fission-track maximum age at ≈ 3430 years, which might be associated to archaeological events.

The obtained ages to Macusanite, Rajchi and Firura were: (4.98 ± 0.57) Ma; (1.19 ± 0.08) Ma; and (1.07 ± 0.07) Ma, respectively. These ages are associated to volcanic events. In the case of Purupurini, due the low density of fossil tracks, only an apparent age, (0.028 ± 0.009) Ma, was obtained. The weighted mean value between the result obtained by Bigazzi *et al.* [9] and the Purupurini's age in this work is (0.037 ± 0.007) Ma.

Only one of the samples (Sara Sara) could not be dated, because it had inclusions which could be mistaken for tracks.

The comparison between the ages of Macusanite, Rajchi, Purupurini and Firura samples, obtained both through the size-correction method plus the Jankov Moldavite and the plateau-age technique plus absolute dosimetry, shows that the answer provided for both methods are coherent.

Acknowledgements

This work were financially supported by Brazilian foundations CNPq and FAPESP. We are indebted to the staff of the nuclear reactor of IPEN/CNEN, São Paulo, Brazil, for performing the neutron

irradiations. We wish to dedicate this work to Pedro José Iunes, a beloved brother with whom we shared this life.

References

- [1] R.F. Galbraith, *Statistics for Fission Track Analysis*, Chapman and Hall/CRC, 2005.
- [2] W. Zeil, *The Andes: A Geological Review*, Gebreder Borntraeger, 1979.
- [3] P.J. Iunes, J.C. Hadler, G. Bigazzi, C.A. Tello, S. Guedes and S.R. Paulo, *Chem. Geol.* **187** (2002) 201.
- [4] C.W. Naeser, G.A. Izett and J.D. Obradovitch, *Geol. Surv. Bull.* **1489** (1980) 1.
- [5] D.S. Miller and G.A. Wagner, *Nucl. Tracks* **5** (1981) 147.
- [6] D.S. Miller, N. Eby, R. McCorkell, P.E. Rosemberg and M. Suzuki, *Nucl. Tracks Rad. Meas.* **17** (1990) 237.
- [7] G. Poupeau, N. Sabil, I.M. Villa, G. Bigazzi, N. Vatin-Perigon, P. Flores, P. Pereyra, G. Salas and G. Arroyo, *Tectonophys.* **205** (1992) 295.
- [8] G. Poupeau, E. Labrin, N. Sabil, G. Bigazzi, P.G. Arroyo and N. Vatin-Perigon, *Nucl. Tracks Rad. Meas.* **21** (1993) 499.
- [9] G. Bigazzi, J.C. Hadler, P.J. Iunes, A.M. Osorio, *Rad. Meas.* **39** (2005) 585.
- [10] V.E. Barnes, G. Edwards, W.A. McLaughlin, I. Friedman and O. Joensuu, *Geol. Soc. Am. Bull.* **81** (1970) 1539.
- [11] M. Pichavant, J.V. Herrera, S. Boulmier, L. Briquieu, J.L. Joron, M. Juteau, A. Marin, L. Michard, S.M.F. Shepard, M. Treuil and M. Vernet, *Geochim. Soc. Special Publ.* **1** (1987) 359.
- [12] P.B. Price and R.M. Walker, *Phys. Lett.* **76** (1962) 113.
- [13] P.B. Price and R.M. Walker, *J. Appl. Phys.* **33** (1962) 3400.
- [14] P.B. Price and R.M. Walker, *J. Appl. Phys.* **33** (1962) 3407.
- [15] P.B. Price and R.M. Walker, *Nature* **196** (1962) 732.
- [16] R.L. Fleischer and P.B. Price, *J. Appl. Phys.* **34** (1963) 2903.
- [17] P.B. Price and R.M. Walker, *J. Geophys. Res.* **68** (1963) 4847.
- [18] G. Bigazzi, *Earth Planet. Sci. Lett.* **3** (1967) 313.
- [19] G.A. Wagner, *Earth Planet. Sci. Lett.* **4** (1968) 411.
- [20] D. Storzer and G.A. Wagner, *Earth Planet. Sci. Lett.* **5** (1969) 463.
- [21] D. Storzer and G. Poupeau, *Comp. Rend, Acad. Sci. Paris* **276D** (1973) 137.

- [22] M.L. Balestrieri, G. Bigazzi, V.J. Bouska, E. Labrin, J.C. Hadler, N. Kitada, A.M. Osorio, G. Poupeau, G. Wadatsumi and A. Zúñiga, in *Advances in Fission-Track Geochronology*, (eds. P. van den Haute and F. de Corte), Kluwer Academic Publishers (1998), p. 287.
- [23] M.A. Laurenzi, G. Bigazzi, M.L. Balestrieri and V. Bouska, *Meteoritics & Planet. Sci.* **38** (2003) 887.
- [24] A.S. Sandhu and J.A. Westgate, *Earth Planet. Sci. Lett.* **131** (1995) 289.
- [25] G. Bigazzi, J.C. Hadler, P.J. Iunes, M. Oddone, S.R. Paulo and A. Zúñiga, *Nucl. Instr. Meth. Phys. Res.* **A352** (1995) 588.
- [26] G. Bigazzi, M.A. Laurenzi, M. Soligo and P. Tuccimei, *J. Volc. and Geotherm. Res.* **177** (2008) 244.
- [27] A.H. Jaffey, K.F. Flynn, L.E. Glendenin, W.C. Bentley and A.M. Essling, *Phys. Rev.* **C4** (1971) 1889.
- [28] D. Galliker, E. Hugentobler and B. Hahn, *Am. Helv. Phys. Acta* **43** (1970) 593.
- [29] N.E. Holden and D.C. Hoffman, *Pure Appl. Chem.* **72** (2000) 1525.
- [30] S. Guedes, J.C. Hadler, P.J. Iunes, S.R. Paulo and A. Zúñiga, *J. Radioanal. Nucl. Chem.* **245** (2000) 441.
- [31] S. Guedes, J.C. Hadler, P.J. Iunes, A. Zúñiga, C.A. Tello and S.R. Paulo, *Nucl. Instr. Meth. Phys. Res.* **A496** (2003) 215.
- [32] S. Guedes, J.C. Hadler, Sarkis, Oliveira, Kakazu, Iunes, Saiki, Tello, Paulo. *J. Radioanal. Nucl. Chem.* 258:117-122, 2003.
- [33] N.E. Holden, *Pure Appl. Chem.* **61** (1989) 1483.
- [34] G.A. Cowan and H.H. Adler, *Geochim. Cosmochim. Acta* **40** (1976) 1487.
- [35] F. de Corte, F. Bellemans, P. van den Haute, C. Ingelbrecht and C. Nicholl, in *Advances in Fission-Track Geochronology*, (eds. P. van den Haute and F. de Corte), Kluwer Academic Publishers (1998), p. 67.

Reactivity, energetics and molecular structure: A theoretical and experimental approach

Juan Z. Dávalos Prado*

Instituto de Química-Física "Rocasolano"-CSIC, c/ Serrano 119-28006 Madrid, Spain

E-mail: jdavalos@iqfr.csic.es

Abstract: We study chemical reactivity, energetics, interactions and structure of neutral and ionic species – in the gas phase- of fundamental, biological and technological relevance. For this purpose we use a number of experimental (Fourier Transform Ion Cyclotron Resonance FT-ICR; calorimetry of combustion; Knudsen's effusion; photoelectron-photoion coincidence spectroscopy PEPICO) and theoretical techniques. The combination of the experimental results with those obtained by means of quantum-mechanic calculations (*ab-initio*, DFT) allow us to: i) obtain quantitative information on thermodynamic and kinetic of a variety of reactions in the phase gas, ii) determine interesting and novel relationship of reactivity-chemical structure, iii) determine the thermodynamic stability of neutral and ionic species, iv) to discover new species and new types of chemical bonds.

Dedicated to Prof. Holger Valqui C. on the occasion of his 80th birthday.

Key words: Thermodynamic stability, proton-exchange, molecular-structure, DFT, *ab-initio*, combustion calorimetry, FT-ICR, PEPICO.

1. Introduction.

Our research lines focus on the quantitative study of the energetics, structure, chemical reactivity and interaction among neutral and ionic species (organic, inorganic and organometallic) in the absence of the disturbing effect of the solvent. Medium effects on reactivity are often quite large. The determination of thermodynamic state functions for neutral and ionic species, as well as of rate constants for ion-molecule reactions and unimolecular dissociations, all of them in the gas phase, allows the study of intrinsic (i.e., solvent-unperturbed) reactivities. When necessary, quantitative information on the solvation of individual species (notably ions) can then be obtained through appropriate thermodynamic cycles. Frequently, these studies lead to the discovery of new reaction mechanisms or chemical bonds as well as of chemical species of unforeseen structures.

Fourier Transform Ion Cyclotron Resonance Spectroscopy (FT-ICR); combustion calorimetry, Knudsen Effusion and vaporization techniques, Photoelectron-Photoion Coincidence Spectroscopy (PEPICO) and theoretical methods (*ab initio*, DFT) are usually combined in our studies, which are structured into two interconnected lines: The first line, is basically related with small molecular systems where we used 4.7 T FT-ICR, combustion calorimetry/Knudsen effusion and PEPICO techniques. In the second line, recently implemented, our group tries to approach very current topics of chemical-physics interest, with

* Dr. J.Z. Dávalos has been student and after Lecturer in Physics at the Science Faculty-UNI.

particular incident in the macromolecular field (proteomics, metabolomics, polymeric materials). For this purpose we are using the 7.0T FT-ICR spectrometer provided with external ionization ESI/MALDI sources.

We believe that our investigation contributes, in narrow multidisciplinary collaboration, to the development of a chemical-physics of neutral and ionic species in the gas phase, "parallel" and frequently different from the observed one in condensed phases.

2. Experimental tools.

2.1. FT-ICR (*"Fourier Transform Ion Cyclotron Resonance Spectrometry"*).¹ It is a mass spectrometry based on the effect produced by an intense magnetic field (generated by superconducting solenoids) over the trajectories of charged particles (cyclotronic motion, Fig. 1a). As regards other spectrometries, FT-ICR presents the best sensitivity and resolution currently available.

Our group has used (from 1988) a modified Bruker CMS-47 FT-ICR mass spectrometer equipped with a 4.7 T superconducting magnet to the study of ion-molecule processes where take part "small" volatile species. At present, we have a new FT-ICR spectrometer, Agilent/Varian-920 (Fig. 1b), provided with a 7.0 T actively shielded superconducting magnet and equipped with the external ionization sources: i/ MALDI (*Matrix-Assisted Laser Desorption Ionization*) and, ii/ ESI (*Electrospray Ionization*), which is in line with a Triple-quadrupole Agilent/Varian-320 mass spectrometer. This hybrid-spectrometer is provided with fragmentation techniques such as SORI-CID (*Sustained Off-Resonance Irradiation- Collision-induced Dissociation*), IRMPD (*IR Multiphoton Dissociation*) or ECD (*Electron Capture Dissociation*), and it can be also coupled to chromatographic interfaces (GC or liquid HPLC).

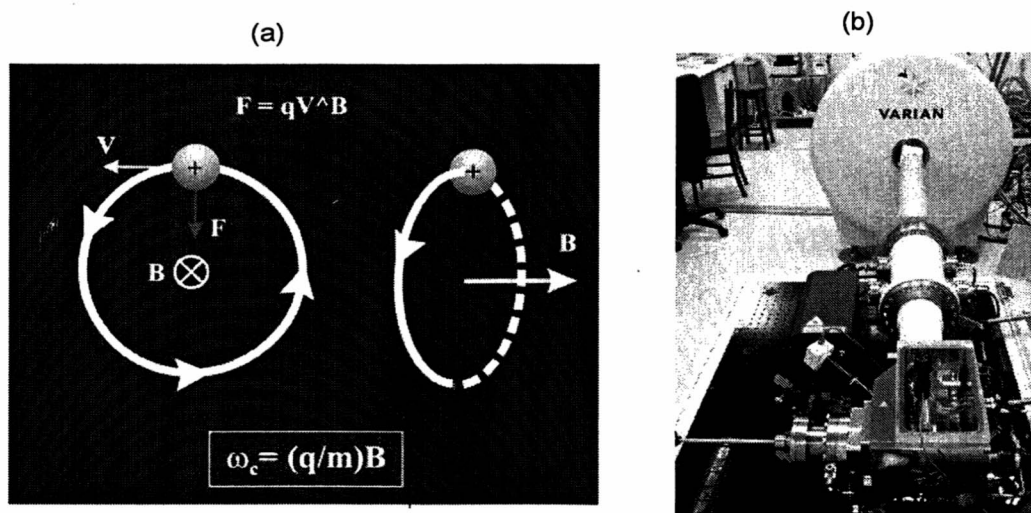
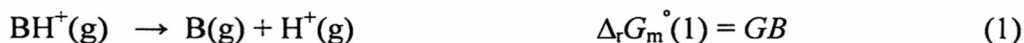


Fig.1. (a) Cyclotronic motion. (b) Hybrid FT-ICR of 7.0T with ESI/MALDI sources

Basicity (GB) and acidity (GA) in the gas phase. The basicity of a base B and the acidity of an acid AH, in gas phase can be defined in terms of standard Gibbs energy changes for reactions (1) and (2), respectively: $GB(B) = \Delta_r G_m^0(1)$ and $GA(AH) = \Delta_r G_m^0(2)$.



2.2. Photodissociation electron-ion coincidence techniques: PEPICO (*"Photoelectron-Photoion Coincidence"*), using basically Synchrotron Radiation lines (SLS-Paul Scherrer Institut-Switzerland, LNLS-Campinas-Brazil). It can provide quantitative and qualitative information on: i) Heats of formation [$\Delta_f H_m^0(\text{g})$] and binding energies (*BDE*) of ionic and neutral species practically impossible to obtain by means of the conventional techniques, such as calorimetries. ii) Dynamic of photo-fragmentation processes (in energy ranges of UV and soft X-Ray) of energy-selected ions.

2.3. The traditional calorimetry techniques (calorimetry of combustion, differential scanning calorimetry DSC) and those of vaporization/sublimation (Knudsen's Effusion), suitably modified and updated, continue providing quantitative precise information on: i) Thermodynamic stability (heats of formation) of organic neutral molecules including C, H, O, N, S and halogens (Cl, Br, I). ii) Thermophysical properties such as calorific capacities, phase transitions. iii) Energetic properties of organic samples.

3. Theoretical methods

The quantum chemical calculations are carried out using the Gaussian 03 and 09 packages.² The geometries of the compounds under investigation are optimized by using *Density Functional Theory* (DFT), with the Becke 3 parameter and the Lee, Yang, Parr (B3LYP)³ and Truhlar functionals (M05, M06),⁴ usually without symmetry restrictions. The *ab-initio* calculations employed in several of our works are related with Møller-Plessed (MP2)⁵ and Gn (n=2, 3, 4)⁶ level of theories, which ones predict the energies quite well.

4. Results

Here are described some of the relevant results obtained by our group:

4.1. The $\text{P}_4\cdots\text{Li}^+$ ion in the gas phase: A planetary system.⁷

We have explored experimentally and computationally the existence, thermodynamic stability, and properties of the seemingly unknown ion $\text{P}_4\cdots\text{Li}^+$ (tetraphosphorus-Li) in the gas phase.



The Gibbs energy of dissociation is known as lithium cation basicity, *LCB*, of P_4 and it was determined by experimental (FT-ICR) and computational (G2-level) methods.

Our results were the following:

- Bracketing experiments show that $\text{P}_4\text{Li}^+(\text{g})$ suffers irreversible transfer of Li^+ . We have estimated a $LCB(\text{P}_4) = 19.4 \pm 4.8$ kcal/mol,
- *Ab initio* calculations (at G2 level) lead to a value of $LCB(\text{P}_4)$ of 17.5 kcal/mol

The study of isomerization barriers reveals the existence of 4 groups of stationary points (Fig. 2a) in the potential energy surface PES: Three of them corresponding respectively to the

attachment of the Li^+ cation to a corner or apex (1), an edge (2), and a face (3) of the tetrahedron. The fourth structure, in which Li^+ occupied the center of tetrahedron, was found to be a saddle point.

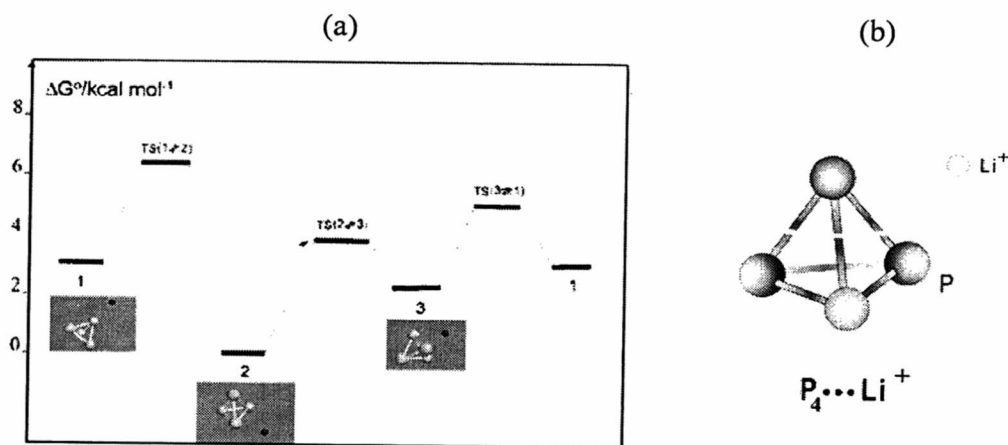


Fig. 2 (a) Energetic diagram showing the stable structures 1, 2, and 3 for P_4Li^+ as well as the transition states of these structures. (b) P_4Li^+ cation

All these results imply that P_4Li^+ in the gas phase provides a nice of a “planetary system” (Fig. 2b) wherein Li^+ cation can easily move around P_4 along “orbits” connecting the points above the middle points of the P-P edges with points above the center of PPP faces and points of corners. Obviously, the motion of Li^+ will be favored by increasing the temperature of system. At room temperature RT, this orbiting will likely take place in a nanosecond time scale.

4.2. Discover new species and chemical bonds: A Carbon atom covalently bound to five ligands, the case of $\text{Si}_2(\text{CH}_3)_7^+$ cation.⁸

We have experimentally shown, by means FT-ICR spectrometry, that $[\text{Me}_3\text{Si}-\text{SiMe}_3]^+$ (I^+ , m/z 161) cation (Fig. 3) is asymmetric, fluxional ion. Theoretical calculations at a very substantial level show that the optimized geometry [B3LYP/6-311+G(3df,2pd)] of the ion I^+ corresponds to a C_{3h} -symmetric structure, in which a totally planar CH_3 group is symmetrically bonded to two eclipsed $\text{Si}(\text{CH}_3)_3$ moieties. Accordingly, the most important conclusion of our theoretical study is that in I^+ , the central carbon atom is pentacoordinated. The calculations also rationalize the fluxionality of this ion.

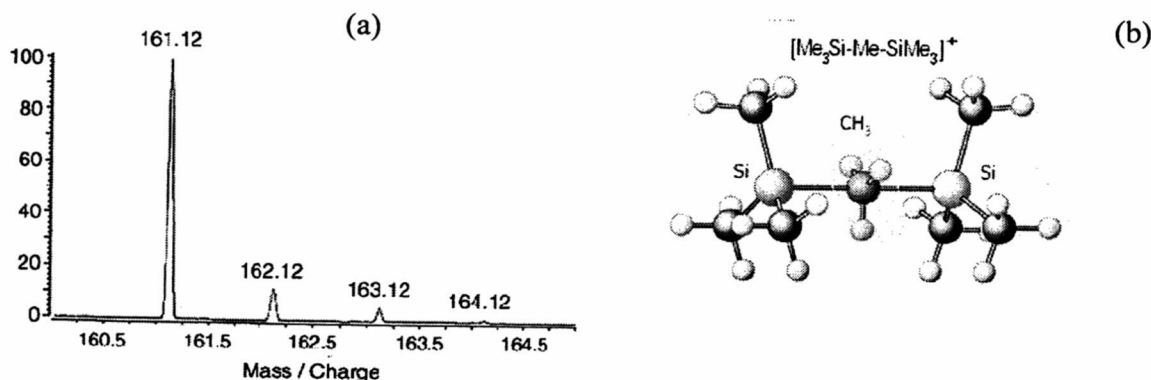


Fig. 3 (a) FT-ICR mass spectrum of cation I^+ , (b) optimized geometry for I^+ .

4.3. Neutral, Ion Gas-Phase Energetics and Structural Properties of Hydroxybenzophenones.⁹

We have carried out studies of the energetics, structure, and physical properties of *o*-, *m*-, and *p*-hydroxybenzophenone neutral molecules ($C_{13}H_{10}O_2$) and their corresponding anions (obtained by de-protonation of the OH group). In particular, we determined the heat of formation, $\Delta_f H_m^0(g)$, for all of these species. A reliable experimental estimation of the enthalpy associated with intramolecular hydrogen bonding in chelated species was also experimentally obtained. The gas-phase acidities (*GA*) of benzophenones, substituted phenols, and several aliphatic alcohols are compared with the corresponding aqueous acidities (pK_a), covering a range of 278 kJ/mol in *GA* and 11.4 in pK_a (Fig. 4).

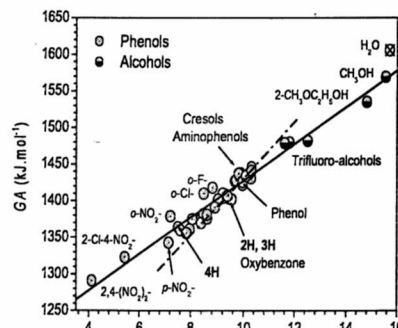
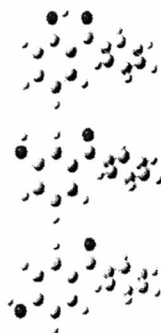


Fig. 4 Molecular geometries and correlation of acidity in the gas phase (*GA*) and in solution (pK_a)

4.4. Towards superacidity in the gas phase: Enhanced Acidity of Phosphine- and Amine-Boranes.^{10,11}

Our experimental and theoretical study on series of phosphines, amines and their phosphine-, amine-borane adducts showed that BH_3 attachment leads to a substantial increase in the intrinsic acidity of the system: from 80 to 110 kJ/mol for $R-PH_2BH_3$ and almost twice of this range, for $R-NH_2BH_3$.

This acidity-enhancing effect of BH_3 is enormous, between 13 and 18 orders of magnitude in terms of ionization constants. This indicates that the acidity enhancement of protic acids by Lewis acids observed in solution also occurs in the gas phase. It is interesting to remark that for $R-NH_2BH_3$ case, the most significant finding is that typical nitrogen bases, such as aniline, react with BH_3 to give amine-borane complexes, which in the gas phase have acidities as high as those of either phosphoric, oxalic, or salicylic acid; their acidity is higher than many carboxylic acids, such as formic, acetic and propanoic acid.



Fig. 5 Complexes of phosphine- and amine-boranes

4.5. Electronic transfer reactions (ET) on “cage” type hydrocarbons.¹²

It has been experimentally established that the radical cation of adamantane ($C_{10}H_{16}$) and adamantylidene-adamantane ($C_{20}H_{28}$) are stable (albeit very reactive) species. Use of the natural ^{13}C “labelling” of these compounds allowed experimental determination of the rate constants for isergonic electron exchange between these hydrocarbons and their corresponding radical cations. These rate constants are extremely high and practically collision-controlled in both cases.

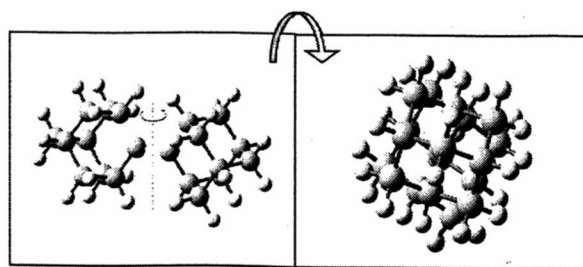


Fig. 6 Structure of adamantylidene-adamantane

Theoretical computations exploring different possible reaction paths have given insight into the role that intermediate complexes may play in this ET reaction.

4.6. Photodissociation dynamics and energetics of Tin species.¹³

The photodissociation dynamics and energetics of tetramethyltin (Me_4Sn) and hexamethylditin (Me_6Sn_2) (Fig. 7) has been studied by TPEPICO spectrometry. Ions are energy-selected, and their 0 K dissociation onsets are measured by monitoring the mass spectra as a function of ion internal energy. Me_4Sn^+ dissociates rapidly by methyl loss, whereas the hexamethylditin ion dissociates slowly on the time scale of the experiment so that dissociation rate constants are measured as a function of the ion energy. Rice–Ramsperger–Kassel–Marcus (RRKM) theory and the simplified statistical adiabatic channel model (SSACM) are used to extrapolate the measured rate constants for methyl and radical $\text{Me}_3\text{Sn}^\bullet$ loss to their 0 K dissociation onsets. Updated values for the heats of formation of the neutral Me_4Sn and Me_6Sn_2 are used to derive the following 298.15 K gas-phase standard heats of formation, $\Delta_f H_m^0(\text{g})$, in kJ/mol: for Me_3Sn^+ , 746.3 ± 2.9 ; Me_5Sn_2^+ , 705.1 ± 7.5 ; $\text{Me}_3\text{Sn}^\bullet$, 116.6 ± 9.7 ; Me_2Sn , 123.0 ± 16.5 ; MeSn^+ , 877.8 ± 16.4 . These energetic values also lead to the following 298.15 K bond dissociation enthalpies, in kJ/mol: $\text{BDE}(\text{Me}_3\text{Sn}-\text{Me}) = 284.1 \pm 9.9$; $\text{BDE}(\text{Me}_3\text{Sn}-\text{SnMe}_3) = 252.6 \pm 14.8$.

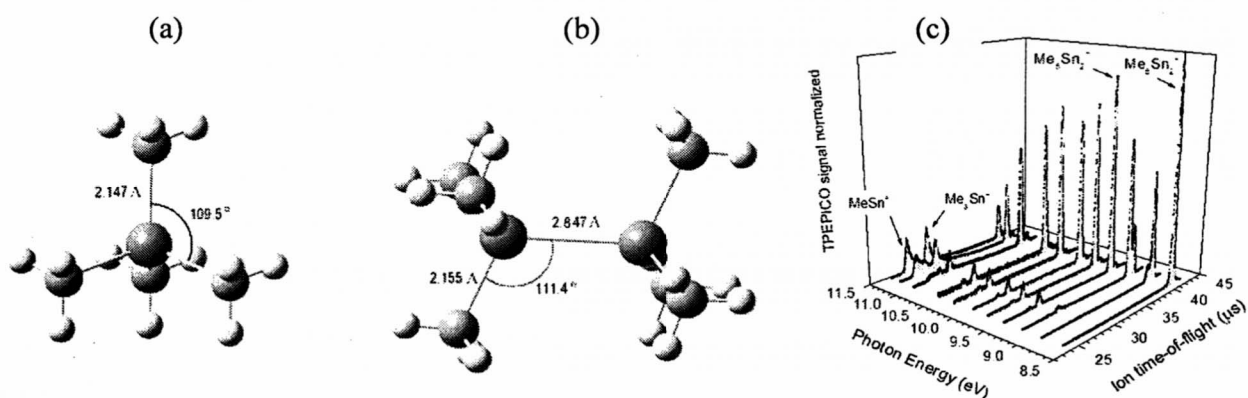


Fig. 7 (a) Molecular geometry for Me_4Sn and (b) staggered Me_6Sn_2 , (c) TPEPICO ion time-of-flight (linear TOF) distributions for Me_6Sn_2 at several selected photon energies.

4.7. Energetics of *trans*-hydroxycinnamic acids.¹⁴

Phenolic acids, such as hydroxycinnamic and hydroxybenzoic acids, are a diverse group of aromatic secondary plant metabolites. They are ubiquitously distributed, in both edible and non-edible plants, as esters or glycoside derivatives. In recent years these acids and derivatives have attracted much attention due to their various biological, photobiological and pharmaceutical activities and also to their industrial or technological applications. Many of these activities are related to their antioxidant properties, which may be due to their ability to scavenge free radicals and/or to synergistic effects with physiological antioxidants and several enzymes. In fact, these actions could prevent oxidative damage of biomolecules (proteins, membrane lipids, and nucleic acids) related to various diseases such as cancer, cardiovascular risks or

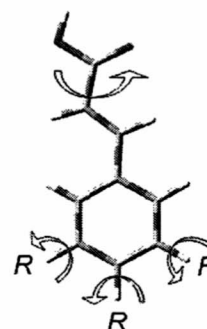


Fig. 8 Hydroxycinnamic acids

diabetes. In this work, we have studied the energetics and structural properties of *trans*-cinnamic, coumaric, caffeic, ferulic and sinapic acids. It has been determined the heat of formation, $\Delta_f H_m^0(g)$, for all of these species. These values have permitted to study the energetic effects of π -donor substituents ($-\text{OH}$ and $-\text{OCH}_3$) in cinnamic acid derivatives and in the respective benzene analogues. Our results indicate that the interaction between $-\text{OCH}_3$ and/or $-\text{OH}$ groups in hydroxycinnamic acids takes place without significant influence of the propenoic fragment.

4.8. Study of oligomerization in icosahedral *closo*-(car)boranes Lithium clusters.¹⁵

We have carried out an experimental and computational study of the oligomerization in icosahedral *closo*-(car)boranes lithium-based clusters, $\text{Li}_2\text{B}_{12}\text{H}_{12}$ (**I**) and $\text{LiCB}_{11}\text{H}_{12}$ (**II**). The experiments were performed on a hybrid ESI-TQ-FT-ICR mass spectrometer. According to the experimental results, the dianion $\text{B}_{12}\text{H}_{12}^{2-}$ shows a bigger affinity for lithium than the monoanion $\text{CB}_{11}\text{H}_{12}^-$. Signals of oligomer anions of **I** singly and doubly charged have been detected (depending on the loss of one or two Li^+ ions), which were identified, respectively, as $[\text{I}_n - \text{Li}]^-$ ($n = 1, 2, 3, 4$) and $[\text{I}_n - 2\text{Li}]^{2-}$ ($n = 1, 4, 5, 6, 7$). As for **II**, only monomer and dimer singly-charged anions $[\text{II}_n - \text{Li}]^-$ ($n = 1, 2$), were detected. These results are explained and rationalized by high-level computational studies, finding the most stable structures of **I**, **II** and their derived cluster anions. We have also estimated the bond dissociation energy (*BDE*) between Li and the icosahedral borane cages in the processes $[\text{X}_2 - \text{Li}]^- \rightarrow \text{X} + [\text{X} - \text{Li}]^-$ with $\text{X} = \text{I}, \text{II}$. Comparison of experiments and computations show an excellent agreement for the bond dissociation energy in process involving **II**, with $\Delta E = 1.5$ eV.

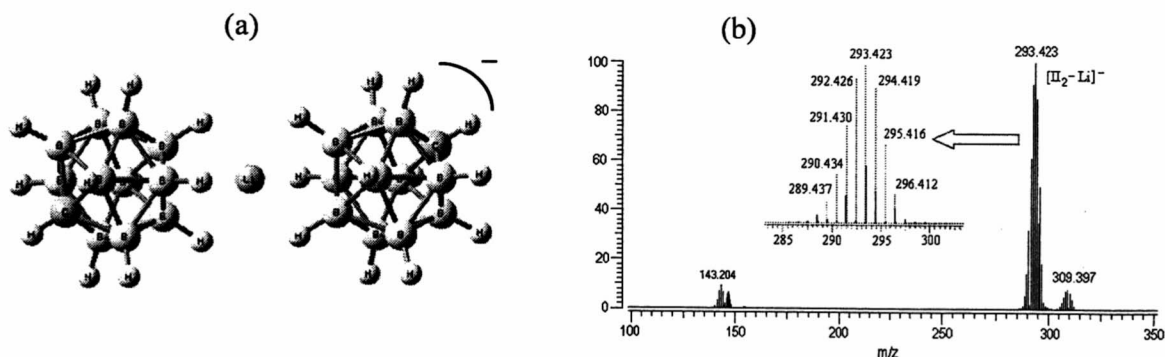


Fig. 9 (a) C_{2h} optimized geometry of the homodimer anion $[\text{II}_2 - \text{Li}]^-$ ($[\text{CB}_{11}\text{H}_{12} \cdots \text{Li} \cdots \text{CB}_{11}\text{H}_{12}]^-$), obtained at B3LYP/6-311++G(d,p) level. (b) ESI-FT-ICR mass spectrum of cluster $[\text{II}_2 - \text{Li}]^-$ in the negative mode.

REFERENCES

- [1] (a) A.G. Marshall. *Adv. Mass Spectrom.* 11, **1989**, 651. (b) A.G. Marshall, C.L. Hendrickson, G.S. Jackson. *Mass Spectrom. Rev.* 17, **1998**, 1.
- [2] Frisch, M.J., et al., *Gaussian 03, Revision C.02*, 2003 and *Gaussian 09*, Inc., Wallingford CT; A.02 ed. 2009.
- [3] (a) A.D. Becke, *J. Chem. Phys.* **1993**, 98, 5648. (b) C. Lee, W. Yang, R.G. Parr, *Phys. Rev.* **1988**, B37, 785.
- [4] Y. Zhao, N.E. Schultz, D.G. Truhlar, D. G. *J. Chem. Theory Comput.* 2, **2006**, 364.

-
- [5] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, In "Ab Initio Molecular Orbital Theory" John Wiley, New York, 1986.
- [6] L. A. Curtiss, P. C. Redfern, K. Raghavachari, *J. Chem. Phys.* 126, **2007**, 084108.
- [7] J.-L.M. Abboud, I. Alkorta, J.Z. Dávalos, J.-F. Gal, M. Herreros, P.-C. María, O. Mó, M.T. Molina, R. Notario and M. Yáñez, *J. Am. Chem. Soc.* 122 (**2000**) 4451-4454
- [8] J.Z. Dávalos, R. Herrero, J.-L.M. Abboud, O. Mó, M. Yáñez, *Angew. Chem. Int. Eng. Ed.* 46 (**2007**) 381-385.
- [9] J.Z. Dávalos, A. Guerrero, R. Herrero, P. Jimenez, A. Chana, J.L.-M. Abboud, C.F.R.A.C. Lima, L.M.N.B.F. Santos, A.F. Lago, *J. Org. Chem.* 75 (8) (**2010**) 2564-2571.
- [10] M. Hurtado, M. Yáñez, R. Herrero, A. Guerrero, J.Z. Dávalos, J.-L.M. Abboud, B. Khater, J.-C. Guillemin, *Chem. A Eur. J.* 15 (**2009**) 4622-4629.
- [11] A. Martín-Sómer, M. Lamsabhi, M. Yáñez, J.Z. Dávalos, J. González, R. Ramos and J.-C. Guillemin, *Chem. A Eur. J.* 18 (**2012**) 15699-15705.
- [12] A. Guerrero, R. Herrero, E. Quintanilla, J.Z. Dávalos, J.-L. M. Abboud, P.B. Coto, D. Lenoir, *Chem. Phys. Chem.* 11 (**2010**) 713-721.
- [13] J.Z. Dávalos, R. Herrero, N. Shuman, T. Baer, *J. Phys. Chem. A.* 115 (**2011**) 402-409.
- [14] J.Z. Dávalos, R. Herrero, A. Chana, A. Guerrero, P. Jiménez, J.M. Santiuste, *J. Phys. Chem. A.* 116 (**2012**) 2261-2267.
- [15] J.Z. Dávalos, J. González, A. Guerrero, D. Hnyk, J. Holub, J.M. Oliva, *J. Phys. Chem. C*, In press (dx.doi.org/10.1021/jp3102354)

Proton-fountain Electric-field-assisted Nanolithography (PEN): Fabrication of polymer nanostructures that respond to chemical and electrical stimuli.

**An overview in the context of the top-down and bottom-up approaches to
nanotechnology**

Andres La Rosa¹, Mingdi Yan², Rodolfo Fernandez¹, Xiaohua Wang¹ and Elia Zegarra^{1,3}

¹Portland State University, Portland, Oregon 97207, USA

²University of Massachusetts Lowell, Lowell, MA 01854, USA

³Universidad Nacional de Ingeniería, Facultad de Ciencias, Lima-PERU

I. ABSTRACT

The development of chemically functionalized materials, such that their physical properties can vary in response to external mechanical, chemical, or optical stimuli, offers potential applications in a wide range of fields, namely microfluidics, electronic memory devices, sensors and actuators. In particular, patterned structures built with stimuli-responsive *polymer* materials are attractive due to their inherent lower cost production and for building soft scaffolds that mimic closer natural bio-environments. In addition, harnessing the construction of patterns with *nanoscale dimensions* would not only *a)* allow building lab-on-a-chip devices that require minimal chemical reactants volumes, but also *b)* find applications in the area of nano-electronics for fabricating flexible, low-cost, and low-voltage-operation integrated logic circuits devices. To address these potential applications of stimuli-responsive polymer nanomaterials in the bio and nano-electronics arena, this article provides first a brief review of radiation and non-radiation based lithography methods used for fabricating nanopatterns. This introduction helps to put in context a more general description of the *Proton-fountain Electric-field-assisted Nanolithography* (PEN) technique, a recently introduced scanning-based method able to fabricate patterns of nanoscale dimensions using responsive polymer films. We also outline potential avenues for the outgrowth of PEN by replacing its current top-down fabrication approach with a bottom-up modality. The proposed outgrowth is to improve the fabrication speed and the lateral dimensions of the patterns. More specifically, we address the fact that, since PEN capitalizes on the reversible swelling-response of poly(4-vinylpyridine) (P4VP) films upon spatially-localized injection of protons (hydronium ions H_3O^+), the diffusion of the positive charges inside the polymer film matrix limits the patterns lateral resolution. This shortcoming can be remediated by the integration of ultra-fast optical activation into the PEN technique in order to gain much finer control over the functionalized sample area where the polymer molecules are selectively attached to the substrate, which would allow implementing a diffusion free, nanometer resolution, self-assembly method for fabricating erasable polymer nanostructures.

I. The development of PEN in the context of the emerging fields of biomimetic and nano-electronics

There exists a long standing interest in the biomimetic field for unraveling the inner working principles of biological systems for, ideally, building devices that closely mimic their functioning. Lipid bilayers membranes that separate cells from their external environments constitute specific example of the sophisticated mechanisms observed in biological machineries, where embedded membrane-proteins facilitate the communication between the interior and exterior sides carrying out a multitude of tasks, namely signal transduction, transportation of small molecules, catalytic reactions, etc. (See Ref. 1 provides an interesting review of intriguing mechanisms by which these molecular machines operate.)¹ Interesting enough, this fascinating concerted bio-chemical work is revealed at different dimension scales. Indeed, nature provides with vast examples of highly organized hierarchical structures whose dimensions range from the micron-size (cell) to the nano-sized scale.² An example of the latter constitutes the flagella of bacteria, which rotates at over 10,000 r.p.m. driven by a protons-flow caused by electrochemical potential differences across the membrane; the diameter of the bearing is approximately 30 nm, with an estimated clearance of approximately 1 nm.³

On the other hand, there are increasing evidences that living cells are not just chemical factories (as tacitly implicit in the lines above) but they can also be conceived as mechanical devices,^{4,5} for it is found that cell membranes are very sensitive to the mechanical properties of its surrounding matrix (affecting their growth, differentiation, migration, and, eventually, apoptosis.)^{6,7} Incidentally, a recent report indicates that stem cells do not regenerate efficiently in vitro environment unless the surrounding medium is made out of soft materials.^{8,9} This new mechanical sensitivity characteristic has fueled further the development of synthetic polymer scaffold for regenerative medicine.^{10,11,12,13} In addition, there exists special interest for micron- and nano-sized structures. Their advantage can be contrasted within the field of tissue engineering where, for example, hydrogel scaffolds¹⁴ are found to cause premature death of cells (necrosis) due to diffusion limitations, even though these artificial scaffolds closely mimic the chemical and mechanical properties of natural extracellular matrix.¹⁵ What happens is that in large hydrogels it is difficult to control the three-dimensional architecture and cell-cell interactions, which makes it difficult to replicate the complexity of real tissues. Micron-sized hydrogels, in contrast, have no such limitations. For example, by using hydrogels of controlled sizes and shapes¹⁶, it is possible to minimize diffusion limitations while fabricating tissues with complex microarchitecture.¹⁷

Inspired by the well-coordinated chemical and mechanical processes displayed by living cell membranes, there has been successful attempts in the design of a variety of building blocks (gels,^{18,19} brushes,²⁰ hybrid systems with inorganic particles²¹) that respond selectively to different (pH,²² temperature,^{23,24} optical,^{15,25} and magnetic²⁶) external stimuli. One particular approach within the biomimetic-materials field constitutes the development of versatile stimuli responsive thin films²⁷ The rationale behind these efforts conceives building complex synthetic hierarchies (needed to eventually mimic *nature*) as a combination of functional-domains separated by stimuli-responsive polymer thin films, the latter regulating the interactions between the domain compartments.²⁸ This vision constitutes one of the motivations for developing *Proton-fountain Electric-field-assisted Nanolithography* (PEN), which aims at

harnessing the fabrication of nanoscale patterning of responsive polymer films, and whose description is the subject of this article.

PEN was introduced very recently²⁹ by demonstrating its ability to pattern features (of nanometer-size in height and micron-size wide) in poly(4-vinylpyridine) P4VP films. The initial selection of this polymer responded to the fact that its swelling properties, triggered by wet exposure to acids, had been previously demonstrated in experiments performed at macroscopic scales.³⁰ PEN has extended this principle to the nanoscale domain,²⁹ aiming to fabricate artificial scaffolds that closely mimic the mechanical properties of natural extracellular. In addition, being a polymer of conjugated type (it contains pyridine groups, with π molecular orbitals, along its chain) P4VP is also being used to explore opto-electronics applications,³¹ which is timely in the current trend of research interest for developing low-cost, low-voltage operation, flexible nano-electronic devices.³² Altogether, the potential applications in the biological and electronics fields have triggered further the interest of our research groups for developing PEN. The first PEN demonstrations were implemented using a scanning probe modality, as described in more detail in the following sections. In this article we also explore new experimental approaches to overcome the inherent slow process associated to its scanning probe implementation and how to improve the capability to fabricate patterns with finer line widths.

This article is organized as follows. Section II provides a brief view on past and current trends in the fabrication of nanomaterials, including specific examples of the top-down approach (briefly describing radiation and non-radiation-based lithographic techniques for building single electron memory devices) and the bottom-up approach (describing the construction of self-assembly complex architectures using P4VP material.) The purpose of section II is to first provide a quick view of the underlying efforts for harnessing the fabrication of materials with nanoscale dimensions. Section III offers a summary of early results obtained using PEN, including technical details of the fabrication process with an emphasis on attaining an understanding of the working principle of the molecular "glue" used in the fabrication process; the latter is used to covalently attach polymer molecules to a substrate. Section IV outlines the integration of ultra-fast optical activation into the PEN technique in order to gain control over the specific area that will be chemically functionalized and where the polymer molecules will be selectively attached to the substrate. Such a technical outgrowth would allow implementing a diffusion free self-assembly method for fabricating erasable polymer nanostructures.

II. General trends in the fabrication of nanomaterials

Nanomaterials fabrication methods can be classified according to whether their assembly followed either *i*) the so called *bottom-up* approach, where smaller components of atomic or molecular dimensions self-assemble together, according to a natural physical principle or an externally applied driving force, to give rise to larger and more organized systems;^{33, 34} or *ii*) the *top-down* approach, a process that starts from a large piece and subsequently uses finer and finer tools for creating correspondingly smaller structures.³⁵ These two approaches are schematically presented in Fig. 1. There are advantages and disadvantages in both approaches, which are estimated according to their speed, reproducibility, and cost.

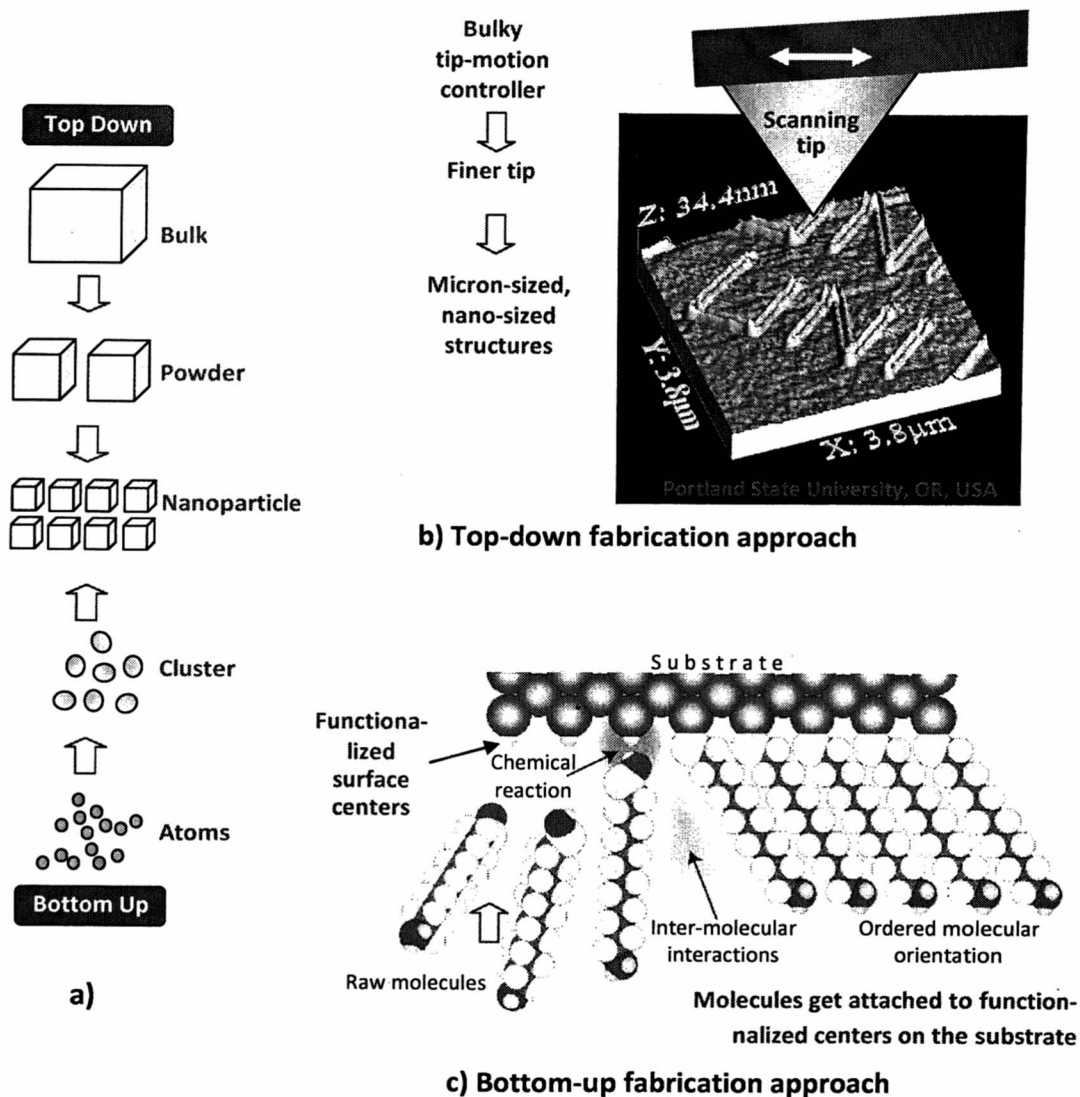


Fig.1 a) Schematic representation of the formation of nanostructures via the top-down and bottom-up approaches.³⁶ **b)** A tapered probe, manipulated by a macroscopic machine, allows “writing” small features by scratching the probe apex on a soft polymer surface. Notice the removed material disperses on the sides of the created grooves. *Image attained at Portland State University, Portland, Oregon, USA; courtesy of Rodolfo Fernandez.* **c)** Example of self-assembling set to occur on previously chemically-functionalized surfaces; chemisorption happens due to interactions between adsorbing molecules and specific sites on the substrate. The resulting nano-electronics material emerges with much more effective functions when proper shapes and microstructures are provided.^{37,38} *Image courtesy of Dr. Hiroyuki Sugimura, Nanoscopic Surface Architecture Laboratory, Department of Materials Science and Engineering Kyoto University, Japan.*

II.1 The “top-down” approach to nanotechnology

II.1A Photolithography silicon technology: Radiation-based method for fabricating electronic devices

One of the best examples of the top-down fabrication procedure is photolithography silicon technology,^{35, 39} which was originally developed for microelectronics (integrated circuits planar technology) but now it is also used to make miniature machines (three dimensional structures obtained by exploiting the preferential material-etching along the silicon’s crystallographic planes.)^{40,41} This revolutionary fabrication process started with the invention of the point-contact transistor in 1947⁴² that heralded the dawn of microelectronics. Since then, the industry has been driven by the demand to build devices that squeeze an ever increasing number of individual circuit elements onto ever smaller pieces of semiconductor materials.⁴³

The ability to shrink the size of the devices depends on the particular lithographic technique used to make the circuit pattern. In the photolithography process outlined in figure 2, pre-designed patterns are transferred from a mask (made, for example, using electron beam lithography) to the target silicon substrate. The process comprises the following typical steps:

- i) Spin-coating a thin layer of light-sensitive polymer resist on a silicon wafer sample (Fig. 2b).
- ii) Subsequently the sample is illuminated with ultra-violet light through the patterned-apertures of a mask (Fig. 2c); hence only some regions of the resist are exposed to the uv-light, which causes changes in the resists’ solubility. There exist two modalities for using the mask. In the *contact mode*, illustrated in Fig. 2 c-1, the mask is placed in almost near-contact with the sample; this modality is prone to cause damage to both, unless very much care is placed in the procedure. This type of setting is widely used in research labs. But for applications involving large volume production, where the integrity of the components becomes an issue, the mask is instead placed at a working-distance from the sample. In this *projection mode* the mask features are projected through an optical system, as shown in Fig. 2 c-2, which offers a safer alternative than the contact mode. Notice however that, due to light diffraction effects, the resist’s areas exposed to radiation will be larger than the corresponding opening areas of the mask (i.e. there is a loss in resolution). This imposes limitations on lithography-based procedures for creating devices of dimensions smaller than the wavelength. Thus the semiconductor industry has also been driven to use radiation of smaller wavelength.
- iii) The procedure continues dipping the sample in a developer in order to remove the resist from the places where it has been exposed to light (Fig. 2d).
- iv) The resist pattern on the wafer is subsequently used to either etch material from the naked regions where the resist exposes the wafer to the etching agent (Fig. 2e) or, alternatively, to deposit additional material as required by the circuit design. Afterwards, the resist is stripped completely off the wafer (Fig. 2f).

Notice that after these four steps, features on the mask (step c) will have been transferred to the Si-substrate (step-f). This whole process is repeated as many times as required by the circuit design complexity.

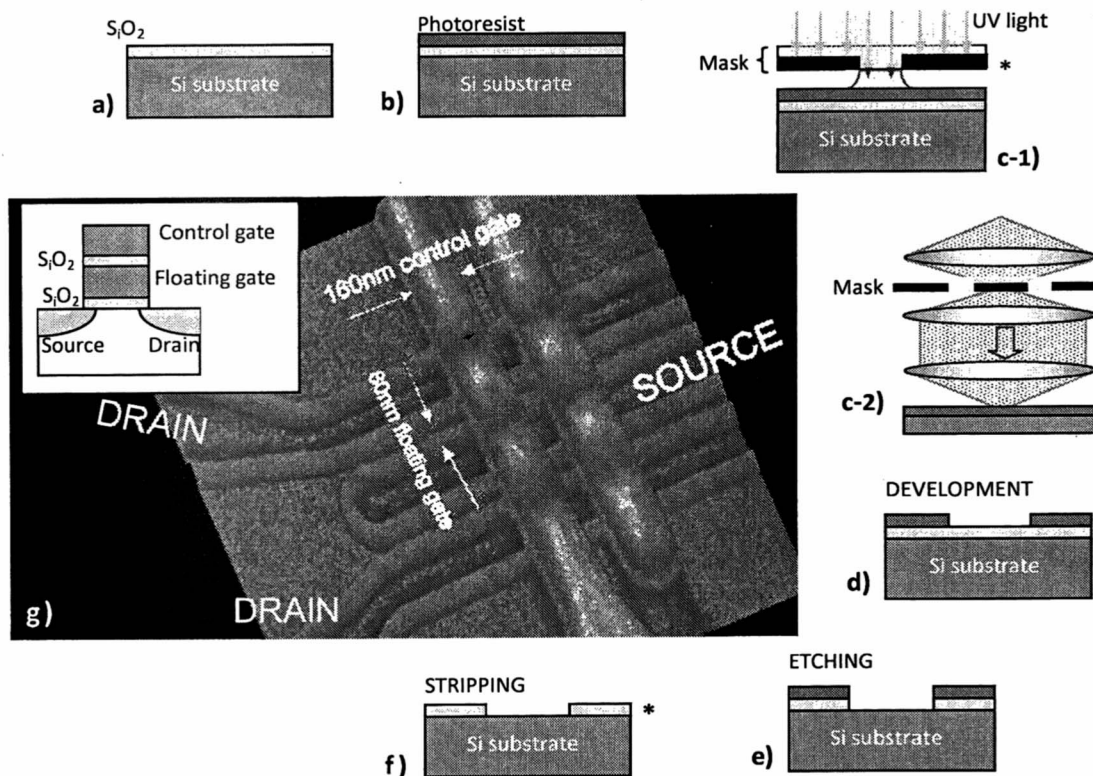


Fig. 2 a) to f) Schematic diagram of a photolithography process. The asteristics on step-c1 and step-f are to highlight the fact that features on the mask are replicated onto the Si-substrate (except for some limitations imposed by light diffraction, which makes the features on the silicon to be larger than the ones in the mask.) **g)** Non-volatile memory cell with a (80 nm wide and 160 nanometers long) "floating gate" central storage data, built at Bell Lab using optical lithography. *Image of the non-volatile memory reprinted with permission of Alcatel-Lucent USA Inc.*

A representative example of a top of the line device manufactured using a lithography process is the non-volatile memory device shown in Fig. 2g, which stores data even when its power supply is turned off. This device is basically a metal-oxide-semiconductor field effect transistor (MOSFET), except for having a modified gate electrode. For completeness, a schematic of both is shown in Fig. 3. The top and right sides displays the working principle of a MOSFET, where a key role is played by the electron conductive channel (separated from the p-type substrate by a depletion region) that forms upon the application of a sufficiently high gate voltage (greater than a threshold voltage V_T); subsequent modification of the channel (in response to different values of the drain voltage V_D) renders transistor operation with very high input impedance and unequal low-power consumption device. The ability to modulate the conductance of the channel through a gate electrode that is isolated from the channel by a thin oxide constitutes one of its main advantages. Fig. 3 also shows a modified MOSFET where a poly-silicon floating gate is fabricated in between the control gate and the channel, separated by a thin oxide. When a positive voltage is applied to the control gate, charge is injected from

the channel region and through the gate oxide into the floating gate. When the applied voltage is removed, the injected charge can be stored in the floating gate for a long time (days). To remove the charge, a negative gate-voltage V_G is applied.

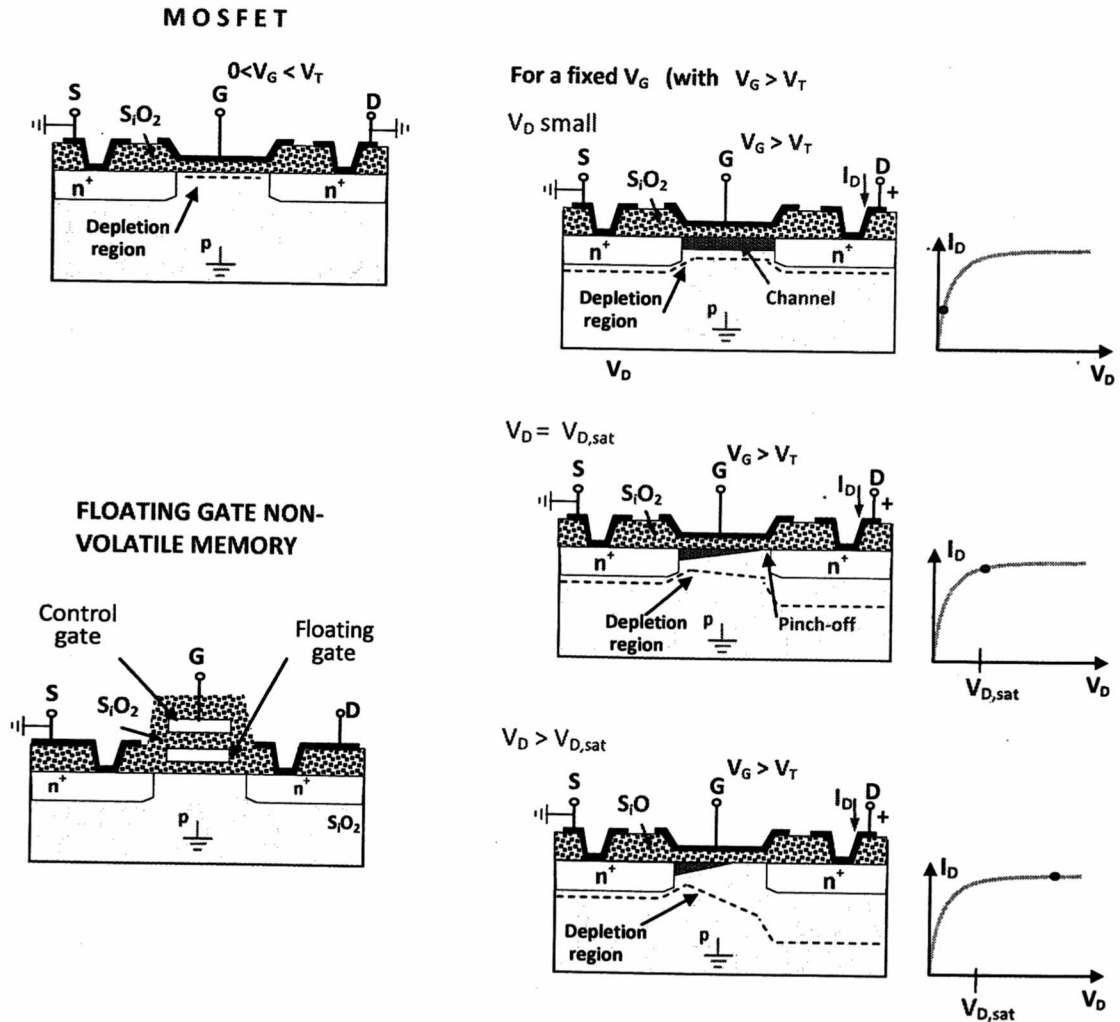


Fig. 3 Top and right sides: A p-type metal-oxide-semiconductor field effect transistor and output I - V characteristics. A conductive channel (or inversion layer) is formed between the source and drain n^+ regions when the applied gate voltage exceeds a threshold voltage V_T (the latter is characteristic of the particular MOSFET device). The current I vs drain voltage V_D response is illustrated for a fixed V_G voltage. As V_D changes, the corresponding modifications of the channel controls the operation of the transistor. **Left bottom side:** Floating gate non-volatile memory. This is basically a MOSFET that has a modified gate electrode that comprises a floating gate surrounded by insulator S_iO_2 . When a positive voltage is applied to the control gate, charge is injected from the channel region through the gate oxide into the floating gate. When the applied voltage is removed, the injected charge can remain stored in the floating gate.

The finest resolution that can be achieved by optical lithography is hampered by diffraction effects (step c-2 in Fig. 2). As a rule of thumb, the smallest feature that can be fabricated using traditional optical methods is of the order of λ/NA , where λ is the wavelength of the light being used ($\lambda \sim 500$ nm for visible light, 200 nm for uv) and $\text{NA} = n \sin\theta$ is the numerical aperture of the final objective lens used to deliver the radiation on the substrate (n is the index of refraction of the medium separating the lens and the sample). Still, the 80 nm features in the memory device shown in fig. 2g was achieved using radiation of 193 nm wavelength. However, this was possible only due to special procedures that employed a phase-shift lithography method,⁴⁴ together with special chemically-amplified photoresist materials.⁴⁵ This example reveals then that, by 1999, the semiconductor industry was already struggling to continue the trend of shrinking the size of electronic devices.

The above mentioned limitations imposed by optical diffraction triggered the development of alternative strategies, including *immersion optics* (using a fluid to fill the gap between the projection-lens and the wafer surface in order to increase the numerical aperture);⁴⁶ *absorbance modulation optical lithography* AMOL (where two different wavelengths, applied simultaneously, control the opacity and transparency of selected regions on the sample);⁴⁷ *electron beam lithography* EBL (computer-controlled SEM with means for blanking the “pencil” electron-beam and “writing” the pattern on the sample one pixel at a time,^{48,49} or using more sophisticated procedures to speed up such a process^{50,51} including hybrid EBL-lithography approaches⁵²); *ion beam lithography* (to capitalize on the reduced lateral scattering of ions, compared to electrons, within the resist film);⁵³ *soft x-ray or extreme uv-lithography* (currently the favorite 13-nm wavelength technology to replace optical lithography.⁵⁴ All these different options are well documented in Ref. 35. For completeness, let’s mention that progress following the “top-down” route also includes the fabrication of stimuli responsive polymer brushes,⁵⁵ growth of polymers from previously patterned templates,⁵⁶ and chain polymerization of monomolecular layer by local stimulation using a sharp tip.^{57,58,59} The use of a tip to delineate high resolution patterns will be addressed in the section below that describes Proton-fountain Electric-field-assisted Nanolithography (PEN).

II.1B Nano imprint lithography: Non-radiation based method to construct single-electron memory devices operating at room temperature

A quite different alternative route to overcome the lateral resolution limitations imposed by diffraction is to use no radiation at all. In this new approach, patterns on the resist are delineated not by chemical reactions triggered by incident radiation but by mechanical means instead. That is the case of *nano imprint lithography* NIL (where the resists are physically deformed using a patterned mold)⁶⁰ and *nanostamping* (where, as its name indicates, the specific features are defined by stamping “ink” to a substrate via a pre-fabricated stamp.)⁶¹ To gain some grasp on how does this approach work, let’s describe the NIL with a bit more detail.

Fig 4a shows schematically the NIL imprinting process,⁶⁰ implicitly illustrating some of its virtues, namely no diffraction limited constraints (spatial resolution limited only by the ability to fabricate a proper high resolution master mold), elimination of radiation sources and optical alignment settings (which would lead to lower fabrication cost), and, very important, a viable alternative for industrial production (many devices constructed at once). In this process, first a

compression molding (coated with a thin anti sticking layer) is used to create a shallow thickness contrast pattern on a thin resists film. Afterwards, the mold is removed, and a very directional reactive ion etching process⁶² takes place to transfer the initially shallow pattern deeply through the entire resist thickness.

The high resolution capability of the NIL constitutes a viable alternative to large scale fabrication of electronic devices with nanometer-sized features. In particular, for example, floating gate non-volatile memory devices described in the previous section (Fig. 3), where data information is represented by storing charges on the floating gate. Such regular MOSFET devices typically use thousands of electrons to define a binary "1" state. The motivation for shrinking the size of these devices is the added benefit of faster operation speed and lower power consumption.⁶³ What would be the ultimate limit in scaling down a floating gate memory? A provocative answer would be to use just one electron to represent a bit.⁶⁴ Working in this direction, early devices of relatively large dimension (~ 100 nm) required cryogenic temperatures to work with such sensitivity. However, the potential of single-electron devices for integrated circuit applications at room temperature was identified as early as in 1988.⁶⁵ A reduction in the size of the device to ~ 10 nm, so that the quantized energy level spacing becomes larger than the thermal noise, lead to an increase in the maximum operating temperature of single-electron devices to room temperature in 1995.⁶⁶ The progress has not occurred, however, without difficulties. At this scale, the normally random distribution of the individual dopant atoms within the semiconductor becomes a critical factor in determining device performance. Particularly, the dopants influence the device-to-device fluctuations in the threshold voltage (V_T in Fig. 3) that turns-on the device. It has been found that an orderly distribution of the dopant results in less fluctuations.⁶⁷ In short, for single electron operation, a small floating gate is needed to significantly increase both the electron quantum energy levels (the narrower the gate, the more spaced the quantized energy levels) and the electron charging energy (due to the small capacitance).

Design, fabrication, and characterization of room-temperature Si single-electron memories using nanoimprint lithography NIL was reported in 2003.⁶⁸ Although other approaches had succeeded using grown Si "dots" or isolated nanocrystal Si as storage dots,⁶⁹ the multi-dot nature and their random location lead to a large fluctuation of the device performance, and, hence, not amenable for large-scale integration. In contrast, in the NIL approach the floating dot is formed and self-aligned with the narrowest part of the channel (see Fig. 4b) thus giving the stored electron maximum screening ability. It is found that in these devices the *threshold voltage* shifts with the presence of additional charges in the floating gate. Also, due to the Coulomb blockade effect,^{70,71} the voltage increments needed to keep charging the gate becomes discrete (and well separated). The ultimate limit in scaling down the floating gate memory is to use one electron to represent one bit.

The NIL fabrication of a single-electron memory device, outlined in Fig. 4, comprises the following steps:⁶⁸

- i) The starting material is a silicon-on-insulator substrate composed of a single crystal silicon layer (35 nm thick) built on an amorphous SiO₂ oxide. The silicon layer is for (housing) the channel.

- ii) The single crystal silicon layer is subjected to a 2.4-nm-thick oxide thermally growth in diluted O_2 . This oxide is to separate the channel from the floating gate to be built on top; that is, it will constitute the tunneling region for electrons to be injected from the channel to the floating gate. Attaining a good quality oxide at this stage appears to be a key step to ensure a long retention time (more than 2 days) memory device. (Previous attempts⁷² using much thinner oxides ended up with a device having just five seconds retention time memory.)
- iii) A 13 nm poly-silicon is deposited by low-pressure chemical vapor deposition, which serves as the basis material to build the floating gate.

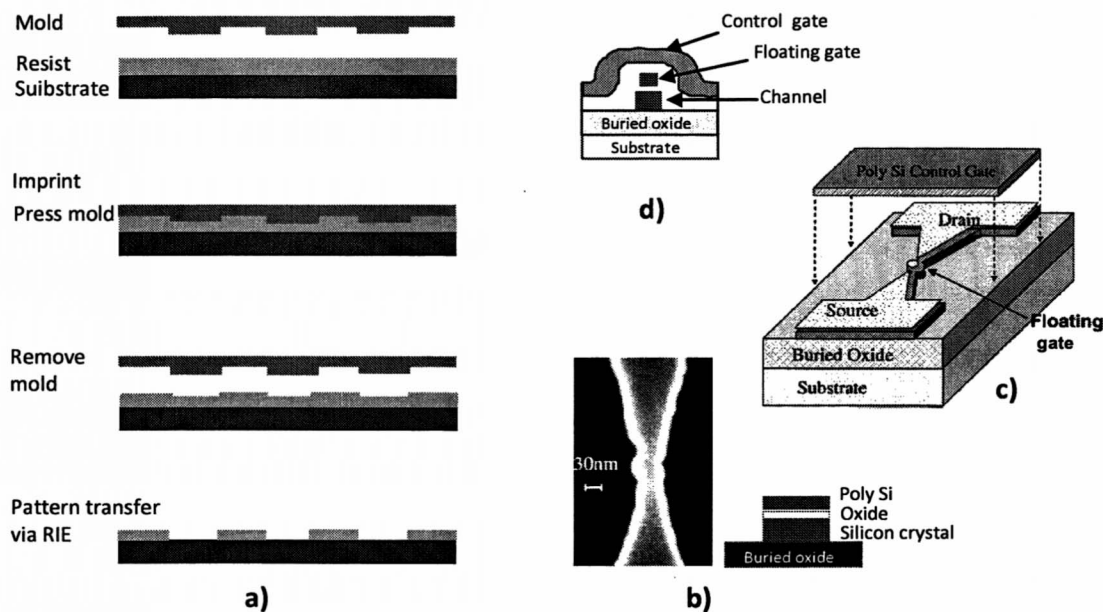


Fig. 4 Construction of single electron transistor devices via NIE. **a)** Schematic diagram of nano imprinting lithography (NIL) process. A pre-fabricated mold establishes an initial topographic contrast on the resist, which is then enhanced through the resist's thickness by a reaction ion etching process. **b)** A layered silicon-oxide-polysilicon material (with a buried oxide as a substrate) is NIE patterned using a high resolution mold. The electron micrograph shown in the figure reveals the resulting planar tapered shape morphology of the triple layer. **c)** Upon thermal oxidation, the top poly-silicon layer is completely consumed except for a 5 nm sized dot (capitalizing on self-limited oxidation). The dot serves as the memory holder. **d)** Cross section of the memory device. *Reprinted with permission from [W. Wu, J. Gu, H. Ge, C. Keimel, and S. Y. Chou, Appl. Phys. Lett. 83, 2268 (2003)]; copyright 2003, American Institute of Physics.*

- iv) The Si layer, the oxide, and the poly-silicon layer are patterned using NIL (as outlined in Fig 4a). The double-neck tapered planar geometry of the pattern (see Figs 4b and 4c) constitutes a clever geometric design that leads, upon oxidation in the next step, to the formation of a small poly-silicon "dot."

- v) A 13.6 nm oxide is thermally grown at 850 °C in dry O₂. During the oxidation of the (Si, SiO₂, poly-Si) patterned layered, the oxide grows faster in the thin poly-Si layer, and, thus, it is completely consumed except for the formation of a small dot. This occurs due to the fact that oxidation is a self-limited process.⁷³ Notice, this way the floating dot ends up self-aligned with the narrowest part of the channel !. The poly-silicon dot serves as the floating memory gate.
- vi) Plasma enhanced chemical vapor deposition of a 33-nm SiO₂ layer, in order to electrically isolate the silicon dot. The samples are then annealed at 850 °C in N₂ to improve the SiO₂ quality.
- vii) Next, a poly-silicon layer is deposited, and the control gate is patterned to a length of 3 μm, which covers the floating gate and part of the narrow channel.
- viii) After source and drain implantation, and final contact connection established, the fabrication of a memory device is achieved.

II.2 Bottom-up approach to nanotechnology

There exists a variety of examples that exploit a bottom-up route to fabricate nanostructures. Here we focused in one involving poly(4-vinylpyridine) P4VP, the same material used in the first demonstrations of PEN. This way we expose a different facet of this interesting material.

II.2A Self-organized P4VP polymer architectures

Self-organizing structural order over several length scales are possible due to several molecular interactions, including hydrophobic and hydrophilic effects, hydrogen bonding, Coulomb and van der Waals forces.⁷⁴ In the case of proteins, for instance, multiple interactions create minima in the local free energy of the molecules as a function of different chain configurations; without these interactions, entropy causes the molecules to adopt irregular shapes with little organization. However, the calculation and measurement of energy and entropy at the molecular level are nontrivial tasks, which expose the need for developing guidelines for successfully fabricating structures at different scales. Basic principles of the spontaneous organization of simple synthetic polymers have been addressed, both theoretically and experimentally.^{74,75} Self-organization typically renders just the local structures (*i. e.* the initial step of a self-organizing process). To fully exploit the opportunities offered by the symmetry of self-organized structures, and obtain materials with directional properties at different scales, additional mechanisms and interactions (hydrogen bonding and hydrophobic interaction that promote phase separation⁷⁶, electric fields,⁷⁷ topographically patterned surfaces,⁷⁸ tailored substrate chemistry,⁷⁹) have to be used. One particular example is presented next.

Fig. 5a shows a general bottom-up strategy for building ordered polymer structures, which combines **recognition** (whereby two molecules with molecularly matching complementary interactions and shapes recognize each other, thus initially forming a receptor-substrate supramolecule) with **self-organization** (forming long range structural order via additional interactions).³⁴ The process starts with a flexible polymer that has bonding sites along its backbone (in general some precautions need to be taken to avoid coiling.^{80,81}) The centers are then "recognized" by, and connect to, side groups (for example molecules with polar-heads and nonpolar-tails as shown in Fig. 5a) via complementary bonds (hydrogen bond,^{76,82} ionic

interactions,⁸³ etc.) thus forming comb-shaped supramolecules. Subsequently, such macromolecules self-organize.

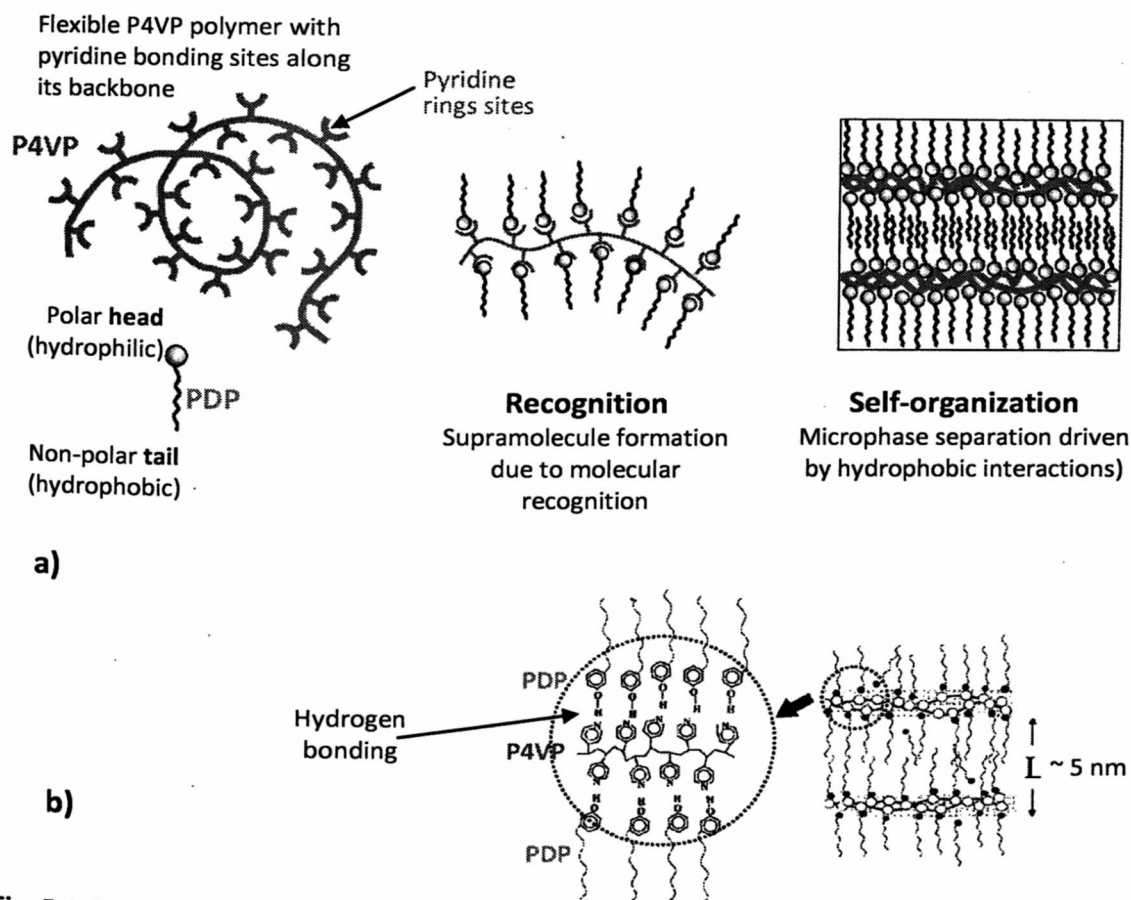


Fig. 5 Schematic description of polymer-architectures formation based on recognition and self-organization steps using poly(4-vinylpyridine) P4VP and pentadecylphenol PDP. **a)** The schematics on the top show *i)* pyridine groups distributed along the P4VP chains, *ii)* P4VP-PDP hydrogen bonding, and *iii)* lamellar structure separation between the non-polar surfactant tail layer and the polar polymer layer to which the surfactants are attached. Adapted from [O. Ikkala and G. Brinke, *Science* 295, 2407 (2002)]; reprinted with permission from AAAS. **b)** Schematic illustrations of P4VP-PDP hydrogen bonding and lamellar microphase-separated structure. Reprinted with permission from [J. Ruokolainen et al, *Macromolecules* 30, 2002 (1997)]; copyright 1997 American Chemical Society.

As a specific example, Fig 5b shows that poly(4-vinylpyridine) P4VP works well under this strategy (see also Fig 6b):

i) The pyridine rings, distributed along the P4VP polymer chains, act as recognition centers.⁷⁶ Optionally, acid molecules (methane sulfonic acid MSA in Fig. 6b) can be used to protonate these rings, which helps to stretch out the P4VP (due to the mutual electrostatic repulsion among the centers) and strengthen their role as receptor centers.⁸²

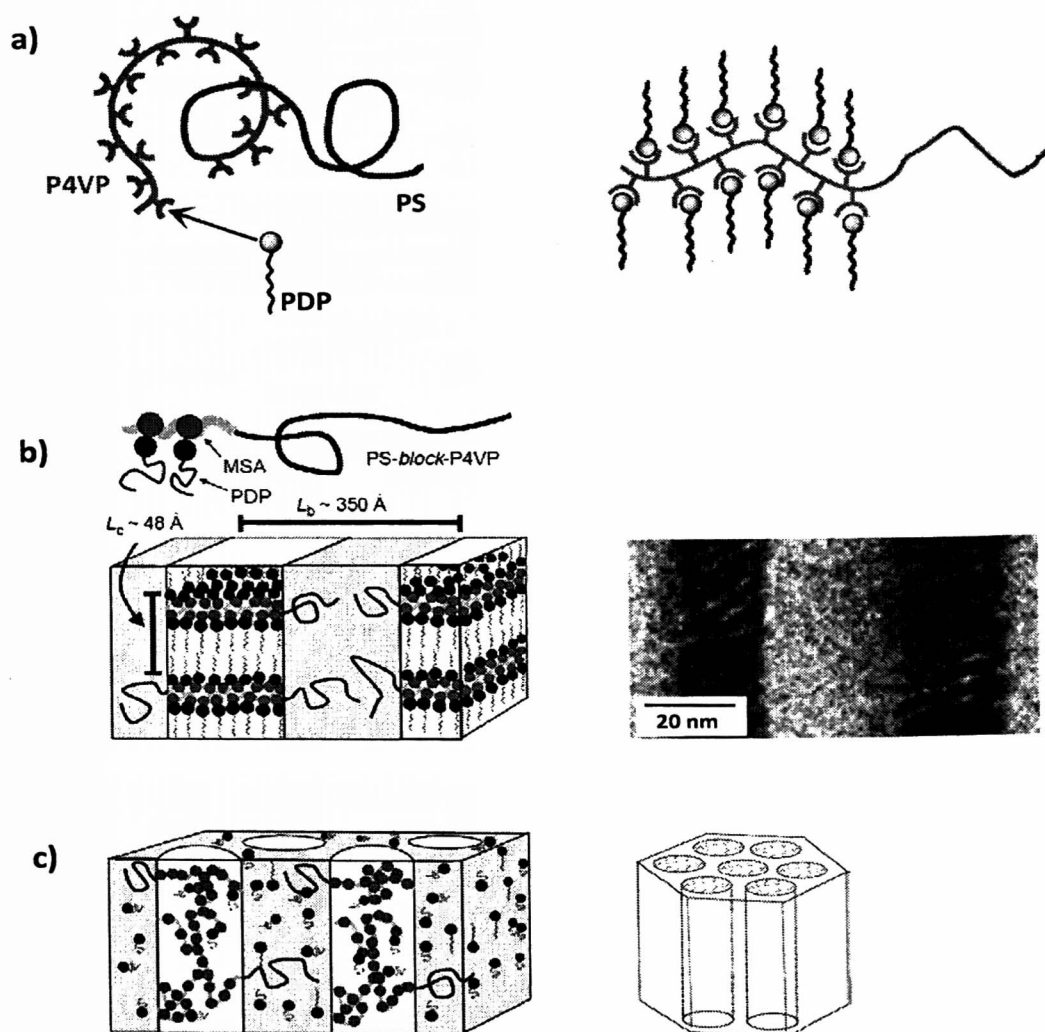


Fig. 6 **a)** Block copolymers (comprising of two or more homopolymer subunits linked by covalent bonds) allow building more complex architectures following similar procedure as the one outlined in Figure 5 above. *Adapted from [O. Ikkala and G. Brinke, Science 295, 2407 (2002)]; reprinted with permission from AAAS.* **b)** Left: Diblock copolymers (consisting of P4VP-MSA-PDP molecular complex) allow microstructural control on two length scales. Right: SEM micrograph of the complex diblock architecture. **c)** Setting the system at 150°C causes disorder in the P4VP-MSA-PDP domains, leading to the formation of hexagonal cylindrical structures (within the rest of the polystyrene and part of PDP materials). *From [J. Ruokolainen et al, Science 280, 557; 1998]; reprinted with permission from AAAS.*

ii) The fabrication process continues by adding amphiphilic surfactants (molecules with polar-heads and less-polar tails, such as pentadecylphenol PDP), which form hydrogen bonds with the polymer, (see also Fig. 5b).

iii) This procedure leads to the formation of structures with mesomorphic order of periodicity $L \sim 5$ nm (Fig. 6b.) The layered structures result from microphase separation between the “*non-polar surfactant PDP tail layer*” and the “*polar polymer layer*” to which the surfactants are attached. (Microphase separation is similar to that of oil and water; being immiscible they phase separate.)

More complex architectures can be obtained using block copolymers (polymers derived from two or more monomeric species, as opposed to a homopolymer composed by only one type of monomer. See Fig. 6a). Block copolymers are interesting because they can “microphase separate.” That is, the different monomeric components have different solubility⁷⁴ and, therefore, they tend to separate out. But, because the blocks are covalently bonded to each other, they cannot get apart macroscopically; instead they self-organize into lamellar fashion (like grill-sheets stacked one over another), thus adding another periodicity to the architecture as the ratio of the polymer length can be arbitrarily selected. As an example, Fig. 6b shows structures whose starting materials are polystyrene(PS)-block-P4VP, with the P4VP chains forming the minority blocks.⁸² Notice the two different periodicities $L_c = 48$ Å and $L_b = 350$ Å.

Additional architectures can be obtained by varying the temperature. Fig. 6c illustrates the case in which one of these ordered structures can be purposely destroyed, hence giving rise to hexagonal-cylinder porous.

III. Top-down *Proton-fountain Electric-field-assisted Nanolithography* (PEN):

Fabrication of erasable polymer nanostructures via chemical and electrical stimuli

The examples in the previous section described the use of poly(4-vinylpyridine) P4VP to build extended and complex architectures. This section focuses on complementary efforts, using Electric-field-assisted Nanolithography (PEN), for harnessing the patterning of raw P4VP films in a very localized fashion, delineating polymeric features of micron- and nano-scale dimensions. PEN could become a useful tool in the materials science and molecular engineering fields for constructing devices that rely on the transduction of environmental signals.⁸⁴ This Section III describes the scanning-based, top-down approach, implementation of PEN in some detail. Its potential outgrowth, incorporating bottom-up approach procedures, aiming at increasing fabrication speed and pattern resolution, is outlined in Section IV.

III.1 PEN working principle

PEN mirrors the dip-pen nanolithography (DPN) technique^{85,86} where patterns are created by anchoring molecules *onto* a surface of proper chemical affinity. But, contrasting such “piling molecule over molecule” approach, PEN was set to instead trigger the formation of nanostructures as a result of the substrate’s swelling in response to an external **chemical** and **electrical** stimuli delivered through a sharp scanning probe.^{29, 87} The procedure is illustrated schematically in Fig. 7. The experimental results suggest that the pattern formation on a 50 nm thick polymer film is triggered by the local injection of protons (hydronium ions H_3O^+ from an acid-coated sharp solid stylus) into the substrate. The transfer of charge from the stylus into the polymer matrix film occurs provided that both the ambient humidity is above 60%⁸⁸ and the stylus is pressed against the film with a sufficient strong force, greater than 1 μ N.²⁹ The high humidity facilitate the formation of a water meniscus around the tip-sample contact, which

works as a bridge assisting in the transportation of the hydronium ions from the tip to the substrate's surface. The high force aids the ions to penetrate and subsequently diffuse into the substrate.

The presence of swollen structures just in the regions where the probe made contact with the surface suggests that the protons, upon diffusing into the polymer matrix, protonate the P4VP pyridine centers. Although the polymer matrix is electrically neutral,⁸⁹ the electrostatic repulsion among the pyridine centers produces a net swelling effect on the polymer,³⁰ as suggested in the inset of Fig. 7.

It is also observed that, in addition to the dependence on the humidity and the contact force, the dimensions of the swollen features are also influenced by the dwell contact time (longer times allow more molecules diffusing into the polymer),²⁹ and by the application of an external electric field (for voltages in the 0 to 5 V range the height of the features vary at ~ 1 nm/V.)⁸⁷

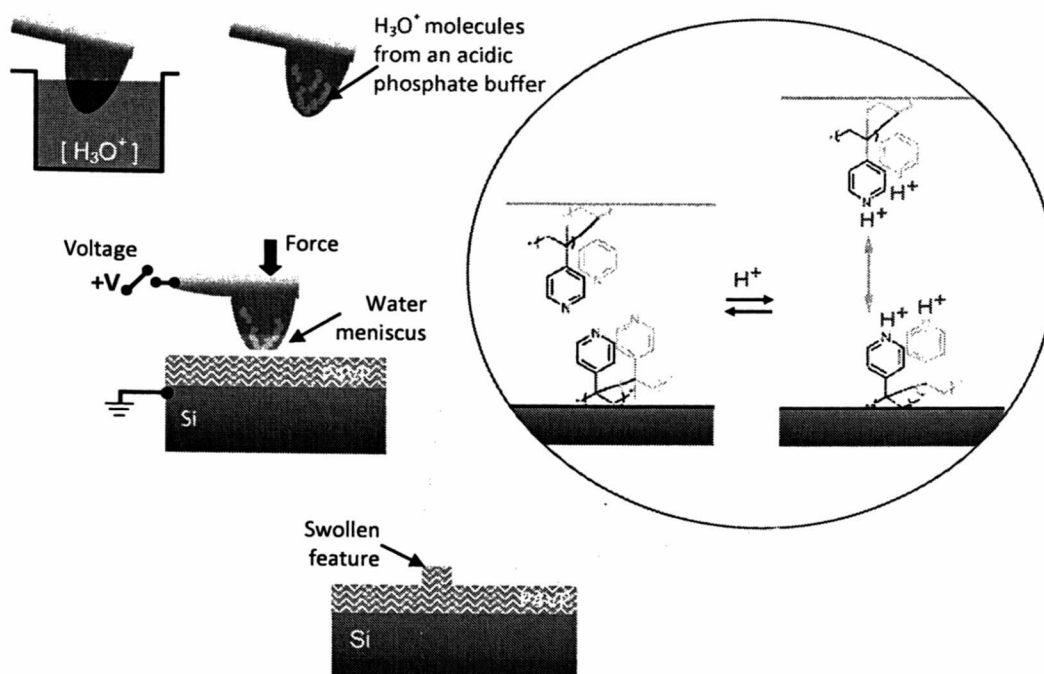


Fig. 7 PEN working mechanism. An tapered stylus (~ 20 nm apex radius) coated with acid is used as a source of hydronium ions H_3O^+ . As the probe is brought into contact with the surface, a water meniscus is formed, which acts as a bridge that facilitates the transport of ions towards the sample surface. The applied force facilitates the penetration of the ions into the polymer matrix; upon diffusion the ions protonate the pyridine groups. The inset suggest that the Coulomb repulsion between the protonated pyridine groups results into a net swelling. This working principle also implies that, upon de-protonation the swelling process can be reversed.

Since the protonation can be counteracted by adding a chemical base, the swelling mechanism can then be made reversible. The latter feature stimulates further the development

of PEN since it opens a newer nanofabrication method for biotechnology applications (possibly for creating on/off switching gates, that would allow manipulating the transport, separation, and detection of bio-molecules; the latter occurring in soft-material scaffolds that mimic closer natural bio-environments).

III.2 Sample preparation: Immobilization of polymer films to flat substrates

III.2A. Piranha-cleaning

Silicon wafers with their native oxide layer are cut into square pieces ($\sim 2\text{cm} \times 2\text{cm}$), and subsequently cleaned either by sonication in isopropyl alcohol for 15 minutes. Alternatively the samples are cleaned by immersion into piranha solution (3:7 v/v ratio of H_2O_2 and concentrated sulfuric acid H_2SO_4) for 60 min at 80°C . For this latter process, organic solvents are kept away from the piranha solution since they react violently. First, hydrogen peroxide H_2O_2 is poured into a beaker and then the H_2SO_4 is added slowly. The mixing is very exothermic and steam will be visible; if boiling signs are noticed, it would mean the pouring is going too fast. The pieces are then washed thoroughly with boiling deionized water for 90 minutes and then dried (ideally under a stream of nitrogen). This cleaning procedure creates a surface rich in functional (-OH) hydroxyl groups⁹⁰ on the oxide surface,⁹¹ which facilitate the subsequent process (described below) that uses functionalized molecular-glue to covalently attach a polymer film to a silicon oxide substrate.

III.2B Immobilization of the polymer film via PFPA molecular glue

Spin coating is the most popular method to prepare polymer thin films on flat substrates; but the physisorption character of this attachment makes the film prone to easy removal. It is highly desirable to develop more robust film immobilization methods that can make them withstand harsh environmental and processing conditions. Dr. Yan's group, at the University of Massachusetts Lowell, has developed versatile "double-side" molecular "glue" that is just one monolayer thick^{92,93} for covalently immobilizing a variety of molecules (including polymers,⁹⁴ carbohydrates,⁹⁵ and electronic materials such as graphene⁹⁶) to a variety of substrates (oxides, metals, and semiconductors).⁹⁷ When applied to polymers, this method produces films $\sim 6\text{ nm}$ in thickness⁹³ after activating of the molecular glue with uv light (but avoiding wavelengths smaller than 280 nm to minimize polymer crosslinks). Additional details of the PFPA covalent attachment process are provided below as a preparation for the route envisioned for the outgrowth of the PEN technique described in Section IV.

a. Monolayer-thick PFPA molecular glue

A general approach for covalent immobilization of polymer films to a silicon oxide substrate uses a family of penta-fluorophenyl compounds that are doubly functionalized with azido^{98,99} and silane^{100,101} groups. The silane group attaches to the substrate and the azido group attaches to a polymer molecule.

- i) The functionalization of penta-fluorophenyl compounds with the azido group (containing three nitrogen atoms) is to exploit the ability of nitrogen to get inserted into the C-H covalent bonds of polymer strands upon uv activation.^{98, 99} This first functionalization step produces perfluorophenyl azides PFPA (see Fig. 8a).

ii) The functionalization with a silane group (Si atom connected with other four functional groups), as shown in Fig. 8b, is to exploit its reactivity with, and thus covalently attach to, hydroxyl groups (-OH) found on a silicon oxide surface.^{102,103} For completeness, it is worth mentioning that detailed arrangement of silane molecules at the Si/SiO interface has been studied by IR spectral analysis.¹⁰⁴ Exploiting this reactivity, a variety of silane groups are widely used to form self-assembled PFPA monolayers.^{37,105}

Further details concerning the preparation of perfluorophenyl azides PFPA can be found in references 98 and 99. Preparation of silane functionalized PFPA is described in Refs. 92 and 93.

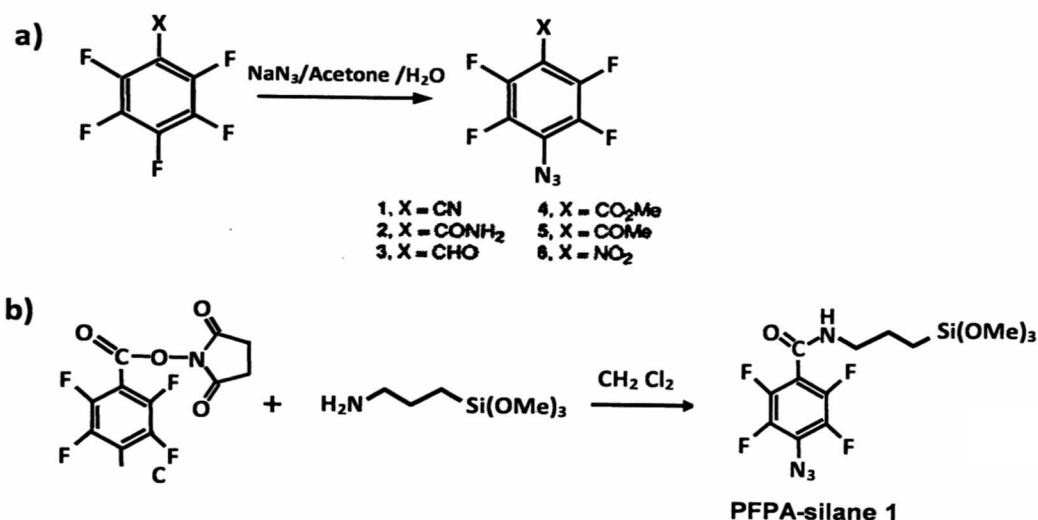


Fig. 8. Synthesis of molecular “glue” containing a reactive nitride group N₃ (which attach well to molecules containing C-H bonds, i.e. polymers, by undergoing C-H and/or N-H insertion reactions upon uv-light activation), and a silane group SiY₄ (which attach well with hydroxyl groups that are present on a piranha-cleaned silicon oxide). **a)** Synthesis of functionalized perfluoro-phenyl azides (PFPA) from their corresponding pentafluorophenyl.⁹⁸ **b)** Synthesis of PFPA-Silane 1.¹⁰⁶

b. Functionalization of silicon oxide surface with monolayer-thick PFPA-silane

A piranha-cleaned wafer (with its native oxide and hydroxyl -OH groups) is soaked in a toluene solution of PFPA-silane (5 mM) for about 4 hours at room temperature. The procedure is performed in sealed vials to minimize contact with air moisture. Subsequently, the wafers are removed from the solution, rinsed with a gentle stream of toluene, dried under nitrogen, and allowed to cure at room temperature for at least 24 hours.⁹⁹

Figure 9 shows a suggested model of the molecular arrangement that silane groups undergo near the proximity of a silicon oxide surface. The initial density of Si-OH groups is important to establish the final Si-O-Si link arrangements.¹⁰⁴ Ellipsometry measurements reveal PFPA layer thickness in the 6 Å to 11 Å range, depending on the soaking time (1 hour to 24 hours).¹⁰¹

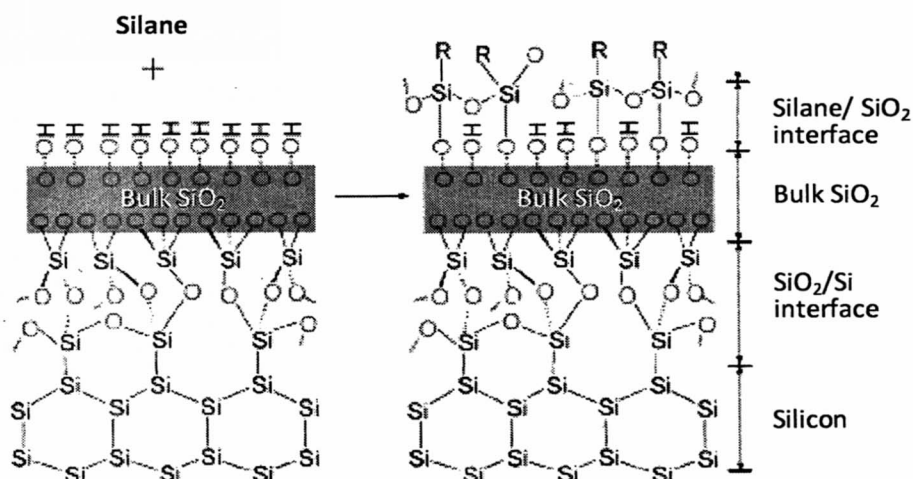


Fig. 9 The silane/SiO₂ interface. Attachment of silane groups to a silicon oxide surface. The self-assembled arrangement responds to interaction between silane groups and -OH hydroxyl groups that are present on a piranha-cleaned silicon oxide surface. *Reprinted with permission from (R. Tian et al, Langmuir 26, 4563; 2010); copyright 2010 American Chemical Society.*

c. Immobilization of polymer films on a PFPA-silane functionalized surface via either uv light or thermal activation

Once a PFPA-silane monolayer is attached to a substrate, the azido groups become exposed to the C-H bonds of the polymer molecules that are subsequently added via a spin coating procedure. Polymers in a proper solvent (toluene for polystyrene PS, chloroform for poly(2-ethyl-2-oxazoline) PEOX or for poly(4-vinylpyridine) P4VP) at typically 10 mg/mL solution concentration are spin coated at 2000 rpm for 60 seconds. The initial formation of thick layers (> 30 nm) are purposely attempted (to minimize the eventual formation of voids). The samples are then blown dry with nitrogen and are ready for the subsequent uv irradiation step.^{93, 99} A medium-pressure mercury-vapor lamp (450 W power, Hanovia) is used to trigger C-H and/or N-H insertion reactions at the PFPA-polymer interface. (A long pass 280 nm filter is used to protect the polymer from crosslinking caused by very short wavelengths).

A minimum time of 3 minutes of uv-irradiation is required to immobilize the polymer film, (the thickness grows rapidly in the first 5 minutes, and no variation occur after 20 minutes).⁹³ The final thickness (~ 2-5 nm) is independent of the starting thickness of the film right after the spin coating, which supports the argument that only the polymer molecules in the neighborhood of the PFPA azido group participate in the reaction. The latter result suggest that, if the process were performed at higher temperatures (thus allowing the polymer molecules to explore more spatial configurations and, hence, more frequently exposed to the azide groups on the surface) the efficiency of the C-H and N-H insertion reactions would increase. Indeed, that turns out to be the case. In the absence of uv radiation, just heating the spin coated polymer in an oven at temperatures in the range of 140 to 180, similar immobilization results

are obtained.⁹⁴ The formation of a polymer film, involving the insertion of N into the C-H polymer bonds is schematically described in Fig. 10.

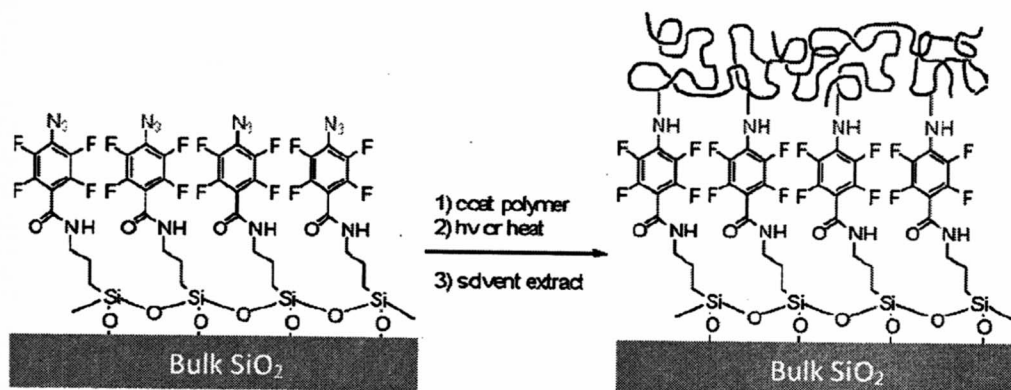


Fig. 10 Schematic description of the attachment of the PFPA azido (nitrogen) groups to polymer molecules. *Reprinted with permission from [L. Liu et al, J. Am. Chem. Soc. 128, 14067; 2006]; copyright 2006 American Chemical Society.*

III.2C Polymer films fabrication via uv-activation of crosslinks with uv light (< 280 nm)

The azido PFPA-silane functionalization method described above is suitable to immobilize monolayer-thick polymer films to silicon oxide surfaces. When thicker polymer layers are required there exists a relatively simpler alternative method¹⁰⁷ that comprises spin coating the polymer and subsequent uv light irradiation to generate crosslinks within the polymer matrix (no molecular glue is required). This method renders films of thickness ranging from a few nm to 80 nm depending on the initial polymer concentration, molecular weight, and irradiation time, which turns out to attach well not only on oxides but also on metallic surfaces.¹⁰⁷ In this process, after spin coating a polymer-solvent solution (10 mg/mL) on a clean surface, irradiation with uv light (mercury lamp, 10 mW/cm² intensity at the sample's surface, no long pass filter used in this case) for 20 minutes is sufficient to strongly attach the film. The medium-pressure Hg lamp has a 222-3673 nm spectral emission with a maximum at 366 nm. For testing purposes, when using a 280 nm long pass optical filter no polystyrene film remains after extraction with toluene, which indicates high-energy deep uv light (< 280 nm) is responsible for the immobilization of the films. The film is robust since it remains on the surface after the unattached polymer is removed by extensive extraction with the solvent. Even after 30 minutes of continuous boiling the film survives. Since the procedure occurs in oxides as well as metal substrates, the working mechanism has been attributed to a crosslinking process.

For the PEN applications we have consistently used poly(4-vinylpyridine) (P4VP). A 10 mg/mL solution of P4VP (Mw ca. 160,000) in CH₂Cl₂ is spin-coated onto the piranha-cleaned silicon wafer at 2000 rpm for 60 seconds, which yield films of thickness in the range of 50 to 80 nm, as revealed by ellipsometry measurements. The film is irradiated with a 450 W medium-pressure Hg lamp at ambient conditions for 5 minutes. The light intensity of the lamp, measured by a sensor with peak sensitivity at 254 nm, is 12 mW/cm² at the sample-plane. (The 5 minute irradiation time includes a 2 minute warm-up for the lamp to reach its full intensity.) The

unbound polymer is removed by soaking the sample in CH_2Cl_2 for 24 hours to give the cross-linked P4VP film. The thickness of the film was measured with a Gaertner ellipsometer (model L116A, with a 2 mW He/Ne laser at an incident angle of 70°), assuming a refractive indices 1.465 for SiO_2 and 1.581 for P4VP in the final estimations.

III.3 PEN top-down approach

The PEN nanoscale patterning process constitutes an extension of previously reported swelling of UV-crosslinked P4VP large-area films under acidic stimulation.³⁰ In that experiment, measurements performed before and after soaking cross-linked P4VP films in a pH 4 buffer solution indicated a 20% increase in the films' thickness (measured by ellipsometry of the dried films), while a shift in the (IR-spectra) frequency-band, associated to the pyridyl ring-stretching modes, provided a signature of the protonation. The increase in film thickness is attributed to the protonation of pyridyl groups, caused by the added acidic component to form pyridinium ions, whose mutual repulsions cause the film to expand (see the inset in Fig. 7). Triggering the polymer swelling mechanism via the delivery of acid through a sharp stylus, as outlined in Fig. 7 establishes a new way to create nanostructures. Here we outline the necessary instrumentation and typical procedure.

Atomic force microscopy (AFM) system with scanning lithography capabilities: Currently we use an AFM IX-20 model from Park Scientific, which controls the vertical position of the probe (a cantilever with pyramid-shape tip at one end, as shown in Fig. 7) while being laterally moved, in a raster fashion, along the sample's surface. The system is able to control the force that the tip exerts over the sample, as well as to establish an electrical potential difference between the sample and the tip (when using metalized probes and metallic substrates.) In a typical procedure, a first step constitutes to test the lithographic capability of the system by creating patterns via scratching a soft polymer film with the silicon tip. Fig 11 shows 'UNI' and 'VALQUI' patterns written at ambient conditions. Both were "written" in contact mode using an n-type silicon tip (MikroMasch, NSC15, with the top side coated with aluminum to increase the reflectivity of the AFM laser system), a probe-sample contact force of $1.8 \mu\text{N}$, and writing speed of 100 nm/s. For the regions where not writing was performed (*i.e.* going from one letter to the next) a velocity of $5 \mu\text{m/s}$ and 80 nN (*i.e.* less contact force). In the case of the VALQUI pattern the writing force was changed to 1250 nN. The reason for the decrease was to prevent excessive accumulation of polymer on the tip during the process, since the writing length was doubled with respect to the first pattern.

Preparation of the acidic fountain tip. A source of hydronium ions H_3O^+ is prepared out of phosphate buffered solutions, which have the remarkable property that can be diluted and still keep the same concentration of H_3O^+ .¹⁰⁸ Different pH values can be obtained by dissolving corresponding quantity ratios of sodium dihydrogen phosphate (NaH_2PO_4) and sodium hydrogen phosphate (Na_2HPO_4) in distilled water. For example, mixing 13.8 g/l and 0.036 g/l of the two salts, respectively, gives 0.1 M buffer solution of pH equal to 4.0.¹⁰⁹ This acidic solution serves as the source of hydronium ions. Subsequently an AFM tip of relatively high spring constant ($k = 40 \text{ N/m}$) is coated by simply soaking the probe into the buffer solution for about 1 minute (Fig. 7a) and then allowing it to dry in air for 10 minutes or, alternatively, by blowing it

with nitrogen (Fig. 7b.) Such fountain tip constitutes the probe for delivering hydronium ions very locally over targeted sites on a responsive material substrate.

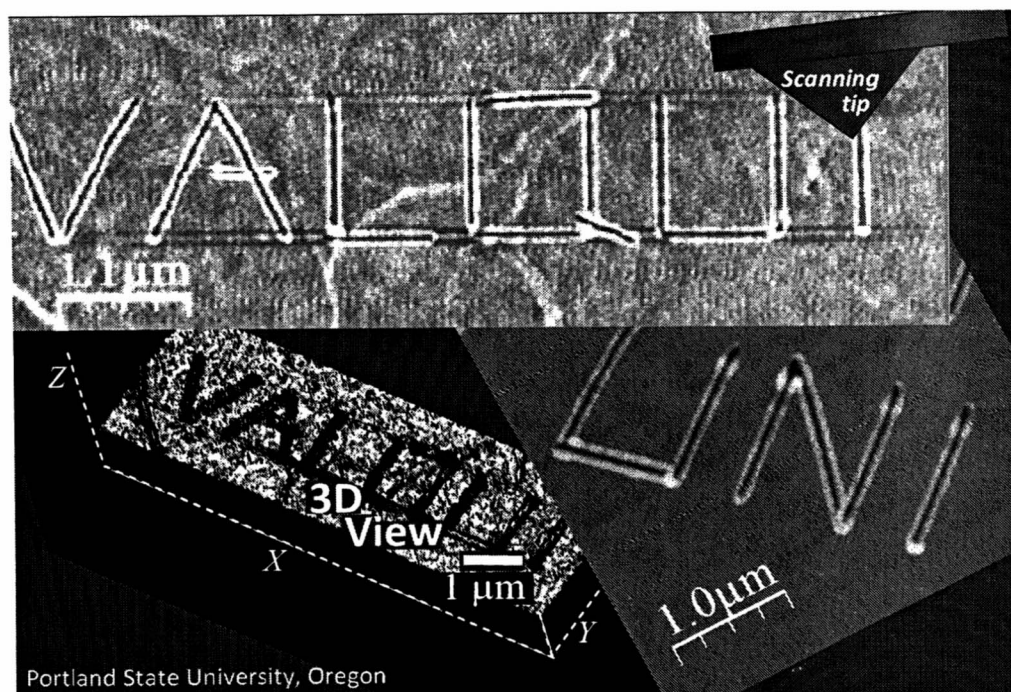


Fig. 11. Pattern formation by scratching a soft polymer film with an AFM probe. The logos are aimed to provide tribute to Professor Holger Valqui, in recognition to his long standing contribution to science and education at Universidad Nacional de Ingeniería, Lima-Perú.

Pattern Fabrication. The experiments described below used as starting material to 50 nm thick P4VP polymer film fabricated by crosslink activated by uv radiation (*i. e.* no PFPA molecular glue was used).

i) First, the atomic force microscope (AFM), with the “ink” loaded tip, is set in “tapping imaging-mode,” in which the cantilever probe undergoes oscillations perpendicular to the sample’s surface. To obtain an initial knowledge of the “blank paper” (a UV cross-linked P4VP polymer film), an image is taken at a relatively fast lateral scanning rate of 6 $\mu\text{m/s}$ (0.5 Hz line scan rate, 6 μm line width, 100 pixels per line). It is found that the use of such modest high rates prevent the transfer of buffer molecules into the substrate.

ii) Once it is verified that the scanned area is free of unwanted topographic features, the AFM microscope is switched to “contact imaging-mode” for patterning formation under physical parameters controlled by the operator, namely contact forces of the order of 1 μN , ‘writing’ speeds up to 400 nm/sec (0.05 Hz line scan rate), and fixed bias voltages up to 5V. Patterns of different planar morphologies can be generated with pre-programmed software designs, which guide the voltage-controlled XY lateral scanning of the tip while an electronic feedback-control

keeps the probe-sample contact force constant. In our case, using a Park Scientific AFM IX-20, the sample rests on a XY piezo scanner stage (equipped with strain-gauge sensors for overcoming piezoelectric hysteresis via another internal feedback control⁴³), while the tip is held by an independent piezoelectric z-stage. This is a convenient way of decoupling the sample's horizontal XY scanning motion from the probe's vertical z-displacements, which avoids potential mutual "cross talk" between the lateral and vertical motion of the tip.

iii) Finally, once the pattern process is completed, the AFM microscope is switched back to the tapping mode for topography imaging in order to verify whether the patterns have been formed.

Figure 12 shows schematically a sequential process of the pattern formation, as well as a 3D view of an actual line-pattern created exploiting the responsive characteristics of P4VP. The cross-linked film swelled only in the area where the phosphate "ink" was intended to be delivered by the scanning probe while applying a contact force F of 1500 nN and a bias voltage V of +5 Volts. The experiment was performed at ambient condition (local humidity of 60%). It is widely accepted that a ~ 20 nm thick layer of water (and other potential contaminants) is present on the surface of any sample at ambient conditions. The image demonstrates the ability for sequentially creating an initial pattern (while scanning the probe in contact mode) and then imaging (in tapping mode) the resulting sample topography with the same ink-loaded tip.

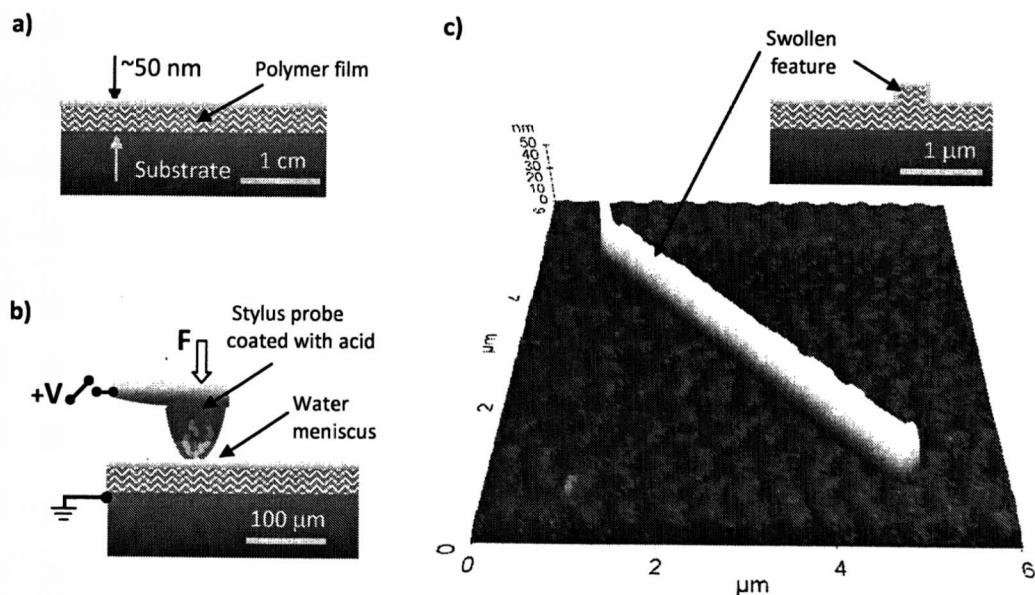


Fig. 12 Top-down PEN patterning with responsive polymers. The diagram shows the process at different length scales. **a)** Fabrication of a 50 nm thick film of macro-scale dimension area (cm dimension). **b)** Creation of fine structures using a sharp stylus (apex of ~ 20 nm) loaded with acid "ink", and scanned across the sample surface in contact mode under a pressing force and a bias voltage. Swollen features of desired lateral morphologies are created as protons are transferred from the probe's tip to the substrate. **c)** Imaging of the resultant pattern in tapping-mode imaging modality using the same ink-loaded tip.

The line profile in Fig. 13 reveals that the line-feature is approximately 25 nm in height and 500 nm in width. This height represents a 50% local increase in thickness, a substantial swelling that reveals the hydrogel characteristics of the pattern. A hydrogel refers to a flexible (typically) hydrophilic cross-linked polymer network and a fluid filling the interstitial spaces of the network. The entire network holds the liquid in place thus giving the system a solid aspect. But contrary to other solid materials, these wet and soft systems are capable of undergoing very large deformation (sometimes greater than 100%). Hence, the hydrogel-type structures fabricated via PEN with P4VP constitute an advantage of the technique for replicating structures that closely mimic natural bio-environments. Indeed, the role of gels in living organism cannot be exaggerated. As it is well put by Osada and Gong,¹¹⁰ living organisms are largely made of gels (mammalian tissues are aqueous materials largely composed of protein and polysaccharide networks,) which enables them to transport ions and molecules very effectively while keeping its solidity.

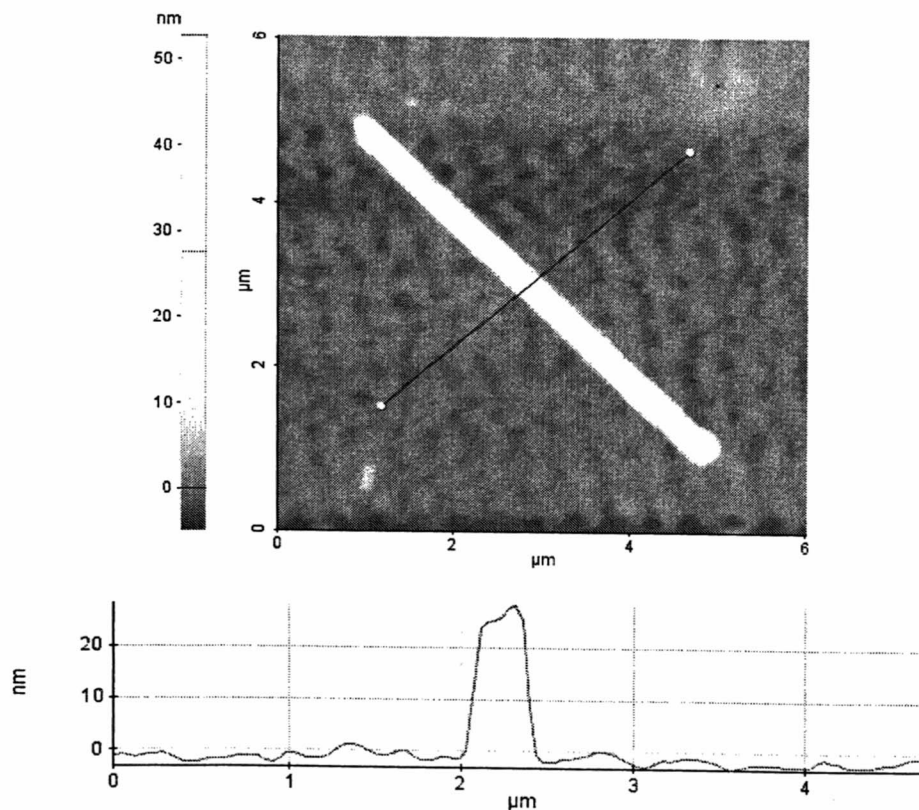


Fig. 13 Hydrogel characteristics of the pattern. Top view and line profile of the line-feature created via PEN with a 50 nm thick P4VP polymer film as starting material. The substantial swelling of the feature reveals the hydrogel characteristic of the pattern, which is suitable for creating structural environments that mimic closer natural biological environments.

The dimension of the PEN patterns is influenced by the environmental humidity. By increasing humidity, one expects an increase in size of the water meniscus around the tip,¹¹¹ and hence an increase in the amount of the hydronium ions ink that can be transported from the tip towards the sample's surface. PEN pattern formation in P4VP films under different humidity levels has been well documented by C. Maedler.⁸⁸ Dot-like structures were created by keeping the probe in contact with the sample during different dwell times and at different relative humidity inside of the chamber containing the AFM system. By starting at 60% RH and subsequent continuous decrease of the RH, dots were written at 60%, 50% and 40% RH, respectively. No structures could be created for RH below 30%. The results are summarized in the graph in Fig. 14. The formation of no features below 40% of relative humidity suggests that the transfer of the buffer molecules from the tip to the surface requires a sufficient water meniscus between tip and sample.¹¹² Buffer molecules probably dissolve in the water before diffusing into the P4VP polymer. Without dissolution no molecules can be transferred to the substrate. With decreasing RH the height of the structures decreases, which indicates that the water meniscus becomes smaller.¹¹² A smaller water meniscus leads to fewer dissolved buffer molecules and therefore a lower rate of protonation of the pyridyl groups.

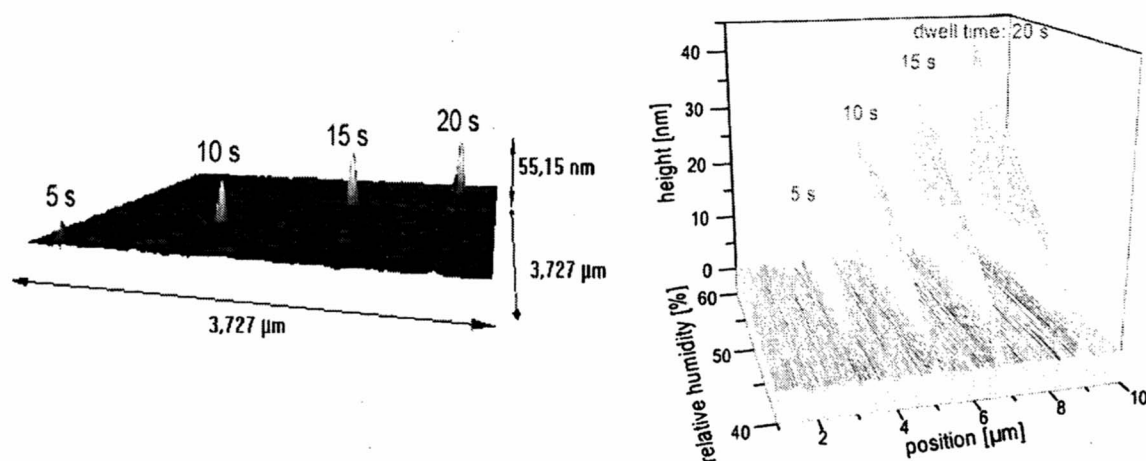


Fig. 14. Humidity dependence on pattern formation a) Dot-like structures form by leaving the acid-coated tip in contact with the substrate for different amounts of time. The dwell time is indicated above the peaks. b) Pattern height against dwell time and relative humidity. No pattern were possible to create below 40% humidity. *Reprinted from (C. Maedler et al, Proc. SPIE, 7364, 736409-1, 2009), and reproduced with permission of the SPIE.*

The fabrication process is also affected by the force that the the tip exerts against the surface as well as by the application of an electric field between the silicon tip and the silicon substrate (see setting in Fig. 12b). This has been well documented by Xiaohua Wang, whose results are summarized in Fig. 15.⁸⁷ The experiments were performed at ambient environment 23-26°C and 45-55% relative humidity. Notice that the application of a bias voltage alleviates the need of applying stronger forces (the latter could damage the polymer). For a 5V bias voltage, structures can be fabricated with forces as small as 0.5 μN , which otherwise wouldn't

be able to be created. We attribute the applied voltage helps to drive the hydronium ions into the polymer matrix; when a negative voltage is applied to the probe no patterns are formed. Typical bias voltages are up to 5 V, which when applied through a 50 nm film sets a strong electric field $\sim 10^6$ V/cm (still leaving the film apparently undamaged.) Provided that the humidity is above 40%, the patterns are consistently fabricated with this PEN method. On the other hand, the application of a force appears to facilitate the insertion of the hydrogen ions into the polymer matrix. Although the high humidity facilitates the transport of the ion from the tip to the surface, at the same time it may cause the ions to diffuse along the sample surface. A gentle indent of the probe into the substrate may create a channel for the ions to get immersed into the polymer matrix and diffuse towards the pyridine center of the P4VP molecules.

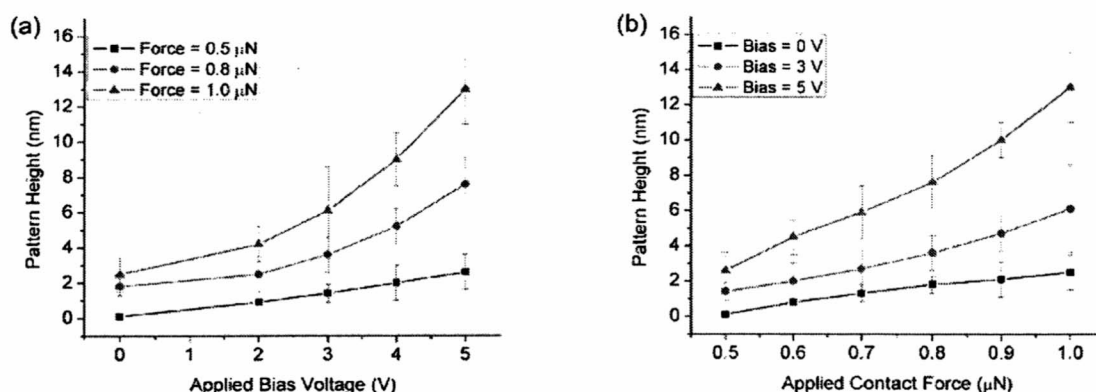
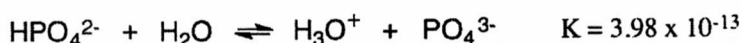
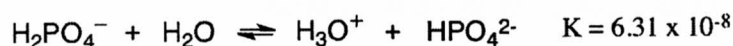


Fig. 15. Force and bias voltage dependence on the pattern formation (a) Pattern height (nm) vs. applied bias voltage (V) at various fixed contact forces. (b) The pattern height (nm) vs. applied contact forces (μN) under different constant bias voltages. Reprinted with permission from [Xiaohua Wang et al, ACS Appl. Mater. Interface **2**, 2904; 2010], copyright © 2010, American Chemical Society.

Although the results given above favor the hypothesis of protonation, would the hydronium ions be the only ones diffusing into the polymer network? To test this hypothesis C. Maedler *et al* investigated further the swollen structures.⁸⁹ Armed with Kelvin Probe Force Microscope as a tool,¹¹³ and following up studies of deposition of charges in silicon,⁸⁹ Maedler *Diplom* they corroborated to have the sensitivity for monitoring the presence of a net charge in the polymer structures, if any. But they found none on the P4VP polymer films. This implies that not only the hydronium ions penetrate the polymer, but they bring with them their corresponding counterions (see expression below) thus keeping the charge neutrality of the sample. If the fixed pyridinium cations were the only ions present there would be an exceedingly large electrostatic repulsion, but such interaction is partially screened by the presence of counterions resulting from the dissociation of the phosphate salts in water,



Here K stands for the corresponding dissociation constants.

IV Bottom-up *Proton-fountain Electric-field-assisted Nanolithography* (PEN):

Self-assembly fabrication of nanostructures with responsive characteristics

The PEN top-down approach described in the previous section offers a convenient platform for fabricating mesoscopic (hydrogel-type) stimuli responsive structures with fine precision. Such an ability to control the dimensions of patterns with 1 nm per volt precision is suitable, for example, to study the size-dependence properties of these structures. But to go beyond the study of fundamental properties and consider massive device production, both the fabrication speed and the still relatively thick dimension of the patterns are issues that need to be addressed in order to make the PEN technique competitive with other nanofabrication techniques. In effect,

- i) PEN nanostructures are currently fabricated one at a time in a serial fashion and at a speed of 400 nm/sec, which is too slow by commercialization and industrial applications standards.
- ii) The patterns lateral dimension is limited by diffusion. Basically, upon their injection into the polymer matrix, the hydronium ions naturally diffuse towards the pyridine centers of the P4VP matrix (then the mutual repulsion of these centers causes the polymer to swell). It is the lateral diffusion, then, what limits the creation of narrow features. Even though features with 60 nm line-width have been created with this top-down approach, better control to create finer line-width features would be desirable.

This section explores strategies to overcome these two limitations. Essentially, we propose to use a fabrication procedure based on the self-assembly of P4VP molecules onto a surface that has been previously chemically-functionalized with high resolution. On one hand, a self-assembly approach would inherently assure a fast and massive fabrication procedure (provided that the chemical functionalization is also fabricated by a fast method). On the other hand, pre-functionalizing the substrate with high lateral resolution would ensure the attachment of responsive polymer material only to pre-determined well-localized sites; hence when exposed to an acid bath the diffusion of the hydronium ions would be intrinsically avoided. In addition, given the reversibility of the P4VP swelling, the nanostructures could be erased upon immersion in a base-chemical bath. In short, we pursue the design of chemical functionalization methods that can be implemented with high resolution and ideally also at very fast pace. Ultimately we envision fast method for fabricating erasable nanostructures.

Addressing the high-resolution and speed fabrication aspects, this section is organized as follows. The first part of this section outlines a general view of some thermal and optical routes available for the chemical functionalization of a substrate with high resolution. For the optical route, conventional (far-field) and near-field illumination modalities are considered. The far-

field option is included for its additional capability to undertake the chemical functionalization at fast pace (when implemented with very short-pulse illumination and exploiting non-linear thermal material responses). The near-field option offers high resolution although its scanning implementation limits its fabrication speed; still different near-field illumination strategies could render the technique very useful. Here far-field refers to conventional optics where the sample is illuminated through a lens located at a distance greater than the wavelength; hence the name 'far-field'. Near-field refers to a setting where the sample is illuminated through a ~ 100 nm aperture located at a distance much smaller than the wavelength from the sample; hence the name "near" field; see Fig. 17. The second part offers a background on the working principles behind the highly localized thermal effects produced by the incidence of femto-second laser pulses on a material. The high-gradient temperature profile produced by mechanic/thermal contact is also mentioned. This background helps to put into context the suggested bottom-up implementation of PEN. The third and fourth parts describe in more detail the implementation of the high resolution chemical functionalization.

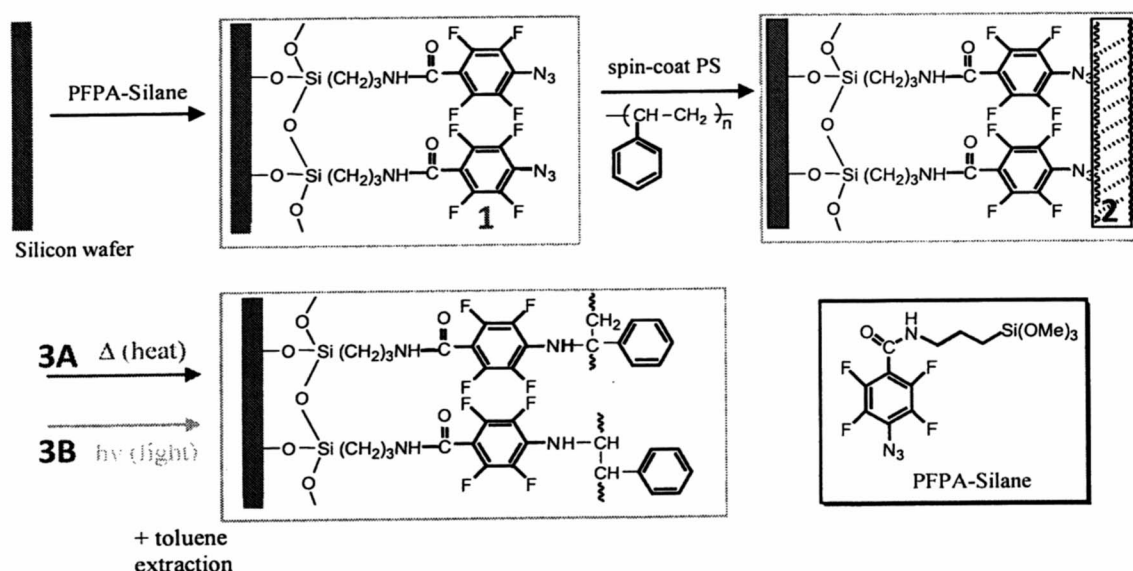


Fig. 16 Schematic (macroscale) procedure for covalent attachment of polymer molecules to a surface functionalized with PFPA monolayer "glue". (1) The silane group reacts with the native SiO_2 built on the silicon wafer substrate thus forming a self-assembled monolayer. (2) Polymer is spin coated. (3) Upon (thermal or uv) chemical activation, perfluorophenyl nitrenes are generated and undergo C-H and/or N-H insertion reactions with neighboring polymer molecules; a polymer coating is formed.^{92,93} In a nano-scale approach, the attachment of polymer molecules at more localized sites can be achieved by implementing step 3 using far- and near-field optical tools (as outlined in Fig. 17).

IV.1 High spatial resolution chemical functionalization

The new strategies proposed here for the outgrowth of the PEN technique build upon the covalent-immobilization techniques described in section III.2B above, particularly the attachment of polymer films to a silicon oxide surface. One of those methods uses PFPA as a monolayer-adhesive sandwiched between the substrate and the polymer film; while the PFPA's

silane group attaches to the sample surface and (under either ultraviolet-radiation or thermal activation) the PFPA's azido group grabs a polymer molecule in the film (as outlined in Figs. 10 and 16).⁹⁹

The variation to be introduced in this fabrication process lies in step-3. By integrating high spatial and temporal resolution optical techniques it would be possible to engineer the functionalization of PFPA (and the P4VP molecules) on well-localized sites on a sample substrate (Fig. 17a):

- One option is to illuminate the sample in a far-field modality with very short pulses (750–840 nm, 9.5 nJ per pulse, 25 fs pulse duration) in order to establish high-gradient temperature profiles on the substrate that is in intimate contact with the PFPA layer. An expected Arrhenius-type temperature dependence of the PFPA chemical activation ($\sim Ae^{-E/kT}$) implies that the thermal gradient could produce a more localized chemically activated PFPA region compared to the illuminated region. In other words the laser power can be selected so that only a small section of the Gaussian diffraction-limited focus exceeds the required intensity to trigger the chemical reaction. This short-pulse illumination method would inherently lead to high lateral resolution. The far-field implementation is for attaining a fast fabrication process since the illuminated spot can be rapidly scanned using resonant mirrors.^{114,115}

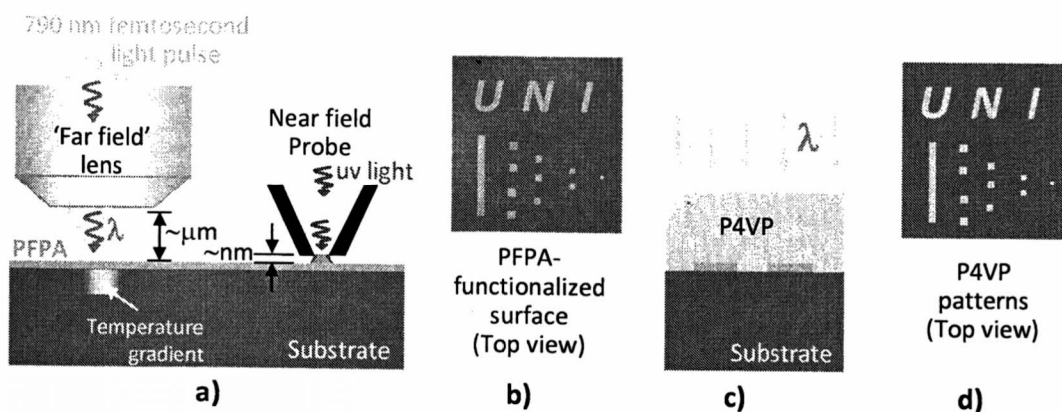


Fig. 17 Schematic description of a self-assembly method for fabricating nanostructures using stimuli-responsive P4VP molecules. **a) High-resolution opto/thermal PFPA functionalization of the substrate.** Sample is illuminated through a conventional (far-field) objective lens with femtosecond pulses to establish non-linear thermal absorption on the substrate. The thermal gradient produces a more localized chemically activated region of the PFPA layer than the illuminated region; hence the higher resolution. Alternatively, the sample is irradiated through a sub-wavelength aperture using a near-field probe (see also Fig. 18). A scanning of the sample relative to the optical probes allows high resolution pattern features. **b)** After sonication, a functionalized surface results. **c, d) Self-assembly organization of stimuli-responsive P4VP molecules.** In step c) the functionalized surface is spin coated with, or simply immersed into, a solution of P4VP; a far-field optical activation leads to a covalent attachment of the P4VP molecules to the PFA centers. **d)** After toluene extraction, only the P4VP molecules covalently attached to the surface remains. Notice the more localized illumination approaches allow attempting the immobilization of P4VP down to, potentially, single molecules.

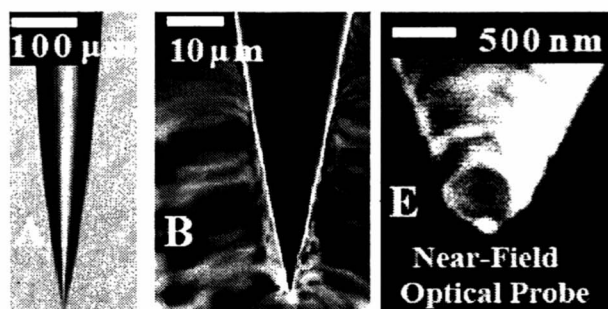


Fig. 18 Sequences of the fabrication of a Near-field optical probe. **A.** Optical image of a commercially available optical fiber glass (120 μm cladding, 8 μm core diameter) that has been tapered via chemical etching.¹¹⁶ **B.** SEM-image of an aluminum coated fiber glass probe. Since the conditions for total internal reflection in the fiber has been spoiled in the tapered region, the metal coating is to prevent the leakage of light and channel the light coupled at the back of the fiber towards the apex region. **C:** A focused ion beam is used to carefully truncate the apex of the probe thus leaving an aperture.¹¹⁷

- Another option is to illuminate the sample with uv-light through an apertures of sub-wavelength dimension (using a near-field probe like the one shown in Fig. 18),¹¹⁷ provided that the aperture is placed very near the surface (~ 10 nm) to avoid diffraction; hence the name “near-field” illumination. This method offers high resolution implementation

It would be interesting to establish a systematic comparison between these two opto/thermal covalent activation methods for immobilizing PFPA molecules onto a substrate. Harnessing the functionalization of a surface with PFPA is key for the success of the project, for the remaining PEN fabrication steps would become straightforward: the post self-assembly attachment of the stimuli-responsive P4VP molecules onto the PFPA centers will simply require a spin coating process or, even simpler, jut dipping the functionalized substrate into a solution containing P4VP molecules (Fig. 17).

In the following sections we describe in a bit more detailed the two far-field and near-field optical approaches to chemical activation. But first we provide some background on the highly localized thermal effects on materials produced by very short pulses in order to provide the context underlying the proposed high-resolution chemical functionalization techniques.

IV.2 Highly spatially-localized thermal effects

Highly localized thermal effects on materials are associated with femtosecond pulses excitations as well as with nonlinear thermal responses as a result simple contact with heated probes. In this section we briefly outline these two aspects.

IV.2A Localized thermal ablation effects produced by femtosecond pulses

The remarkable highly spatially-localized thermal effects produced by ultra-short (sub picosecond) laser pulses was evidenced in the initial experiments of polymer ablation in 1987.

¹¹⁸ Pulses of 160 fs etched holes on polymethylmethacrylate (PMMA) with sharp edges and no sign of collateral thermal damaged. Thermal diffusion was absent and there was virtually no solid deposit of the polymer at the sides of the hole. The interpretation, however, followed some controversial discussions. Since PMMA has negligible absorption at 308 nm wavelength the result was initially attributed to multi-photon excitation. But experiments in 2003 ruled out this hypothesis on the basis that the optical breakdown is independent of the light polarization.

¹¹⁹ Additional explanation would surface later on.

Previous to these experiments, the accepted principle was that laser-induced electron avalanche ionization (where electrons with a sufficiently high energy cause ionization in successive collisions, thus creating a new generation in the avalanche) as the mechanism that determines the breakdown threshold in pure transparent crystalline or amorphous solids and liquids. Once sufficiently dense plasma has been created, the joule heating losses in the strongly colliding electron gas become so large that internally localized melting and/or evaporation takes place. The surrounding matrix may be subsequently damaged by shock waves induced by thermal stresses.¹²⁰ The new paradigm brought by excitation with short pulses is that sub-picosecond duration time is significantly shorter than the time scale for electron energy transfer to the lattice. As a result, damage caused by sub picosecond pulses is characterized by ablation, with essentially no collateral damage.¹²¹

But what produces the first initial seed for avalanche ionization to take place? The question is daunting since multi-photon excitation has already been ruled out. A clue comes from the remarkable feature that more deterministic ablation breakdown thresholds are obtained when using femtosecond laser, which contrast with the more random ablation thresholds measured when using longer pulses;¹²² that is, ablation thresholds measured from place to place along the material is more uniform when using short pulses. For example a row of holes created in the surface of Corning 0211 glass by single femtosecond pulses (at 527-nm wavelength, average pulse energy of 7.5 nJ, and focused by a 1.3 NA objective on the distal side of the cover slip) displayed a striking reproducibility in size. Such reproducibility in the size of the features combined with the fact that each hole was of nanometer- scales indicates that the initial charge carriers that seed for avalanche-ionization must have been created in a very reproducible manner. It is surprising it had occurred in a heterogeneous material such as glass, in which the band gap varies considerably from spot to spot. Pre-existing carriers cannot explain such uniformity, since the ablation of holes of such small scale (< 20 nm) and sharp edges (~4 nm) with regularity and precision would require initial free-electron densities far higher than present in large-band-gap materials (an electron density in excess of 10^{18} e/cm³ would be necessary to ensure at least one seed electron in the volume of a hole). The multi-photon excitation had also been ruled out. Such pre-existent carriers would vary then randomly along the sample, subjected to a Boltzmann statistics; different threshold energies may occur then at different sites, just because this random effect. An accepted explanation for the reproducibility of the features is that when using femtosecond pulses then the tunneling (Zener effect) of electrons from the valence band become more effective to contribute as initial seeds due to the high electric fields; the uniformity of the valence-electrons population is then attributed to be

responsible for the uniformity of the ablated regions. In short, an accepted trend of thought is that the deterministic character of the optical breakdown is dictated by the material valence-electron spatial uniformity; in good-quality optical material the energy gap may vary substantially over small scales but the valence-electron density, which is proportional to the atomic density, is extremely uniform.¹¹⁹

Another important feature to take away from these experiments is that the deterministic breakdown threshold (producing sharply ablated features) and the Gaussian intensity profile of a laser pulse (exponential variation of the light intensity) implies an exponential dependence of the size D of the breakdown zone on the pulse energy. Joglekar et al reported¹¹⁹ that $D = \sigma[8\ln(E/v)]^{1/2}$, where D is the diameter, E is the pulse energy, and v and σ are fitting parameters: v gives the threshold ablation energy. One could extrapolate from these results an exponential dependence of the local temperature established by the fast-pulse excitation, which gives indication on the high-gradient temperature gradient that can be exploited to activate chemical reactions by adjusting the experiment to use lower incident power.

IV.2B Modification of physical properties by sub-picosecond pulses

In addition to the very localized ablation of hard materials by sub picosecond pulses,^{119,123} there exists also evidence of the their effects on just the chemical structure modification of polymer films (*i. e.* with no ablation). In one case, for instance, the fast excitation process led to a change in the film's wettability properties even though there were not changes in the film's roughness (the other possible channel for changing the film's wettability properties).¹²⁴ The non-linear absorption of fs laser pulses was ascribed to photo-degradation, distinct from the thermal-degradation associated with linear absorption when using wider (ns) pulses¹²⁵. The short (fs) pulses allowed reaching power absorption threshold-limits for increasing concentrations of C=O polar bonds (over the non-polar C-O bonds), hence the modification of the film's wettability properties. On the other hand, control of the fs pulse energy can avoid material ablation and still causing changes in the material's physical properties. For example, (20-50 nJ) pulses with energy levels below the ablation threshold where use to locally increase the refractive index on fused silica, the irradiated regions also developing lower thermal conductivity. Such results have been interpreted in terms of a localized densification of the regions exposed to the fs pulses.

IV.2C High-resolution chemical activation via high-gradient temperature profiles

Spatially-confined activation by high temperature gradients has been put into evidence by Cacialli et al,¹²⁶ using a setting where a heated stylus is laterally scanned while in mechanical contact with a soluble polymer film, poly(p-xylene tetrahydrothiophenium chloride)(PXT). It was known that this precursor polymer can be turned into insoluble poly(p-phenylene vinylene) (PPV) upon the application of ultraviolet radiation at the selected places, thus providing an avenue for creating patterns. It turns out that a heated stylus could equally work well. What is surprising in this work is that, despite the blunt morphology of the stylus, nanometer-sized PPV patterns could be created. These results have been explained in terms of the high temperature gradient on the sample and the strong temperature-dependence of the PXT-to-PPV reaction rate. With a perspective of large scale production, a potential drawback of this approach is that the "writing" tip, being in intimate mechanical contact with the polymer sample, would be

subjected to contamination; the scanning process could be detrimental to samples constituted by monolayer films (causing scratches) and impose restriction on the writing speed. As an alternative, we propose establishing high temperature gradients through a far-field illumination of the substrate (stylus-sample mechanical contact is avoided), as indicated in Fig. 17a. In short we will pursue high resolution PFPA functionalization using a far-field optics approach to establish a gradient temperature on the underlying substrate.

IV.3 Thermal activation of PFPA molecules with femtosecond pulses

The above background helps to put in context the proposed ideas for implementing PEN in a top-down approach, which can now be simply listed.

IV.3A Far-field approach to attain high-gradient temperature profiles on the substrate

An avenue to achieve high resolution chemical functionalization of PFPA monolayer is by establishing high temperature gradients on the underlying glass or silicon substrate using far-field illumination optics. As shown in Fig. 17a, fs laser pulses are employed to create nonlinear absorption on the silicon substrate. The non-linear process on the substrate, under conditions below ablation threshold, should render a local high-gradient temperature profile to activate the PFPA very locally. Although not shown in the figure, an added advantage of this far-field optical setting is that the illuminated spot can be rapidly scanned using resonant mirrors.^{114, 115}

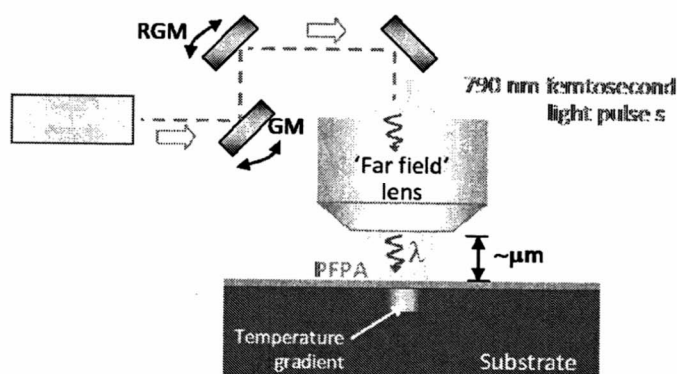


Fig.19 Local activation of PFPA molecules with high lateral resolution by exploiting high-gradients temperature profiles using femtosecond pulses. The spot is scanned in in two dimensions with the use of a galvanometer mirror (GM) and a resonant galvanometer mirror (RGM).¹¹⁵

IV.3.B Local heating activation with a near-field probe

Alternatively one can use metal-coated near-field probes to locally heat the sample. The apex of a NSOM probe (Fig. 18) reaches temperatures well above the ambient conditions depending on on the light intensity coupled into it. Placing the apex at a nanometer distance from the PFPA monolayer (under feedback electronic control) the local high temperature at the apex (~250°C) could activate the PFPA molecules located underneath (see Fig. 17b and Fig. 20). A systematic control to heat the probe was first reported in 1995¹²⁷ Shortly after, Stähelin et al. established that the axial temperature distribution at the probe apex was non-uniform.¹²⁸ The

temperature at the apex of the probe was estimated to be 100 °C when 1 mW input power was coupled into the back of the fiber holding the probe. Coupling 3.3 mW of laser light into the fiber resulted in a maximal temperature of 230 °C at the apex. This temperature range is well within the range of temperatures at which PFPA is thermally activated (as described below), which it can be exploited to create patterned polymer nanostructures by thermal activation. This feature is attractive and important especially for polymers such as isotactic polypropylene and poly(ethylene glycol) where heat treatment is essential for successful film immobilization.¹²⁹

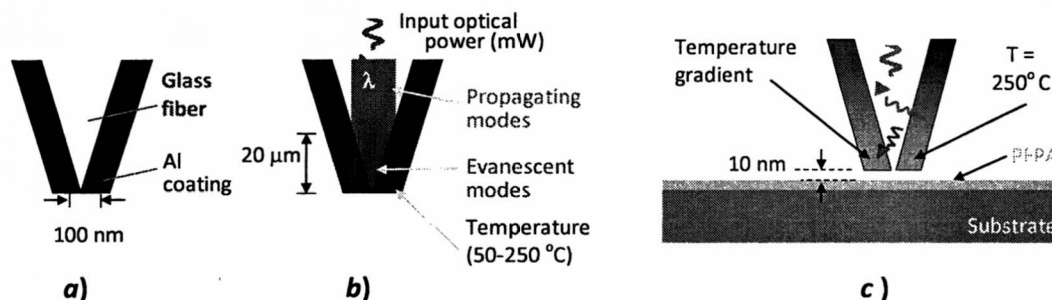


Fig. 20 Heat effects on a NSOM probe. *a)* Schematic of a probe of tapered geometry. *b)* Closer to the apex, the probe's inner cross-section area gradually becomes too narrow for sustaining the light's propagating modes, but evanescent modes are allowed, whose energy is deposited within the skin-depth of the inner metal interface. *c)* The apex of the probe is placed in the proximity of the sample. The application of pulses of light can be used to control the probe temperature and, hence, locally thermally excite the PFPA monolayer.

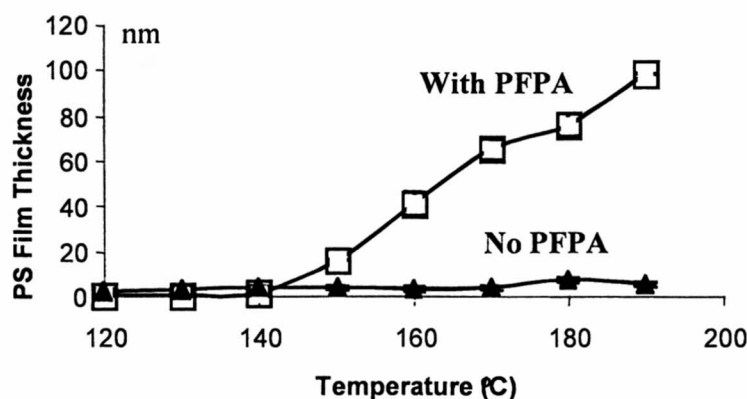


Fig. 21 Immobilized film thickness as a function of temperature. Reprinted with permission from M. Yan and J. Ren, *Chem. Mater.* **16**, 1627 (2004)). Copyright 2004 American Chemical Society.

Studies of PFPA activation at macroscopic scales by thermal means has been reported.⁹⁴ For a 5 minute heating time, the thermal activation of the surface azido groups occurred only at temperatures above 140 °C (see corresponding open squares in Figure 21). Below this temperature, no film immobilization occurred. Accordingly, when using a metal-coated near-

field probe, setting the *peak temperature at its apex above*, the thermal activation temperature of the azido groups, film immobilization should take place only in the region around the area at peak temperature, thus minimizing line broadening. It will be interesting to determine the temperature at which optimal spatial resolution can be achieved without compromising the integrity of the immobilized films. In short this approach would constitute a near-field thermal excitation, exploiting the gradient temperature established on the metallic probe.

IV.3.C The ultraviolet near-field approach

Radiation from a Ti:Sapphire laser system (equipped with non-linear optics for doubling and tripling the operating fundamental frequency at $\lambda=790$ nm) can be coupled to sub-wavelength apertures, which when scanned in the proximity of the sample trigger the PFPA reaction at pre-determined localized sites. This is the realm of near-field uv-lithography.¹³⁰ Different types of sub-wavelength sized apertures can be used for this near-field uv application, including metallic apertures fabricated at the apex of tapered fiber glass probes (as the ones shown in Figs. 18 and 20 above)¹³¹; apertures fabricated (using a focused ion beam) at the pyramid-apex of a semiconductor atomic force microscope probe⁵⁹; or apertures fabricated out of GaAs, which latter displaying extraordinary transmission efficiency (due to the intrinsic nature of the uv radiation).¹³²

In near-field uv-optical *lithography*, light that exits the sub-wavelength aperture (laterally scanned in close proximity, 10 nm, to the sample surface) induces chemical reactions very effectively in the surrounding materials *via* polymerization, crosslinking, or decomposition. Distinct irradiated regions, scanned over with the near-field optical probe, can subsequently be etched out by capitalizing on the differential solubility of the exposed and unexposed regions, thus creating nanometer-sized structures. The experimental protocol will start by treating silicon wafers with PFPA-silane. The sample will then be scanned with a near-field probe. PFPA-silane has a maximal absorption at 256 nm that tails to 350 nm. The un-attached polymer will be removed by soaking the sample in a solvent. Exposure time will be optimized to investigate the attainable resolution of a given near-field probe. The spatial resolution and the topography of patterned films will be characterized using AFM as well as near-field optical microscopy. A number of parameters will be investigated with respect to spatial resolution, immobilization yield and film integrity. These include the aperture size of the probe, tip-sample distance, intensity of the UV laser, and scan rate.

IV.4 Implementation

We will explore PFPA chemical activation via non-linear thermal response of the surrounding substrate using femtosecond laser pulses. The non-linear response (absorption of multiple photons) is associated to the high power intensities resulting from the incidence of pulses of short duration, which results in an intrinsic spatial confinement effects because of the high temperature gradients. While confined effects are reported in the context of material ablation applications (features of 20 nm have been fabricated using radiation of 527-nm wavelength, average pulse energy of 7.5nJ).¹¹⁹ We certainly will not seek material ablation, but instead study the nonlinear absorption effects for chemical bond modification.

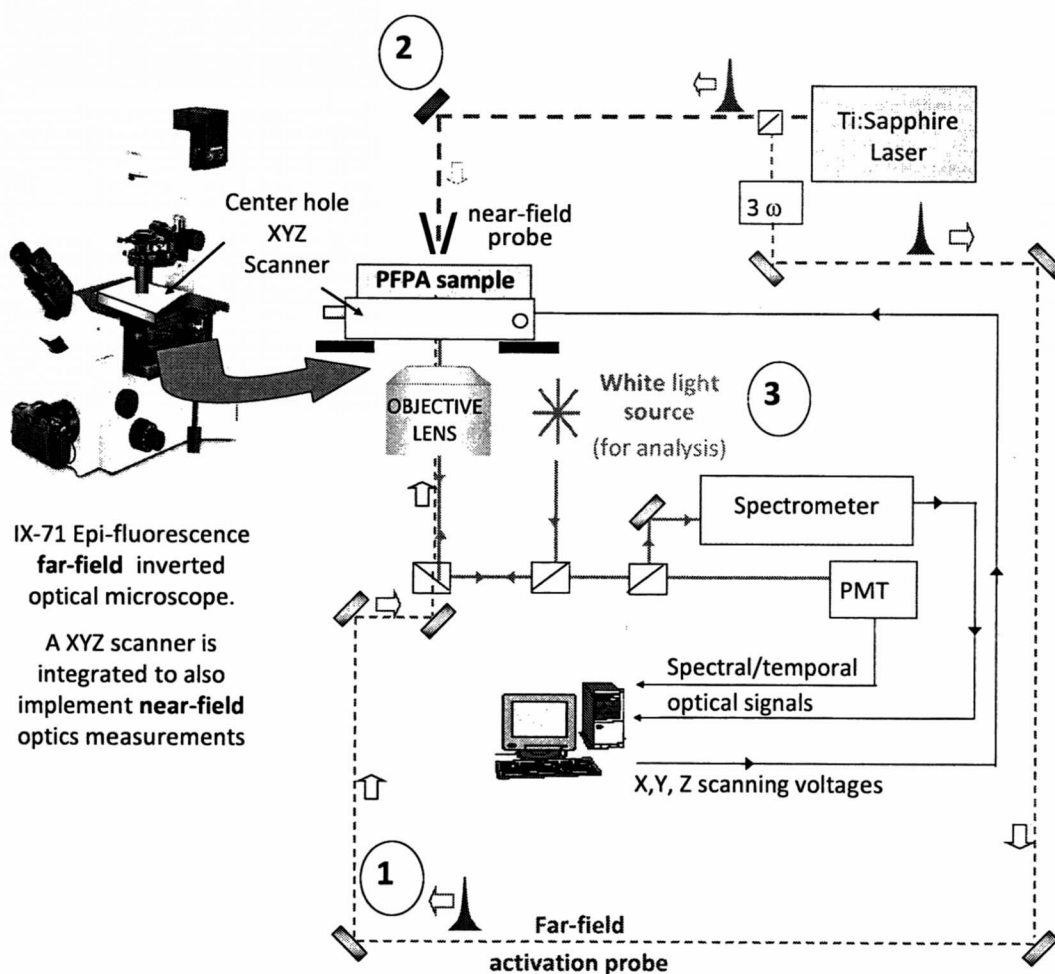


Fig. 21 Experimental setup for activating (using sub-25fs pulses) and characterizing (using white light source) the covalent attachment of PFPA molecules to a working substrate (silicon or glass). Notice the output laser beam is split in order to implement far-field (1) and near-field (2) illuminations. Tripling the laser output frequency is optional. The XYZ scanner is for precise localization for activating a single stimuli-responsive molecule.

V. REFERENCES

- ¹ L. Anson, "Membrane protein biophysics," *Nature* **459**, 343 (2009).
- ² B. Bhushan, "Biomimetics: lessons from nature-an overview," *Phil. Trans. R. Soc. A* **367**, 1445 (2009).
- ³ C. J. Jones, and S. Aizawa, "The bacterial flagellum and flagellar motor: structure, assembly, and functions". *Adv. Microb. Physiol.* **32**, 109 (1991).
- ⁴ C. Cofield, "Cell is mechanical device," *The American Physical Society, APS news, Series II*, **19**, 4 (June 2010).
- ⁵ C. Ainsworth, "Stretching the imagination," *Nature* **456**, 696 (2008).

- ⁶ C. Wu, Y. Li, J. H. Haga, R. Kaunas, J. Chiu, F. Su, S. Usami, and S. Chien, "Directional shear flow and Rho activation prevent the endothelial cell apoptosis induced by micro patterned anisotropic geometry," *PNAS* **104**, 1254 (2007).
- ⁷ C. S. Chen, M. Mrksich, S. Huang, G. M. Whitesides, and D. E. Ingber, "Geometric control of cell life and death," *Science* **276**, 1425 (1997).
- ⁸ P. M. Gilbert, K. L. Havenstrite, K. E. G. Magnusson, A. Sacco, N. A. Leonardi, P. Kraft, N. K. Nguyen, S. Thrun, M. P. Lutolf, and H. M. Blau, "Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture," *Science* **329**, 1078 (2010).
- ⁹ O. Z. Fisher, A. Khademhosseini, R. Langer, and N. A. Peppas, "Bioinspired Materials for Controlling Stem Cell Fate," *Accounts of Chemical Research* **43**, 419 (2010).
- ¹⁰ M. A. Greenfield, J. R. Hoffman, M. Olvera de la Cruz, and S. I. Stupp, "Tunable mechanics of peptide nanofiber gels," *Langmuir* **26**, 3641 (2010).
- ¹¹ M. M. Stevens, J. H. George, "Exploring and engineering the cell surface interface," *Science* **310**, 1135 (2005).
- ¹² C. Ainsworth, "Stretching the imagination," *Nature* **456**, 696 (2008).
- ¹³ A. Khademhosseini, R. Langer, "Microengineered hydrogels for tissue engineering," *Biomaterials* **28**, 5087 (2007).
- ¹⁴ A hydrogel refers to a flexible (typically) hydrophilic cross-linked polymer network and a fluid filling the interstitial spaces of the network
- ¹⁵ C. Rivest, D. W. G. Morrison, B. Ni, J. Rubin, V. Yadav, A. Mahdavi, J. M. Karp, and A. Khademhosseini, "Microscale hydrogels for medicine and biology: synthesis, characteristics, and applications," *Journal of Mechanics of materials and structures* **2**, 1103 (2007)
- ¹⁶ J. Yeh, Y. Ling, J. M. Karp, G. Eng, J. Blumling Iii, R. Langer, and A. Khademhosseini, "Micromolding of shape-controlled, harvestable cell-laden hydrogels," *Biomaterials* **27**, 5391 (2006).
- ¹⁷ Sefton and McGuigan 2006] M. Sefton and A. McGuigan, "Vascularized organoid engineered by modular assembly enables blood perfusion", *PNAS*. **103**, 11461 (2006).
- ¹⁸ S. Maeda, Y. Hara, T. Sakai, R. Yoshida, and S. Hashimoto, "Self-walking gel," *Adv. Mater.* **19**, 3480 (2007).
- ¹⁹ P. Calvert, "Hydrogels for soft machines," *Adv Mater.* **21**, 743 (2009).
- ²⁰ T. K. Tam, M. Ornatska, M. Pita, S. Minko, and E. Katz, "Polymer brush-modified electrode with switchable and tunable redox activity for bioelectronic applications," *J. Phys. Chem. C* **112**, 8438 (2008).
- ²¹ D. Wang, I. Lagzi, P. J. Wesson, and B. A. Grzybowski, "Rewritable and pH-sensitive micropatterns based on nanoparticle 'Inks'," *Small* **6**, 2114 (2010).
- ²² J. Ruhe, M. Ballauff, M. Biesalski, P. Dziezok, F. Grohn, D. Johannsmann, N. Houbenov, N. Hugenberg, R. Konradi, S. Minko, M. Motornov, R. R. Netz, M. Schmidt, C. Seidel, M. Stamm, T. Stephan, D. Usov and H. Zhang, "Polyelectrolyte brushes" in: Polyelectrolytes with defined molecular architecture," M. Schmidt Ed. *Adv. Polym. Sci.* **165**, 79 (2004).
- ²³ C. Liu, H. Qin, P. T. Mather, "Review of progress in shape-memory polymers," *J. Mater. Chem.* **17**, 1543 (2007).
- ²⁴ R. Yoshida, K. Uchida, Y. Kaneko, K. Sakai, A. Kikuchi, Y. Sakurai and T. Okano, "Comb-type grafted hydrogels with rapid deswelling response to temperature changes," *Nature* **374**, 240 (1995).

- 25 T. Suzuki, S. Shinkai, K. Sada, "Supramolecular crosslinked linear poly(trimethylene iminium trifluorosulfonimide) polymer gels sensitive to light and thermal stimuli," *Adv. Mater.* **18**, 1043 (2006).
- 26 B. A. Evans, A. R. Shields, R. Lloyd Carroll, S. Washburn, M. R. Falvo, and R. Superfine, "Magnetically actuated nanorod arrays as biomimetic cilia," *Nano Lett.* **7**, 1428 (2007).
- 27 B. Bhushan, "Biomimetics: lessons from nature-an overview," *Phil. Trans. R. Soc. A* **367**, 1445 (2009).
- 28 I. Tokarev, M. Motornov, and S. Minko; "Molecular-engineered stimuli-responsive thin polymer film: a platform for the development of integrated multifunctional intelligent materials," *J. Mater. Chem.* **19**, 6932 (2009).
- 29 C. Maedler, S. Chada, X. Cui, M. Taylor, M. Yan, A. La Rosa, "Creation of nanopatterns by local protonation of P4VP via dip pen nanolithography," *J. Appl. Phys.* **104**, 014311 (2008).
- 30 B. Harnish, J. T. Robinson, Z. Pei, O. Ramstrom, and M. Yan, *Chem. Mater.* **17**, 4092 (2005).
- 31 J. H. Burroughes, D. D. C. Bradley, A. R. Brown, R. N. Marks, K. Mackay, R. H. Friend, P. L. Burns, and A. B. Holmes, "Light-emitting diodes based on conjugated polymers," *Nature* **347**, 539 (1990).
- 32 T. Sekitani, U. Zschieschang, H. Klauk and T. Someya, "Flexible organic transistors and circuits with extreme bending stability," *Nature Materials* **9**, 1015 (2010).
- 33 L. Grill, M. Dyer, L. Lafferentz, M. Persson, M. V. Peters and S. Hecht, "Nano-architectures by covalent assembly of molecular building blocks," *Nature Nanotechnology* **2**, 687 - 691 (2007).
- 34 O. Ikkala and G. Brinke, "Functional materials based on self-assembly of polymeric supramolecules," *Science* **295**, 2407 (2002).
- 35 R. F. Pease, and S. Y. Chou, "Lithography and Other Patterning Techniques for Future Electronics," *Proceedings of the IEEE* **96**, 248 (2008).
- 36 http://www.gitam.edu/eresource/nano/nanotechnology/role_of_bottomup_and_topdown_a.htm
- 37 H. Jalali and B. D. Gates, "Monitoring and Mapping Imperfections in Silane-Based Self-Assembled Monolayers by Chemical Amplification," *Langmuir* **25**, 9078 (2009).
- 38 <http://www.mtl.kyoto-u.ac.jp/english/laboratory/nanoscopic/nanoscopic.htm>
- 39 S. M. Sze, "Semiconductor Devices, Physics and Technology," John Wiley (1985).
- 40 Marc J. Madou, "Fundamentals of Microfabrication", 2nd Edition, CRC Press (2002).
- 41 M. Elwenspoek and H. V. Jansen; "Silicon Micromachining"; Cambridge University Press; (1998).
- 42 J. Bardeen, W. H. Brattain, *Phys. Rev.* **74**, 230 (1948).
- 43 E. Reichmanis and O. Nalamasu, "Testing the Limits for Resists", *Science* **297**, 349 (2002).
- 44 M. D. Levenson, N. S. Viswanathan, and R. A. Simpson, "Improving resolution in photolithography with a phase-shifting mask," *IEEE Trans. Electron Devices* **ED-29**, 1828 (1982).
- 45 E. Reichmanis, O. Nalamasu, and F. M. Houlihan, "Organic Materials Challenges for 193 nm Imaging," *Acc. Chem. Res.* **32**, 659 (1999).
- 46 S. Owa, H. Nagasaka, K. Nakano, and Y. Ohmura, "Current status and future prospect of immersion lithography," *Proc. SPIE* **6154**, 615408-1 (2006).
- 47 R. Menon and H. I. Smith, "Absorbance modulation optical lithography (AMOL)," *J. Opt. Soc. Amer.* **A23**, 2290 (2006).
- 48 T. W. O'Keefe and R. M. Handy, "An electron imaging system for the fabrication of integrated circuits," *Solid State Electron.* **12**, 841(1969).

- ⁴⁹ T. H. P. Chang, M. G. R. Thomson, E. Kratschmer, H. S. Kim, M. L. Yu, K. Y. Lee, S. A. Rishton, B. W. Hussey, and S. Zolgharnain, "Electron-beam micro columns for lithography and related applications," *J. Vac. Sci. Technol.* **B14**, 3774 (1996).
- ⁵⁰ M. Sugihara, T. Takata, K. Nakamura, R. Inanami, H. Hayashi, K. Kishimoto, T. Hasebe, Y. Kawano, Y. Matsunaga, K. Murakami, and K. Okumura, "Technology mapping technique for throughput enhancement of character projection equipment," in *Proc. SPIE* **6151**, X. Michael and J. Lercel, Eds., Emerging Lithographic Technologies, paper 61510Z-1 (2006).
- ⁵¹ J. A. Liddle, G. M. Gallatin, and L. E. Ocola, "Resist Requirements and Limitations for Nanoscale Electron-Beam Patterning," *MRS Proceedings* **739**: H1.5 (2002).
- ⁵² L. F. Thompson, C. G. Willson, and M. J. Bowden, "Introduction to Microlithography". Oxford Univ. Press (1983).
- ⁵³ R. Kaesmaier, H. Loschner, G. Stengl, J. C. Wolfe, and P. Ruchhoeft, "Ion projection lithography: International development program," *J. Vac. Sci. Technol.* **B17**, 3091 (1999).
- ⁵⁴ J. P. H. Benschop, A. J. J. van Dijsseldonk, W. M. Kaiser, and D. C. Ockwell, "BEUCLIDES: European EUVL Program," *J. Vac. Sci. Technol.* **B17**, 2978 (1999).
- ⁵⁵ M. Kaholek, W. Kyung Lee, B. LaMattina, K. C. Caster, and S. Zauscher, "Fabrication of stimulus-responsive nanopatterned polymer brushes by scanning-probe lithography," *Nano Lett.* **4**, 373 (2004).
- ⁵⁶ X. Liu, S. Guo, and C. A. Mirkin, "Surface and site-specific ring-opening metathesis polymerization initiated by dip-pen nanolithography," *Angew. Chem. Int. Ed.* **42**, 4785 (2003).
- ⁵⁷ Y. Okawa, M. Aono, "Nanoscale control of chain polymerization," *Nature* **409**, 683 (2001).
- ⁵⁸ S. P. Sullivan, A. Schnieders, S. K. Mbugua, and T. P. Beebe Jr, "Controlled polymerization of substituted diacetylene self-organized monolayers confined in molecule corrals," *Langmuir* **21**, 1322 (2005).
- ⁵⁹ Y. Okawa, D. Takajo, S. Tsukamoto, T. Hasegawa and M. Aono, "Atomic force microscopy and theoretical investigation of the lifted-up conformation of polydiacetylene on a graphite substrate," *Soft Matter* **4**, 1041 (2008).
- ⁶⁰ S. Y. Chou, P. R. Krauss, and P. J. Renstrom, "Imprint of sub 25 nm vias and trenches in polymers," *Science*. **272**, 85 (1996).
- ⁶¹ A. Kumar and G. M. Whitesides, "Features of gold having micrometer to centimeter dimensions can be formed through a combination of stamping with an elastomeric stamp and an alkanethio link followed by chemical etching," *Appl. Phys. Lett.* **63**, 2002 (1993).
- ⁶² Banqiu Wu, Ajay Kumar, and Sharma Pamarthy, "High aspect ratio silicon etch: A review," *J. Appl. Phys.* **108**, 051101 (2010).
- ⁶³ K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, T. Hashimoto, T. Kobayashi, T. Kure. and K. Seki, "Single-electron memory for giga-to-tera bit storage," *Proceedings of the IEEE* **87**: 633– 651 (1999).
- ⁶⁴ Zahid Ali Khan Durrani, "Single-Electron Devices And Circuits In Silicon," World Scientific (2009).
- ⁶⁵ K. K. Likharev K.K., "Correlated discrete transfer of single electrons in ultrasmall tunnel junctions." *IBM J Res Develop* **32**, 144 (1988).
- ⁶⁶ Takahashi Y., Nagase M., Namatsu H., Kurihara K., Iwadata K., Nakajima Y., Horiguchi S., Murase K. and Tabe M. (1995). Fabrication technique for Si single-electron transistor operating at room temperature. *Electronics Letters* **31**, 136 (1995).

- ⁶⁷ Takahiro Shinada, Shintaro Okamoto, Takahiro Kobayashi and Iwao Ohdomari, "Enhancing semiconductor device performance using ordered dopant arrays," *Nature* **437**, 1128 (2005).
- ⁶⁸ W. Wu, J. Gu, H. Ge, C. Keimel, and S. Y. Chou, "Room-temperature Si single-electron memory fabricated by nano imprint lithography," *Appl. Phys. Lett.* **83**, 2268 (2003).
- ⁶⁹ B. J. Hinds, T. Yamanaka, and S. Oda, "Emission lifetime of polarizable charge stored in nano-crystalline Si based single-electron memory," *J. Appl. Phys.* **90**, 6402 (2001).
- ⁷⁰ H. Grabert, M. H. Devoret, and North Atlantic Treaty Organization. Scientific Affairs Division, "*Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures*," (Plenum, New York, 1992).
- ⁷¹ T. Tsusumi, H. Hirayam, M. Yacha, and T. Taniyama, "Nanoscale Physics for Materials Science," CRC Press 2010. See Section 2.5.
- ⁷² L. Guo, E. Leobandung, S.Y. Chou, "A Silicon Single-Electron Transistor Memory Operating at Room Temperature," *Science* **275**, 649 (1997).
- ⁷³ H. I. Liu, D. K. Biegelsen, F. A. Ponce, N. M. Johnson, and R. F. W. Pease, "Self-limiting oxidation for fabricating sub-5 nm silicon nanowires," *Appl. Phys. Lett.* **64**, 1383 (1994).
- ⁷⁴ M. Muthukumar, C. K. Ober, E. L. Thomas, "Competing interactions and levels of ordering in self-organizing polymeric materials," *Science* **277**, 1225 (1997).
- ⁷⁵ Frank S. Bates and Glenn H. Fredrickson, "Block Copolymer Thermodynamics: Theory and Experiment," *Annu. Rev. Phys. Chem.* **41**, 525 (1990).
- ⁷⁶ J. Ruokolainen, M. Torkkeli, R. Serimaa, E. Komanshek, G. ten Brinke, and Olli Ikkala, "Order-Disorder Transition in Comblike Block Copolymers Obtained by Hydrogen Bonding between Homopolymers and End-Functionalized Oligomers: Poly(4-vinylpyridine)-Pentadecylphenol," *Macromolecules* **30**, 2002 (1997).
- ⁷⁷ T. Thurn-Albrecht, J. De Rouchey, T. P. Russell, and H. M. Jaeger, *Macromolecules* **33**, 3250 (2000).
- ⁷⁸ C. Park, J. Y. Cheng, M. J. Fasolka, A. M. Mayes, C. A. Ross, E. L. Thomas, and Claudio De Rosa, "Double textured cylindrical block copolymer domains via directional solidification on a topographically patterned substrate," *Appl. Phys. Lett.* **79**, 848 (2001).
- ⁷⁹ E. Huang, P. Mansky, T. P. Russell, C. Harrison, P. M. Chaikin, R. A. Register, C. J. Hawker, and J. Mays, *Macromolecules* **33**, 80 (2000).
- ⁸⁰ M. Knaapila, "Polarized luminescence from self-assembled, aligned, and cleaved supramolecules of highly ordered rodlike polymers," *Appl. Phys. Lett.*, **81**, 1489 (2002).
- ⁸¹ E. Mena-Osteritz *et al.*, *Angew. Chem. Int. Ed. Eng.* **39**, 2680 (2000).
- ⁸² J. Ruokolainen, R. Makinen, M. Torkkeli, T. Makela, R. Serimaa, G. ten Brinke, and O. Ikkala, "Switching Supramolecular Polymeric Materials with Multiple Length Scales," *Science* **280**, 557 (1998).
- ⁸³ L. Wang, J. Yoshida, N. Ogata, S. Sasaki, T. Kajiyama, "Self-Assembled Supramolecular Films Derived from Marine Deoxyribonucleic Acid (DNA)-Cationic Surfactant Complexes: Large-Scale Preparation and Optical and Thermal Properties," *Chem. Mater.* **13**, 1273 (2001).
- ⁸⁴ I. Tokarev and S. Minko, "Stimuli-responsive hydrogel thin films," *Soft Matter* **5**, 511 (2009).
- ⁸⁵ K. Salaita, Y. Wang and C. A. Mirkin, "Applications of dip-pen nanolithography," *Nature Nanotechnology* **2**, 145 (2007).
- ⁸⁶ R. D. Piner, J. Zhu, F. Xu, S. Hong, and C. A. Mirkin, "Dip-Pen Nanolithography," *Science* **283**, 661 (1999).

- ⁸⁷ Xiaohua Wang, Xin Wang, R. Fernandez, L. Ocola, M. Yan, and A. La Rosa; "Electric field-assisted dip-pen nanolithography on poly(4-vinyl pyridine) films," *ACS Appl. Mater. Interface* **2**, 2904 (2010).
- ⁸⁸ C. Maedler, H. Graaf, S. Chada, M. Yan, and A. La Rosa, "Nano-structure formation driven by local protonation of polymer thin films", *Proc. SPIE*, **7364**, 736409-1 (2009).
- ⁸⁹ C. Maedler, "Applying different modes of atomic force microscopy for the manipulation and characterization of spatially localized structures and charges," Diploma Thesis for the academic degree of Diplom Physiker; Faculty of Natural Sciences Institute of Physics, Chemnitz University of Technology (2009).
- ⁹⁰ A functional group is a specific group of atoms within a molecule that is responsible for characteristic chemical reactions of that molecule. The hydroxyl group is a functional group consisting of a hydrogen atom covalently bonded to an oxygen atom.
- ⁹¹ L. T. Zhuravlev, "Concentration of Hydroxyl Groups on the Surface of Amorphous Silicas," *Langmuir* **3**, 316 (1987).
- ⁹² M. Bartlett, M. Yan., "Photochemical Immobilization of Polymer Thin Films on Solid Substrates," *Polym. Mater. Sci. Eng.* **83**, 451 (2000).
- ⁹³ M. Bartlett, M. Yan, "Fabrication of Polymer Thin Films and Arrays with Spatial and Topographical Controls," *Adv. Mater.* **13**, 1449 (2001)
- ⁹⁴ M. Yan and J. Ren, "Covalent Immobilization of Ultrathin Polymer Films by Thermal Activation of Perfluorophenyl Azide," *Chem. Mater.* **16**, 1627 (2004).
- ⁹⁵ X. Wang, O. Ramstrom, M. Yan, "Glyconanomaterials: Synthesis, Characterization, and Ligand Presentation." *Adv. Mater.* **22**, 1946 (2010).
- ⁹⁶ L. -H. Liu, M. M. Lerner, M. Yan, "Derivatization of Pristine Graphene with Well-defined Chemical Functionalities," *Nano Lett.* **10**, 3754 (2010).
- ⁹⁷ L. Liu, M. Yan, "Perfluorophenyl Azides: New Applications in Surface Functionalization and Nanomaterial Synthesis," *Acc. Chem. Res.* **43**, 1434 (2010).
- ⁹⁸ J. F. W. Keana and S. X. Cai, "New Reagents for Photoaffinity Labeling: Synthesis and Photolysis of Functionalized Perfluorophenyl Azides," *J. Org. Chem.* **55**, 3640 (1990).
- ⁹⁹ L. Liu, M. H. Engelhard, and M. Yan, "Surface and Interface Control on Photochemically Initiated Immobilization," *J. Am. Chem. Soc.* **128**, 14067 (2006).
- ¹⁰⁰ L. Liu, M. Yan, "A General Method for the Covalent Immobilization of Single Polymers," *Angew. Chem. Int. Ed.* **45**, 6207 (2006).
- ¹⁰¹ J. Y. Kim, Y. H. Park, J. S. Kim, K. T. Lim, M. Yan, and Y. T. Jeong, "Photochemical immobilization of polymer films on Si wafer via monolayers of perfluorophenyl azide derivatives," *J. Ind. Eng. Chem.* **13**, 781 (2007).
- ¹⁰² Hiroyuki Sugimura, "Self-assembled monolayer in silicon" <http://www.mtl.kyoto-u.ac.jp/groups/sugimura-g/PDF/SAM-on-Si.pdf> A molecule consisting of one Si atom connected with four functional groups, SiX₄, is called as "silane". If one, or more, of these functional groups are substituted with organic functional groups (X=CH₃ for example) it is called *organosilane*. Organosilane molecules react with hydroxyl groups on an oxide surface.
- ¹⁰³ Y. Kim, J. Han, H. Sano, K. Lee, K. Noda, T. Ichii, K. Murase, K. Matsushige, Hiroyuki Sugimura, "Organosilane self-assembled multilayer formation based on activation of methyl-terminated surface with reactive oxygen species generated by vacuum ultra-violet excitation of atmospheric oxygen molecules," *Applied Surface Science* **256**, 1507 (2009).

- 104 R. Tian, O. Seitz, M. Li, Wenchuang (Walter) Hu, Y. J. Chabal, and J. Gao, "Infrared Characterization of Interfacial Si-O Bond Formation on Silanized Flat SiO₂/Si Surfaces," *Langmuir* **26**, 4563 (2010).
- 105 A. Ulman, "Formation and Structure of Self-Assembled Monolayers," *Chem. Rev.* **96**, 1533 (1996).
- 106 M. Yan and J. Ren, "Covalent Immobilization of Ultrathin Polymer Films by Thermal Activation of Perfluorophenyl Azide," *Chem. Mater.* **16**, 1627 (2004).
- 107 M. Yan and B. Harnish, "A simple method for the attachment of polymer films on solid substrates," *Adv. Mater* **15**, 244 (2003).
- 108 Bruce H. Mahan, University Chemistry," 3rd Edition, Adison-Wesley (1975). Section 6.5
- 109 Since all phosphate salts are used in hydrated condition, the molecular weight (MW) should include the corresponding portion of water. For NaH₂PO₄ we should include one molecule of water, hence the MW is 137.99. On the other hand, for the Na₂HPO₄ we should consider 7 molecules of water (heptahydrate,) which gives a MW of 268.07. Hence, if 13.8 g and 0.036 g of NaH₂PO₄ and Na₂HPO₄ are used respectively, then we can quote the concentration of NaH₂PO₄ (the buffer strength) to be practically equal to 0.1M.
- 110 Y. Osada and J. Gong, "Soft and wet materials: Polymer gels," *Adv. Mater.* **10**, 827 (1998).
- 111 S. Rozhok, R. Piner, and C. A. Mirkin, "Dip-Pen Nanolithography: What Controls Ink Transport?," *J. Phys. Chem. B* **107**, 751 (2003).
- 112 M. Schenk, M. Futing, R. Reichelt, "Direct visualization of the dynamic behavior of water meniscus by scanning electron microscopy," *J. Appl. Phys.* **84**, 4880 (1998).
- 113 M. Nonnenmacher, M. P. O'Boyle, and H. K. Wickramasinghe, "Kelvin probe force microscopy," *Appl. Phys. Lett.* **58**, 2921 (1991).
- 114 V. Westphal, M. A. Lauterbach, A. Di Nicola, and S. W. Hell, "Dynamic far-field fluorescence nanoscopy," *New Journal of Physics* **9**, 435 (2007).
- 115 B. G. Saar, C. W. Freudiger, J. Reichman, C. M. Stanley, G. R. Holtom, X. S. Xie, "Video-Rate Molecular Imaging in Vivo with Stimulated Raman Scattering," *Science* **330**, 1368 (2010).
- 116 Fabrication of probes by chemical etching R. Stockle, C. Fokas, V. Deckert, and R. Zenobi, " High-quality near-field optical probes by tube etching," *Appl. Phys. Lett.* **75**, 160 (1999).
- 117 Justin McCollum, "Probe Fabrication for Near-field Optical Microscopy Applications" Physics Master Thesis, Portland State University (2005).
- 118 R. Srinivasan, E. Sutcliffe, and B. Braren, "Ablation and etching of polymethylmethacrylate by very short (160 fs) ultraviolet (308 nm) laser pulses," *Appl. Phys. Lett.* **51**, 1285 (1987).
- 119 A. P. Joglekar, H. Liu, G. J. Spooner, E. Meyhofer, G. Mourou, and A. J. Hunt, "A study of the deterministic character of optical damage by femtosecond laser pulses and applications to nanomachining," *Appl. Phys. B* **77**, 25 (2003).
- 120 Nicolaas Bloembergen, "Laser-Induced Electric Breakdown in Solid," *IEEE Journal of Quantum Electronics*, QE-10, 375 (1974).
- 121 B. C. Stuart, M.D. Feit, A. M. Rubenchik, B. W. Shore, and M. D. Perry, "Laser-Induced Damage in Dielectrics with Nanosecond to Subpicosecond," *Phys. Rev. Lett.* **74**, 2248 (1995).
- 122 D. Du, X. Liu, G. Korn, J. Squier, and G. Mourou, "Laser-induced breakdown by impact ionization in SiO₂ with pulse widths from 7 ns to 150 fs," *Appl. Phys. Lett.* **64**, 3071 (1994).
- 123 R. Gattass and E. Mazur, "Femtosecond laser micromachining in transparent materials," *Nature Photonics* **2**, 219 (2008).

- ¹²⁴ Z. K. Wang, H. Y. Zheng, C. P. Lim, and Y. C. Lam, "Polymer surface wettability modification using femtosecond laser Irradiation," *SIMTech Technical Reports* **11**, 23 (Jan-March 2010).
- ¹²⁵ B. N. Chichkov, C. Momma, S. Nolte, F. von Alvensleben, A. Tunnermann, "Femtosecond, picosecond and nanosecond laser ablation of solids," *Appl. Phys. A* **63**, 109 (1996)
- ¹²⁶ O. Fenwick, L. Bozec, D. Credgington, A. Hammiche, G. Mattia Lazzerini, Y. Silberberg, and F. Cacialli, "Thermochemical nanopatterning of organic semiconductors," *Nature Nanotechnology* **4**, 664 (2009).
- ¹²⁷ A. H. La Rosa, B. I. Yakobson and H. D. Hallen, "Origins and effects of thermal processes on near-field optical probes", *Appl. Phys. Lett.* **67**, 2597 (1995).
- ¹²⁸ M. Stähelin, M. A. Bopp, G. Tarrach, A. J. Meixner, and I. Zschokke-Grahnacher, "Temperature profile of fiber tips used in scanning near-field optical microscopy," *Appl. Phys. Lett.* **68**, 2603 (1996).
- ¹²⁹ M. Yan, J. Ren, "Covalent Immobilization of Ultrathin Polymer Films by Thermal Activation of Perfluorophenyl Azide," *Chem. Mater.* **16**, 1627 (2004).
- ¹³⁰ S. A. Alang Ahmad, L. S. Wong, Ehtsham ul-Haq, J. K. Hobbs, G. J. Leggett, and J. Micklefield, "Protein Micro- and Nanopatterning Using Aminosilanes with Protein-Resistant Photolabile Protecting Groups," *J. Am. Chem. Soc.* **133**, 2749 (2011).
- ¹³¹ S. Sun and G. J. Leggett, "Matching the Resolution of Electron Beam Lithography by Scanning Near-Field Photolithography," *Nano Lett.* **4**, 1381 (2004).
- ¹³² M. A. Vincenti, D. deCeglia, M. Buncick, N. Akozbek, M. J. Bloemer, and M. Scalora, "Extraordinary transmission in the ultraviolet range from subwavelength slits on semiconductors," *J. Appl. Phys.* **107**, 053101 (2010).

Two dimensional BF gravity: A Hamilton-Jacobi analysis

M. C. Bertin¹, B. M. Pimentel², C. E. Valcárcel³

¹*CMCC, Universidade Federal do ABC,*

Rua Santa Adélia, 166, Santo André, SP Brazil.

^{2,3}*Instituto de Física Teórica, UNESP - São Paulo State University,
P. O. Box 70532-2, 01156-970, São Paulo, SP, Brazil.*

Abstract

We analyse the constraint structure of the BF model for two dimensional gravity via the Hamilton-Jacobi formalism. This analysis consists in finding the complete set of involutive Hamiltonians that assure the integrability of the system. We then calculate the characteristic equations of the system, also establishing the equivalence between these equations and the field equations.

Dedicated to Professor H. G. Valqui on the occasion of his 80th birthday.

Keywords: Hamilton-Jacobi formalism, BF model.

1 Introduction

Until now, there is no satisfactory quantum theory for the four dimensional gravity. In order to understand some properties of gravity at quantum level, it has been found at lower dimensional models of gravity interesting topics of research (see [1] and references therein). There are some trivialities on those models, for example, the three dimensional pure gravity has no degrees of freedom. For the two dimensional case, it is well known that the Einstein's tensor is identically zero, then the Einstein's equation will be always satisfied for pure gravity.

Because of this trivial behavior of two dimensional gravity it is customary to make some reformulations in the Einstein-Hilbert action. One of the most used models is due to Jackiw and Teitelboim [2], in which the action is given by

$$I_{JT} = \int d^2x \sqrt{g} \psi (R - k),$$

where ψ is a dilaton field used as a lagrangian multiplier, and k is the cosmological constant. The Euler-Lagrange (EL) equations for the Jackiw-Teitelboim (JT) model are given by

$$R - k = 0, \quad \nabla_\mu \nabla_\nu \psi + \frac{1}{2} k g_{\mu\nu} \psi = 0,$$

the first is the constant curvature equation, and the second one is the equation of motion for the dilaton field which is determined without making further restrictions on the metric. For an extensive review of two dimensional dilaton models of gravity, we refer to [3].

The JT model can be reformulated as an $SO(2, 1)$ gauge theory that can be reproduced through a BF action [4]. The Hamiltonian analysis of the BF model has been studied in [5]. The aim of this article is to analyse the constraint structure of the BF model for two dimensional gravity by means of the Hamilton-Jacobi (HJ) formalism [6].

¹mcbertin@ift.unesp.br

²pimentel@ift.unesp.br

³valcarcel@ift.unesp.br

The HJ formalism is an alternative method to analyse constrained systems. It was born as a generalization of Carathéodory's work [7] on variational principles and first-order partial differential equations. In this formalism we obtain a set of Hamilton-Jacobi Partial Differential Equations (HJPDE) also known as Hamiltonian densities, and the integrability of this system is achieved if the Frobenius' Integrability Condition (IC) is satisfied. There are several applications and improvements of this formulation [8]. Concerning two dimensional gravity, The HJ formalism has been used to study the Polyakov's model in the front-form dynamics [9], the usual Einstein-Hilbert action [10], and also the linearized gravity [11].

In the next section we briefly discuss the HJ formalism. In section 3 we show an introduction to the BF model. In section 4 we make the proper constraint analysis which consists in finding the complete set of involutive Hamiltonians of the theory. In section 5 we compute the characteristic equations (CE) and analyse the dynamical evolution along the independent parameters of the theory. Finally we discuss the results.

2 The Hamilton-Jacobi Formalism

Let us consider a system described by a Lagrangian density $L = L(x^i, \dot{x}^i, t)$, $i = 1, 2, \dots, N$, with a singular Hessian matrix of rank P . In this case we separate the variables $x^i = (x^a, x^z)$, where $a = 1, 2, \dots, P$ and $z = 1, 2, \dots, R$ with $P + R = N$. The variables x^a are related to the invertible part of the Hessian, while x^z are related to its zero-modes. Following the Carathéodory's variational approach, the necessary and sufficient condition for extremizing the action is given by $p_0 + p_a \dot{x}^a + p_z \dot{x}^z - L = 0$, where the canonical momenta are defined by $p_a = \partial S / \partial x^a$, $p_z = \partial S / \partial x^z$ and $p_0 = \partial S / \partial t$.

The singularity of the Hessian matrix assures that there are R canonical constraints $H'_z \equiv p_z - \partial L / \partial \dot{x}^z = 0$. This allows us to define the canonical Hamiltonian $H_0 = p_a \dot{x}^a + p_z \dot{x}^z - L$, such that we have a set of $R + 1$ first-order HJPDE

$$H'_\alpha \equiv p_\alpha - \partial L / \partial \dot{t}^\alpha = 0, \quad \alpha = 0, 1, 2, \dots, R, \quad (1)$$

where $t^\alpha = x^0 \equiv t, x^z$. The Cauchy's method allows to relate (1) to a set of total differential equations, the characteristic equations (CE)

$$dx^a = \frac{\partial H'_\alpha}{\partial p_a} dt^\alpha, \quad dp_a = -\frac{\partial H'_\alpha}{\partial x^a} dt^\alpha, \quad dS = p_a dx^a - H_\alpha dt^\alpha. \quad (2)$$

Solutions of the first two equations are curves of $R + 1$ parameters t^α on the phase space defined by the conjugated variables (x^a, p_a) . From (2), and considering t^α as independent parameters, the evolution of any phase space function $F = F(x^a, t^\alpha, p_a, p_\alpha)$ is given by the fundamental differential

$$dF = \{F, H'_\alpha\} dt^\alpha, \quad (3)$$

where the Poisson Brackets (PB) are defined on the extended phase space of the variables $(x^a, t^\alpha, p_a, p_\alpha)$. Therefore the canonical constraints H'_α play the role of generators of the dynamical evolution of the system: they are considered as the Hamiltonians of the system. Note that for a regular system, the CE (3) becomes the Hamilton's equations.

A complete solution of the set of HJPDE is given by a family of surfaces ortogonal to the characteristic curves. The Frobenius' Integrability Condition (IC) assures the existence of this solution, as well as the independence between the parameters t^α and it is expressed in terms of vector fields tangent to the family of surfaces. In terms of the Hamiltonians H'_α , which are generators of these vectors, the IC is written as

$$\{H'_\alpha, H'_\beta\} = C_{\alpha\beta}^\gamma H'_\gamma, \quad (4)$$

i.e., the Hamiltonians must close a Lie algebra with the PB. Hamiltonians that satisfy (4) are called involutive Hamiltonians. The presence of Hamiltonians that do not satisfy the above condition, the so called non-involutive Hamiltonians, is an indicator that the system of HJPDE is not complete, or the parameters are not independent. In this case, the equivalent IC $dH'_\alpha = 0$ should provide new constraints to the system, as well as indicate the dependence between the parameters [13]. The HJ constraint analysis is actually the search for a complete set of involutive Hamiltonians of the system.

It is possible that after a complete set of HJPDE is found, there remains a subset of non-involutive Hamiltonians H'_x . In this case we may proceed with the method outlined in [13], and build the matrix with elements $M_{xy} \equiv \{H'_x, H'_y\}$. This way redefines the dynamics by eliminating the parameters related to these Hamiltonians with the use of the Generalized Brackets (GB). If the M matrix is singular of rank $K \leq R$, we have a regular submatrix $M_{\bar{a}\bar{b}}$, with $\bar{a} = 1, 2, \dots, K$, and we define the GB by

$$\{A, B\}^* = \{A, B\} - \{A, H'_a\}(M_{\bar{a}\bar{b}})^{-1}\{H'_b, B\}. \quad (5)$$

We can write the fundamental differential as

$$dF = \{F, H'_\alpha\}^* dt^{\bar{\alpha}}, \quad \bar{\alpha} = 0, K+1, \dots, R. \quad (6)$$

After the procedure of finding possible new Hamiltonians and eliminating possible dependence between the parameters, the Hamiltonians H'_α become the complete set of involutive Hamiltonians of the system, this time with the PB substituted by the GB.

3 The BF Model

Let us consider a gauge group G acting on fields of a two dimensional manifold \mathcal{M} . The BF theory consists on the gauge connection 1-form A and a scalar field B , also called Background field, whose action is given by

$$I_{BF}[B, A] = \int_{\mathcal{M}} \text{Tr}(B \wedge F), \quad (7)$$

where the trace is taken on the adjoint representation of G , and the field strength F is related to the gauge connection 1-form A by the covariant exterior derivative $F = DA = dA + A \wedge A$. The corresponding EL equations are

$$F = 0, \quad DB = 0. \quad (8)$$

The interpretation of these EL equations is straightforward: we have a flat connection A , since it has zero field strength, while the B field is parallel to A . With this machinery we may proceed to make a model for the gauge field of the Poincaré group $ISO(1; 1)$, in which we write the connection A as a combination of the 1-form zweibein e^I and the 1-form spin connection ω

$$A = e^I P_I + \omega \Lambda,$$

where $I = 0, 1$, and P_I and Λ are the generators of translations and Lorentz boost respectively. The zweibein and the spin connection are considered independent, and the Poincaré algebra is given by

$$[\Lambda, P_I] = \epsilon_I^J P_J, \quad [P_I, P_J] = 0, \quad (9)$$

the Levi-Civita symbol is defined such that $\epsilon_{01} = 1$.

The Killing metric is defined by these generators as

$$g_{ab} \equiv \text{Tr}(J_a J_b),$$

with $J_0 = P_0$, $J_1 = P_1$, and $J_2 = \Lambda$. In the two dimensional case, the Killing metric is degenerated, then, it is not possible to build a consistent gauge theory. To solve this problem it is customary to

consider the presence of a non-vanishing cosmological constant k , so we can deform the Poincaré algebra (9) to the (anti) de Sitter algebra

$$[\Lambda, P_I] = \epsilon_I^J P_J, \quad [P_I, P_J] = k\epsilon_{IJ}\Lambda. \quad (10)$$

In this case the Killing metric results to be invariant and non-degenerate:

$$g_{ab} = \begin{pmatrix} k\eta_{IJ} & 0 \\ 0 & 1 \end{pmatrix}.$$

The (anti) de Sitter algebra is expressed in terms of the generators J_a by $[J_a, J_b] = f_{ab}^c J_c$, where $f_{ab}^c \equiv \epsilon_{abc}g^{cd}$, and $\epsilon_{012} = 1$.

In terms of the zweibein and the spin connection we have the field strength

$$F = (de^I - \omega\epsilon^I_J \wedge e^J) P_I + \left(d\omega + \frac{k}{2}\epsilon_{IJ}e^I \wedge e^J \right) \Lambda \equiv T^I P_I + \mathcal{R}\Lambda,$$

where T^I and \mathcal{R} represent the torsion and curvature 2-forms of the zweibein field in the first order formalism. The flat connection equation $F = 0$ becomes

$$de^I - \omega\epsilon^I_J \wedge e^J = 0, \quad (11a)$$

$$d\omega + \frac{k}{2}\epsilon_{IJ}e^I \wedge e^J = 0. \quad (11b)$$

Equation (11a) is called the torsion-free equation $T^I = 0$, and along with (11b), and considering an invertible zweibein, it is possible to compute the spin connection in terms of e^I . In this case equation (11b) becomes the equation of constant Ricci's scalar curvature $R = k$. This is the standard procedure to show the equivalence between the two dimensional BF gravity with the JT model (see [4] and [12]).

4 The Hamilton-Jacobi analysis

Instead of using the differential forms, it is preferred to use its components $A = A_\mu^a J_a dx^\mu$ and $B = B^a J_a$, so the action can be written as

$$I_{BF}[B, A] = \int dx^2 B_a (\partial_0 A_1^a - \partial_1 A_0^a + f_{bc}^a A_0^b A_1^c). \quad (12)$$

In terms of the components, the flat connection equation becomes

$$\partial_0 A_1^a - \partial_1 A_0^a + f_{bc}^a A_0^b A_1^c = 0, \quad (13)$$

while the equation for the Background field becomes

$$\delta_0^\mu D_1 B_a - \delta_1^\mu D_0 B_a = 0, \quad (14)$$

where $D_\mu B_a \equiv \partial_\mu B_a + f_{ab}^c A_\mu^b B_c$. It is clear that (14) is equivalent to $D_\mu B_a = 0$.

It is well known that the elimination of the divergence terms in the Lagrangian density does not modify the EL equations, but it changes the functional form of the canonical momenta. Therefore we consider the equivalent action

$$I_{BF}[B, A] = \int d^2x \mathcal{L} = \int d^2x (B_a \partial_0 A_1^a + A_0^a D_1 B_a). \quad (15)$$

The absence of the $\partial_0 B_a$ term as well as the linearity of the velocity of the field A_1^a , make of (15) a first order action. Since first order actions are written in the canonical form $L = \dot{q}^i p_i - H_0$, so we clearly identify the canonical conjugated momenta

$$\pi_a^\mu \equiv \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu^a)} = \delta_1^\mu B_a, \quad \Pi^a \equiv \frac{\partial \mathcal{L}}{\partial(\partial_0 B_a)} = 0,$$

as canonical constraints of the theory. The canonical Hamiltonian density $\mathcal{H}_0 \equiv \pi_a^\mu \partial_0 A_\mu^a + \Pi^a \partial_0 B_a - \mathcal{L}$ is given by

$$\mathcal{H}_0 = -A_0^a D_1 B_a.$$

We have then the set of HJPDE

$$\mathcal{H}' \equiv \pi + \mathcal{H}_0 = 0, \quad \rightarrow x^0 \equiv t, \quad (16a)$$

$$\mathcal{H}_a'^0 \equiv \pi_a^0 = 0 \quad \rightarrow A_0^a \equiv \lambda_0^a, \quad (16b)$$

$$\mathcal{H}_a'^1 \equiv \pi_a^1 - B_a = 0 \quad \rightarrow A_1^a \equiv \lambda_1^a, \quad (16c)$$

$$\mathcal{H}'^a \equiv \Pi^a = 0 \quad \rightarrow B_a \equiv \epsilon_a. \quad (16d)$$

The left hand side of these equations are the Hamiltonian densities for which we have related the set of parameters $(x^0, \lambda_\mu^a, \epsilon_a)$. The Fundamental PB of the theory are given by

$$\{A_\mu^a(x), \pi_b^\nu(y)\} = \delta_b^a \delta_\mu^\nu \delta(x-y), \quad \{B_a(x), \Pi^b(y)\} = \delta_b^a \delta(x-y), \quad (17)$$

and the fundamental differential is given by

$$dF = \int dy [\{F, \mathcal{H}'(y)\}dt + \{F, \mathcal{H}_a'^\mu(y)\}d\lambda_\mu^a(y) + \{F, \mathcal{H}'^a(y)\}d\epsilon_a(y)]. \quad (18)$$

The fundamental differential is used to test the IC of the system in order to find another possible Hamiltonian densities to complete the constraint analysis constraint of the system. We may proceed in another way: we analyse the algebra of the previous Hamiltonian densities, build the GB, and then find the remaining constraints. For this particular case, the later method simplifies calculations and clarifies certain aspects of the formalism.

Notice that the subset of Hamiltonian densities $h^0 \equiv \mathcal{H}'^a$ and $h^1 \equiv \mathcal{H}_a'^1$ is not involutive, then we build the matrix $M(x, y)$ with components

$$M^{rs}(x, y) \equiv \{h^r(x), h^s(y)\} = \begin{pmatrix} 0_{3 \times 3} & 1_{3 \times 3} \\ -1_{3 \times 3} & 0_{3 \times 3} \end{pmatrix} \delta(x-y),$$

where $r, s = 0, 1$. The inverse matrix $M^{-1}(x, y)$ is given by

$$M_{rs}^{-1}(x, y) = \begin{pmatrix} 0_{3 \times 3} & -1_{3 \times 3} \\ 1_{3 \times 3} & 0_{3 \times 3} \end{pmatrix} \delta(x-y).$$

With this matrix, we eliminate the parameters (t_1^a, ϵ_a) related to the non-involutive Hamiltonian densities by building the GB

$$\{F(x), G(y)\}^* = \{F(x), G(y)\} - \int dz dw \{F(x), h^r(z)\} M_{rs}^{-1}(z, w) \{h^s(w), G(y)\}.$$

The only non-zero GB are given by

$$\{A_\mu^a, \pi_b^\nu\}^* = \delta_b^a \delta_\mu^\nu \delta(x-y), \quad \{A_\mu^a(x) B_b(y)\}^* = \delta_b^a \delta_\mu^1 \delta(x-y). \quad (19)$$

The reduction of the phase space is now evident, since B_b and Π^b are not canonical conjugated anymore. In fact, now B_b plays the role of π_b^1 . It is still necessary to test the integrability of the

Hamiltonian \mathcal{H}_a^0 . For this purpose we notice that after the construction of the GB the fundamental differential is now given by

$$dF = \int dy [\{F, \mathcal{H}'\}^* dt + \{F, \mathcal{H}_a^{\prime\mu}\}^* d\lambda^a],$$

where we have renamed $\lambda^a \equiv \lambda_1^a$.

Since $\{\mathcal{H}_a^0, \mathcal{H}'\}^* = -D_1 B_a$, to achieve integrability we have to consider the new Hamiltonian density

$$C'_a \equiv D_1 B_a = 0. \quad (20)$$

It is straightforward to see that the IC for C'_a is satisfied they are in involution with the other Hamiltonians under the GB. In particular, we have the algebra

$$\{C'_a(x), C'_b(y)\}^* = f_{ab} {}^c C'_c(x) \delta(x - y), \quad (21)$$

and the IC programme is closed for this system.

5 The Characteristic Equations

In the HJ description, we consider the complete set of involutive Hamiltonians \mathcal{H}' , \mathcal{H}_a^0 and C'_a as generators of the dynamical evolution of the system. Note that the functions \mathcal{H}' and \mathcal{H}_a^0 are related to the parameters t and λ^a respectively. However, C'_a is not related to any variable of the theory. To consider this Hamiltonian as a generator we expand the phase-space with a new set of variables ω^a . The fundamental differential of the system is then given by

$$dF = \int dy [\{F, \mathcal{H}'(y)\}^* dt + \{F, \mathcal{H}_a^0(y)\}^* d\lambda^a(y) + \{F, C'_a(y)\}^* d\omega^a(y)]. \quad (22)$$

From (22) we obtain the CE

$$dB_a = -f_{ab} {}^c B_c [A_0^b dt - d\omega^b], \quad (23a)$$

$$d\Pi^a = 0, \quad (23b)$$

and

$$dA_\mu^a = \delta_\mu^0 d\lambda^a + \delta_\mu^1 D_1 A_0^a dt - \delta_\mu^1 D_1 d\omega^a, \quad (24a)$$

$$d\pi_a^\mu = \delta_\mu^0 [D_1 B_a] dt - \delta_\mu^1 f_{ab} {}^c B_c [A_0^b dt - d\omega^b]. \quad (24b)$$

These equations should be equivalent to the EL equations (13) and (14). Because of the presence of the parameters ω^a , we need to clarify under what conditions the equivalence of these equations are valid.

The integrability assures independence between the parameters (t, λ^a, ω^a) , so the dynamics in the direction of any parameter is independent of the others. In this case, we have that temporal evolution alone gives

$$D_0 B_a = 0, \quad (25a)$$

$$\partial_0 \Pi^a = 0, \quad (25b)$$

and

$$\partial_0 A_\mu^a = \delta_\mu^1 D_1 A_0^a, \quad (26a)$$

$$\partial_0 \pi_a^\mu = \delta_\mu^0 D_1 B_a - \delta_\mu^1 f_{ab} {}^c B_c A_0^b. \quad (26b)$$

We see that (25a) is one of the EL equations for the B field (14). Eq. (25b) represents the IC for the Hamiltonian (16d), note that this relation is related to the fact that the B_a fields serve as Lagrange multipliers in the action. The component $\mu = 0$ for eq. (26a) states that A_0^a has no dynamics, while the component $\mu = 1$ is the EL equation (13). For $\mu = 0$, (26b) yields $\partial_0 \pi_a^0 = D_1 B_a = C_a'$, which is zero by imposition of integrability. Therefore, this equation reproduces the EL equation $D_1 B_a = 0$. At last, for $\mu = 1$ (26b) gives

$$\partial_0 \pi_a^1 + f_{ab} {}^c B_c A_0^b = 0.$$

Using (25a) it becomes $\partial_0 \pi_a^1 = \partial_0 B_a$, which corroborates the interpretation of B_a as a conjugate momenta related to A_1^a . Therefore, we showed that the time evolution of the characteristic equations, (25-26), is completely equivalent to the EL equations (13) and (14).

The evolution in the direction of the parameters (λ^a, ω^a) , on the other hand, gives the following set of infinitesimal transformations

$$\delta A_\mu^a = \{A_\mu^a, \mathcal{H}_b^0\}^* \delta \lambda^b + \{A_\mu^a, C_b'\}^* \delta \omega^b = \delta_\mu^0 \delta \lambda^a - \delta_\mu^1 D_1 \delta \omega^a, \quad (27a)$$

$$\delta B_a = \{B_a, C_b'\}^* \delta \omega^b = f_{ab} {}^c B_c \delta \omega^b. \quad (27b)$$

These variations left invariant the action (12) if we choose

$$\delta \omega^a = -\xi^a, \quad \delta \lambda^a = D_0 \xi^a, \quad (28)$$

in other words, the BF theory is invariant under the gauge transformation

$$\delta A_\mu^a = D_\mu \xi^a, \quad (29a)$$

$$\delta B_a = f_{bc} {}^a B_b \xi^c. \quad (29b)$$

Since the set of involutive Hamiltonians (\mathcal{H}_b^0, C_b') generates the transformation (27), we have that the generator of the gauge transformation is given by

$$\begin{aligned} G &= \int dy [\mathcal{H}_a^0(y) d\lambda^a(y) + C_a'(y) d\omega^a(y)], \\ &= \int dy [\mathcal{H}_a^0(y) D_0 \xi^a(y) - C_a'(y) \xi^a(y)], \end{aligned} \quad (30)$$

and it can be directly verified from the GB that

$$\begin{aligned} \{B_a, G\}^* &= f_{ab} {}^c B_c \delta \omega^b, \\ \{A_\mu^a, G\}^* &= \delta_\mu^0 \delta \omega^a - \delta_\mu^1 D_1 \delta \omega^a. \end{aligned}$$

6 Final Remarks

In this work we have dealt with the Hamilton-Jacobi constraint structure of the two dimensional BF gravity. The Frobenius' integrability conditions are a cornerstone of the mathematical structure of this formalism. We saw in section 4 that the Poisson algebra of the canonical constraints of the BF model allowed the construction of the GB, due to the presence of the two non-involutive Hamiltonian densities $(\mathcal{H}_a^1, \mathcal{H}^a)$. We built the GB and by proceeding with the integrability we obtained the set $(\mathcal{H}', \mathcal{H}_a^0, C_a')$ as the complete set of involutive Hamiltonian densities of the theory.

We built the fundamental differential (22), from where we obtained the characteristic equations (23-24) of the system. These equations depend explicitly of the parameter ω^a related to the Hamiltonian C_a' , and at principle this dependence does not spoil the equivalence between the CE and the EL equations of the system. This is so because integrability implies that the parameters must be independents, therefore time evolution is independent of the dynamics of the other parameters.

As a final remark we notice that all the involutive Hamiltonians contribute to the construction of the gauge generator (30) and now we can count the degrees of freedom of this theory. The dimension of the phase-space of the BF model is eighteen, six of these dimensions are related to the (B_a, Π^a) canonical variables, and twelve are related to (A_μ^a, π_a^ν) . The construction of the GB reduces the phase-space since we identify the variable B_a with one conjugated momenta π_a^1 and eliminate Π^a . Our reduced phase-space has now the dimension of twelve. On the other hand, we have six generators of gauge transformation, three for each involutive Hamiltonians. These generators reduce the number the dynamical variables in twelve. As a result we have zero degrees of freedom for BF gravity.

7 Acknowledgements

M. C. Bertin was supported by FAPESP. B. M. Pimentel was partially supported by CNPq and CAPES. C. E. Valcárcel was supported by CAPES.

References

- [1] J. D. Brown, *Lower Dimensional Gravity*, World Scientific Pub. Co. Inc. (1988).
- [2] C. Teitelboim, Phys. Lett. **B126** 41, 46, 49 (1983);
R. Jackiw, C. Teitelboim, Quantum Theory of Gravity, S. Christensen (editor), Bristol (1984);
R. Jackiw, Nucl. Phys. **B252**, 343 (1985).
- [3] R. Jackiw, Two Lectures on two-dimensional gravity, gr-qc/9511048.
D. Grumiller, W. Kummer, D. V. Vassilevich, Phys. Rept. **369**, 327 (2002).
- [4] K. Isler, C. A. Trugenberger, Phys. Rev. Lett. **63**, 834 (1989).
- [5] Clisthenis. P. Constantinidis, Jose Andre Lourenco, Ivan Morales, Olivier Piguet, Alex Rios, Class. Quant. Grav. **25**, 125003 (2008);
Clisthenis. P. Constantinidis, Olivier Piguet, Alejandro Perez, Phys. Rev. **D79**, 084007 (2009).
- [6] Y. Güler, Il Nuovo Cimento **B100**, 251 (1987);
Y. Güler, J. Math. Phys. **30**, 785 (1992);
Y. Güler, Il Nuovo Cimento **B107**, 1398 (1992).
- [7] C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the First Order*, American Mathematical Society; 3rd edition (1999).
- [8] B. M. Pimentel, R. G. Teixeira, Il Nuovo Cimento **B111**, 841 (1996);
B. M. Pimentel, R. G. Teixeira, Il Nuovo Cimento **B113**, 805 (1998);
B. M. Pimentel, R. G. Teixeira, J. L. Tomazelli, Ann. Phys. **267**, 75 (1998);
B. M. Pimentel, P. J. Pompeia, J. F. da Rocha-Neto, R. G. Teixeira, Gen. Rel. Grav. **35**, 877 (2003);
M. C. Bertin, B. M. Pimentel, P. J. Pompeia, Mod. Phys. Lett. **A20**, 2873 (2005);
B. M. Pimentel, P. J. Pompeia, J. F. da Rocha-Neto, Il Nuovo Cimento **B120**, 981 (2005);
M. C. Bertin, B. M. Pimentel, P. J. Pompeia, Ann. Phys. **323**, 527 (2008);

- M. C. Bertin, B. M. Pimentel, C. E. Valcárcel, *Ann. Phys.* **323**, 3137 (2008);
M. C. Bertin, B. M. Pimentel, C. E. Valcárcel, G. E. Zambrano, *AIP Conf. Proc.* **1296**, 402 (2010).
- [9] D. Baleanu, Y. Güler, *Nuovo Cimento B* **117**, 917 (2002);
S. I. Muslih, *Gen. Rel. Grav.* **36**, 1151 (2004).
- [10] M. C. Bertin, B. M. Pimentel, P. J. Pompeia, *Ann. Phys.* **325**, 2499 (2010).
- [11] M. C. Bertin, B. M. Pimentel, C. E. Valcárcel, G. E. R. Zambrano, *Class. Quantum Grav.* **28**, 175015 (2011).
- [12] M. Weis, *Topological Aspects of Quantum Gravity*, Thesis accepted for the Ph. D. degree in Physics at the Faculty of Science, University of Copenhagen 1997.
- [13] M. C. Bertin, B. M. Pimentel, C. E. Valcárcel, *Ann. Phys.* **323**, 3137 (2008).

Higgs-Strahlung process in the $SU(3)_L \otimes U_N(1)$ Model with Heavy Leptons

D. Romero^a and O. Pereyra ^a

^a Facultad de Ciencias
Universidad Nacional de Ingeniería
Av. Túpac Amaru, 210, Rimac
Lima - Perú

Abstract

In this work we study the Higgs-Strahlung process because this is one of the major candidate for lead the Higgs boson in the prospective International Linear Collider (ILC). We analyze this process in the frame of the electroweak extension of the Standard Model $SU(3)_L \otimes U_N(1)$ with heavy leptons. In that order we need to develop the scalar sector of the model after the Spontaneous Electroweak Symmetry Breaking (EWSB) to identify the potential mediators of the interaction. From there we calculate the total cross section for this process and examine its variation as a function of the C.M. energy for the range of values established by existing colliders.

Dedicated to the Professor Holger Valqui

1 Introduction

Based on the construction of the proposed (ILC) International Linear Collider, we investigate one of the main processes that would account for the existence of the Higgs boson, the Higgs-strahlung process [1]. The experimental discovery of this particle (Higgs Boson) would set the basis for the mass generation mechanism, spontaneous electroweak symmetry breaking (EWSB) proposed almost fifty years ago by Higgs, Brout, Englert, Guralnik, Hagen and Kibble [2], since according to this, the Higgs boson is responsible for generate mass to the other particles which includes the model. Currently the only collider that is trying to detect the Higgs is the LHC(Large Hadron Collider) at CERN. In this accelerator the collisions are given between protons, hoping to reach energies up to 14 TeV in the CM. There are no official data about the detection of the Higgs in the two detectors CMS and ATLAS, but there is speculation that they will soon have conclusive results for a Higgs mass of 124 GeV around [1]. One of the latest attempts to perform an accurate detection of the Higgs has been planning what would be the International Linear Collider (ILC), is predicted that this new collider begin construction by 2020, which consist in two linear accelerators about 31 km in length in which collide electrons e^- and positrons e^+ with energies of 500 GeV in the CM, allowing in a second stage of the project to extend the accelerator to 50 km, reaching energies of 1 TeV. [3]. Although the effective collision energy at the LHC is greater, the measurements in the ILC would be more accurate. The collisions between electrons and positrons are much simpler to analyze than those in which energy is distributed among their constituents (quarks, antiquarks, gluons, baryonic particles, etc.) as in the LHC. As one of the main roles of the ILC would be to precise measurements of particles discovered at the LHC. We perform the study of this process in the extension $SU(3)_L \otimes U_N(1)$ of the electroweak Standard Model. We find such extension useful because realizes the appearance of a set of new particles like $U^{\pm\pm}$, V^\pm and Z' bosons. We want to highlight before boarding the calculation in the $SU(3)_L \otimes U_N(1)$ electroweak model, we reproduced the total cross section of Higgs-strahlung process in the SM as shown in the literature [4]; [5].

2 The model

We are working in the “ $SU(3)_L \otimes U(1)_N$ electroweak extension with heavy leptons” proposed by Pleites and Tonasse [6]. In this model They enlarge the group symmetry, so that compared with the SM its basic structure is not altered. This leads to an increase in the content of particles, such as Z'_0 and the bileptons (V^\pm, U^\pm).

We are going to divide the full electroweak Lagrangian in sectors, depending on particle content and interactions:

$$\mathcal{L}_{331} = \mathcal{L}_L + \mathcal{L}_B + \mathcal{L}_H + \mathcal{L}_Y$$

where \mathcal{L}_L is the leptonic Lagrangian density, \mathcal{L}_B is the bosonic Lagrangian density, \mathcal{L}_H is the Higgs Lagrangian density and \mathcal{L}_Y is the Yukawa Lagrangian density. We consider important to note that both the leptonic sector and the bosonic sector will not be treated here because these have already been developed in previous work [6], [7], but we will focus on the scalar

sector, due to its dependence the terms of interaction on needed to analyze Higgs-strahlung process.

2.1 Scalar Sector

As mention at the beginning, the full development of this sector is necessary to perform the computation of the total cross section of Higgs-Strahlung process; which is the objective of this work. This requires explicitly determine the kinetics of the Lagrangian and obtain the relationship between the fields of the symmetry and the physical fields of the model. In this context, it is necessary to generate mass to all particles in the model, three scalar triplets with the following representation [8]:

$$\eta = \begin{pmatrix} \eta^0 \\ \eta_1^- \\ \eta_2^+ \end{pmatrix} (3, 0); \quad \rho = \begin{pmatrix} \rho^+ \\ \rho^0 \\ \rho^{++} \end{pmatrix} (3, 1); \quad \chi = \begin{pmatrix} \chi^- \\ \chi^{--} \\ \chi^0 \end{pmatrix} (3, -1) \quad (2.1)$$

This model proposes a scalar Lagrangian based on the introduced fields:

$$\mathcal{L}_H = T + V \quad (2.2)$$

where:

$$T = (D_\mu \eta)^\dagger (D^\mu \eta) + (D_\mu \rho)^\dagger (D^\mu \rho) + (D_\mu \chi)^\dagger (D^\mu \chi) \quad (2.3)$$

the more general potential V has the form [9]:

$$V = V^{(2)} + V^{(3)} + V^{(4)}$$

with

$$V^{(2)} = \mu_1^2 \eta^\dagger \eta + \mu_2^2 \rho^\dagger \rho + \mu_3^2 \chi^\dagger \chi \quad (2.4)$$

$$V^{(3)} = \frac{f_1}{2} \epsilon^{ijk} \eta_i \rho_j \chi_k + \text{h.c.} \quad (2.5)$$

$$\begin{aligned} V^{(4)} = & \alpha_1 (\eta^\dagger \eta)^2 + \alpha_2 (\rho^\dagger \rho)^2 + \alpha_3 (\chi^\dagger \chi)^2 + \alpha_4 (\rho^\dagger \rho) (\eta^\dagger \eta) \\ & + \alpha_5 (\chi^\dagger \chi) (\eta^\dagger \eta) + \alpha_6 (\chi^\dagger \chi) (\rho^\dagger \rho) + \alpha_7 (\rho^\dagger \eta) (\eta^\dagger \rho) + \alpha_8 (\chi^\dagger \eta) (\eta^\dagger \chi) + \\ & + \alpha_9 (\chi^\dagger \rho) (\rho^\dagger \chi) + [\alpha_{10} (\chi^\dagger \eta) (\rho^\dagger \eta) + \text{h.c.}] \end{aligned} \quad (2.6)$$

the local gauge invariance of the Lagrangian requires that the covariant derivative takes the form:

$$D^\mu \varphi_k = \left[\partial^\mu - \frac{ig}{2} \lambda_j A_j^\mu + ig' B^\mu N_\varphi \right] \varphi_k \quad (2.7)$$

where $\varphi_k = \eta, \rho, \chi$.

The vacuum expectation values for the neutral scalar components are define as[8]:

$$\langle \eta^0 \rangle = \frac{v_\eta}{\sqrt{2}}, \quad \langle \rho^0 \rangle = \frac{v_\rho}{\sqrt{2}}, \quad \langle \chi^0 \rangle = \frac{v_\chi}{\sqrt{2}}$$

based on the EWSB we expand the scalar fields around the vacuum:

$$\eta = \frac{1}{\sqrt{2}} \begin{pmatrix} v_\eta + H_\eta(x) \\ 0 \\ 0 \end{pmatrix}, \quad \rho = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_\rho + H_\rho(x) \\ 0 \end{pmatrix}, \quad \chi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ v_\chi + H_\chi(x) \end{pmatrix}, \quad (2.8)$$

2.2 EWSB for the Kinetic term in the Scalar Lagrangian

After the EWSB and to develop the covariant derivative in function of the gauge fields in (2.3) we obtain:

$$\begin{aligned} T = & \frac{1}{2} (\partial_\mu H_\eta) (\partial^\mu H_\eta) + \frac{1}{2} (\partial_\mu H_\rho) (\partial^\mu H_\rho) + \frac{1}{2} (\partial_\mu H_\chi) (\partial^\mu H_\chi) + \Delta_\eta + \Delta_\rho + \Delta_\chi \\ & + \left[\frac{g^2}{4} (v_\eta^2 + v_\rho^2) \right] W_\mu^- W^{\mu+} + \left[\frac{g^2}{4} (v_\eta^2 + v_\chi^2) \right] V_\mu^- V^{\mu+} + \left[\frac{g^2}{4} (v_\rho^2 + v_\chi^2) \right] U_\mu^{--} U^{\mu++} \\ & + \frac{1}{2} (A_\mu^3, A_\mu^8, B_\mu) \frac{g^2}{4} \begin{pmatrix} v_\eta^2 + v_\rho^2 & \frac{1}{\sqrt{3}}(v_\eta^2 - v_\rho^2) & -2\frac{g'}{g}v_\rho^2 \\ \frac{1}{\sqrt{3}}(v_\eta^2 - v_\rho^2) & \frac{1}{3}(v_\eta^2 + v_\rho^2 + 4v_\chi^2) & \frac{2}{\sqrt{3}}\frac{g'}{g}(v_\rho^2 + 2v_\chi^2) \\ -2\frac{g'}{g}v_\rho^2 & \frac{2}{\sqrt{3}}\frac{g'}{g}(v_\rho^2 + 2v_\chi^2) & 4\frac{g'^2}{g^2}(v_\rho^2 + v_\chi^2) \end{pmatrix} \begin{pmatrix} A^{3\mu} \\ A^{8\mu} \\ B^\mu \end{pmatrix} \end{aligned} \quad (2.9)$$

where Δ_η , Δ_ρ y Δ_χ are the interaction terms.

$$\begin{aligned} \Delta_\eta = & \frac{g^2}{8} \left(\frac{c_{2W}}{c_W} \right)^2 Z_\mu Z^\mu (v_\eta^2 + 2v_\eta H_\eta + H_\eta^2) \\ & - \frac{g^2}{4\sqrt{3}} \sqrt{1 - 4s_W^2} \left(\frac{c_{2W}}{c_W} \right) Z_\mu Z^{\mu'} (v_\eta^2 + 2v_\eta H_\eta + H_\eta^2) \\ & + \frac{g^2}{8} \frac{(1 - 4s_W^2)}{3c_W^2} Z'_\mu Z'^{\mu'} (v_\eta^2 + 2v_\eta H_\eta + H_\eta^2) \\ & + \frac{g^2}{4} W_\mu^+ W^{\mu-} (v_\eta^2 + 2v_\eta H_\eta + H_\eta^2) \\ & + \frac{g^2}{4} V_\mu^+ V^{\mu-} (v_\eta^2 + 2v_\eta H_\eta + H_\eta^2) \end{aligned} \quad (2.10)$$

$$\Delta_\rho = 2g^2 s_W^2 A_\mu A^\mu (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2)$$

$$\begin{aligned}
& + \frac{g^2}{8} \left(\frac{c_{2W}}{c_W} \right)^2 Z_\mu Z^\mu (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2) \\
& + g^2 \left(\frac{s_W c_{2W}}{c_W} \right) A_\mu Z^\mu (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2) \\
& + \frac{g^2}{4\sqrt{3}} \frac{(1 - 10s_W^2)}{\sqrt{1 - 4s_W^2}} \left(\frac{c_{2W}}{c_W^2} \right) Z_\mu Z'^\mu (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2) \\
& + \frac{g^2}{\sqrt{3}} \frac{(1 - 10s_W^2)}{\sqrt{1 - 4s_W^2}} \left(\frac{s_W}{c_W} \right) A_\mu Z'^\mu (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2) \\
& + \frac{g^2}{24} \frac{(1 - 10s_W^2)^2}{(1 - 4s_W^2)} \left(\frac{1}{c_W^2} \right) Z'_\mu Z'^\mu (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2) \\
& + \frac{g^2}{4} W_\mu^- W^{\mu+} (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2) \\
& + \frac{g^2}{4} U_\mu^{--} U^{\mu++} (v_\rho^2 + 2v_\rho H_\rho + H_\rho^2)
\end{aligned} \tag{2.11}$$

$$\begin{aligned}
\Delta_\chi & = 2g^2 s_W^2 A_\mu A^\mu (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2) \\
& + \frac{g^2}{8} \left(\frac{s_W^2}{c_W} \right)^2 Z_\mu Z^\mu (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2) \\
& - g^2 \left(\frac{s_W^3}{c_W} \right) A_\mu Z^\mu (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2) \\
& + \frac{2g^2}{4\sqrt{3}} \frac{(7s_W^2 - 1)}{\sqrt{1 - 4s_W^2}} \left(\frac{s_W^2}{c_W^2} \right) Z_\mu Z'^\mu (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2) \\
& - \frac{2g^2}{\sqrt{3}} \frac{(7s_W^2 - 1)}{\sqrt{1 - 4s_W^2}} \left(\frac{s_W}{c_W} \right) A_\mu Z'^\mu (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2) \\
& + \frac{4g^2}{24} \frac{(7s_W^2 - 1)^2}{(1 - 4s_W^2)} \left(\frac{1}{c_W^2} \right) Z'_\mu Z'^\mu (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2) \\
& + \frac{g^2}{4} V_\mu^- V^{\mu+} (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2) \\
& + \frac{g^2}{4} U_\mu^{--} U^{\mu++} (v_\chi^2 + 2v_\chi H_\chi + H_\chi^2)
\end{aligned} \tag{2.12}$$

from [8]:

$$s_W = \frac{t}{(1 + 4t^2)^{1/2}}, \quad t = g'/g \tag{2.13}$$

From the interaction terms we group those containing the coupling between the fields $H_\varphi ZZ$, ($\varphi = \eta, \rho, \chi$) necessary to calculate the cross section of the process that interests us, resulting in:

$$\mathcal{L}_{H_\varphi ZZ} = \frac{g^2}{4} \left(\frac{s_W^2}{c_W} \right)^2 Z_\mu Z^\mu (v_\eta H_\eta + v_\rho H_\rho + v_\chi H_\chi) \tag{2.14}$$

Now, the diagonalization of the mass matrix shown in (2.9) was already developed in [7], which reproduce the results shown in [8], where it was found the following relationship between the physical fields and the gauge fields:

$$\begin{aligned} A_\mu^3 &= s_W A_\mu + c_W Z_\mu \\ A_\mu^8 &= -\sqrt{3} s_W A_\mu - \sqrt{3} s_W t_W Z_\mu - \frac{\sqrt{1-4s_W^2}}{c_W} Z'_\mu \\ B_\mu &= \sqrt{1-4s_W^2} A_\mu - t_W \sqrt{1-4s_W^2} Z_\mu - \sqrt{3} t_W Z'_\mu \end{aligned} \quad (2.15)$$

where $t = g'/g \equiv \tan \theta$.

2.3 EWSB for the Potential in the Scalar Lagrangian

The expansion of scalar fields for potential (2.5),(2.6) on the basis H_η , H_ρ y H_χ allows us to find the mass matrix for neutral fields [9]:

$$M^2 = \frac{1}{2} \begin{pmatrix} 8\alpha_1 v^2 - f_1 u \omega / v & 4\alpha_4 v u + f_1 \omega & 4\alpha_5 v \omega + f_1 u \\ 4\alpha_4 v u + f_1 \omega & 8\alpha_2 u^2 - f_1 v \omega / u & 4\alpha_6 u \omega + f_1 v \\ 4\alpha_5 v \omega + f_1 u & 4\alpha_6 u \omega + f_1 v & 8\alpha_3 \omega^2 - f_1 v u / \omega \end{pmatrix} \quad (2.16)$$

where:

$$v_\eta \equiv v, \quad v_\rho \equiv u, \quad v_\chi \equiv \omega$$

The diagonalization of this mass matrix allow us to find the physical states and their respective masses. Using the approximations and the results obtained in [9] and [10] we have:

$$m_{01}^2 \approx 4 \left(\frac{\alpha_2 u^4 - \alpha_1 v^4}{v^2 - u^2} \right), \quad (2.17)$$

$$m_{02}^2 \approx \left(\frac{v^2 + u^2}{2vu} \right) \omega^2, \quad (2.18)$$

$$m_{03}^2 \approx -4 \alpha_3 \omega^2, \quad (2.19)$$

and the relationship between the physical fields and the symmetry fields:

$$H_\eta \approx \frac{v}{(v^2 + u^2)^{1/2}} H_1^0 + \frac{u}{(v^2 + u^2)^{1/2}} H_2^0 \quad (2.20)$$

$$H_\rho \approx \frac{u}{(v^2 + u^2)^{1/2}} H_1^0 - \frac{v}{(v^2 + u^2)^{1/2}} H_2^0 \quad (2.21)$$

$$H_\chi \approx H_3^0 \quad (2.22)$$

From the values obtained for the Higgs masses, we note that one of them Eq (2.17) does not depends on the value of ω . Moreover we know that the expectation values for the scalar fields satisfy the relation $\omega \gg u, v$ [8] so we can identify the field H_1^0 with mass m_{01} as the Higgs boson predicted by the SM. If we replace the equations (2.20), (2.21) in (2.14), we obtain:

$$\mathcal{L}_{HZZ} = \frac{g^2}{4} \left(\frac{c_{2W}}{c_W} \right)^2 (v_\eta^2 + v_\rho^2)^{1/2} Z_\mu Z^\mu H_1^0 \quad (2.23)$$

Based on this Lagrangian, we calculate the total cross section of scattering for the non polarized process, obtaining:

$$\sigma(e^+e^- \rightarrow ZH) = \frac{\pi\alpha^2}{6} \frac{K}{\sqrt{s}} \frac{(K^2 + 3M_Z^2)(1 - 2x_W)^4(1 - 4x_W + 8x_W^2)}{(s - M_Z^2)^2 x_W^2(1 - x_W)^2} \quad (2.24)$$

where:

$$K = \sqrt{\frac{s^2 + (M_H^2 - M_Z^2)^2 - 2s(M_H^2 + M_Z^2)}{4s}}, \quad x_W = \sin^2 \theta_w$$

We have denoted s as the square of the CM energy.

3 Discussion

The experimental conditions in the proposed International Linear Collider (ILC)[3] provide us with an ideal environment for precise studies on the production of the Higgs with well defined initial conditions and cleaning signals without parallel. The Higgs-Strahlung process $e^- + e^+ \rightarrow Z + H$ is one of the main mechanisms for Higgs production at the ILC. With this process it is possible to detect the Higgs mass (M_H) either by directly decay or by reconstruction of the Z Boson decay in leptons e^-e^+ , $\mu^-\mu^+$ [11].

3.1 Constraints on the Higgs Mass

At present, the SM is strongly supported by experiments. However, the Higgs boson has not been observed experimentally. Nevertheless there are both experimental and theoretical restrictions on this fundamental parameter.

From direct searches of the Higgs boson, the lower limit is $M_H \geq 114$ GeV a 95% confidence level given by LEP2 [12], [11]. The recent combined CDF and $D\phi$ excludes the mass range between 160 and 170 GeV at 95% confidence level[13], [11].

From the theoretical standpoint there is an upper limit to which the behavior of the perturbation theory is valid, and a lower limit of stability derivative from the Higgs potential. If M_H is very large, the Higgs coupling diverges at a scale Λ below the Planck scale and if M_H is very small, then the scalar potential develops a second minimum (global) [14].

3.2 Results and Phenomenology

Based on the results obtained in the previous chapter, equation (2.24), we can express the cross section as a function of the cross section of SM, as follows:

$$\sigma_{HZ}^{331} = (1 - 2 \sin^2 \theta_W)^4 \sigma_{HZ}^{SM} \quad (3.1)$$

where σ_{HZ}^{SM} is the cross section of the SM with a Higgs boson of mass M_H , and θ_W is the Weinberg angle. Our goal now is to study the behavior of both cross sections, as a function of the CM energy (\sqrt{s}), for this, we plot both magnitudes for fixed values of the Higgs mass M_H , taking into account the restrictions detailed in the previous section.

Below is shown a graph of the total cross section σ (b) versus CM energy \sqrt{s} (GeV) for a fixed value of the Higgs mass M_H , in models 331 and SM.

Note that the range of energies that have been considered, are directly related to the possible values that could come in the future linear collider (ILC), which will cover energies from 500 GeV to 1 TeV. On the other hand we have excluded the energy values already verified in the LEP2 (0 – 200 GeV).

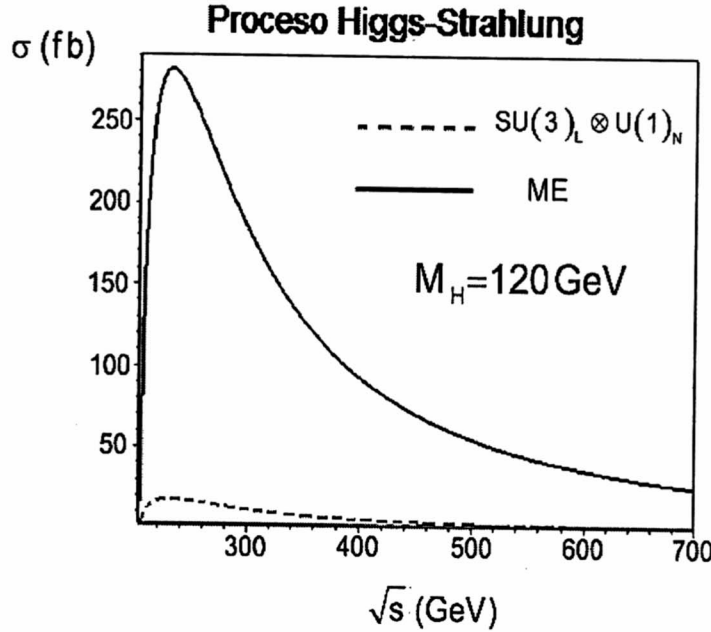


Figure 1: Comparative Plot of σ vs \sqrt{s} for the Higgs-Strahlung process in the $SU(3)_L \otimes U(1)_N$ and SM models, for the Higgs mass of $M_H = 120$ GeV.

Since the SM is the current reference, it was expected that the model $SU(3)_L \otimes U(1)_N$ present a total cross section of the same order of magnitude or otherwise the total cross section depends on parameters which when carried to the limit of SM effectively reproduce the same section. However this does not happen, because the Weinberg angle θ_W is a parameter that is independent of the model and that takes a value of approximately 0.23 [7].

One possible solution to this inconsistency could be solved, if we consider another possible Higgs boson couplings to the Z' . This was not considered initially because this boson, the Z' , is a hypothetical particle predicted by the model, but has not yet been detected experimentally.

Below we show a graph in which we analyze the variation of σ vs \sqrt{s} in the model $SU(3)_L \otimes U(1)_N$, for different values of the Higgs masses, according to the restrictions of these values bounded above:

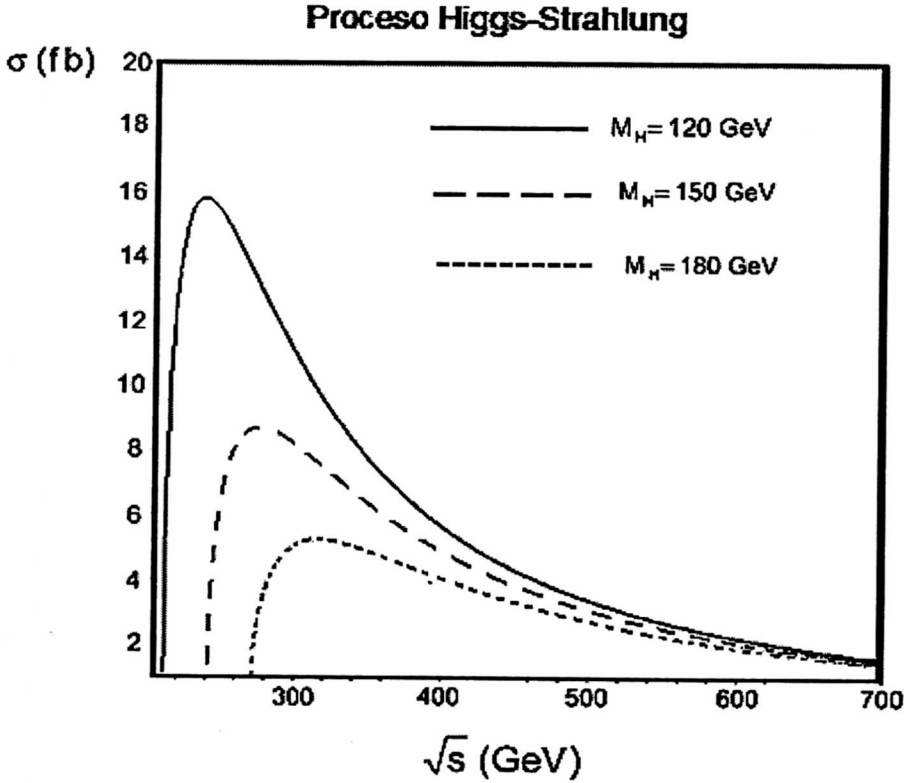


Figure 2: Plot of σ vs \sqrt{s} for the Higgs-Strahlung process in the $SU(3)_L \otimes U(1)_N$ model considering Higgs mass values of $M_H = 120, 150$ y 180 GeV.

The figure shows that as the Higgs mass increases from 120 GeV, the cross section decreases. Looking at one of the graphs for a fixed value of the Higgs mass, we note that the effective section [equation (2.24)] increases sharply at the threshold of a maximum at $\sqrt{s} \sim M_Z + \sqrt{2}M_H$, henceforth, is reduced as $1/s$ [15],[11].

Now we show a graph in which we analyze σ vs M_H in the model $SU(3)_L \otimes U(1)_N$. In this case we will keep fixed CM energy \sqrt{s} and vary the Higgs mass M_H .

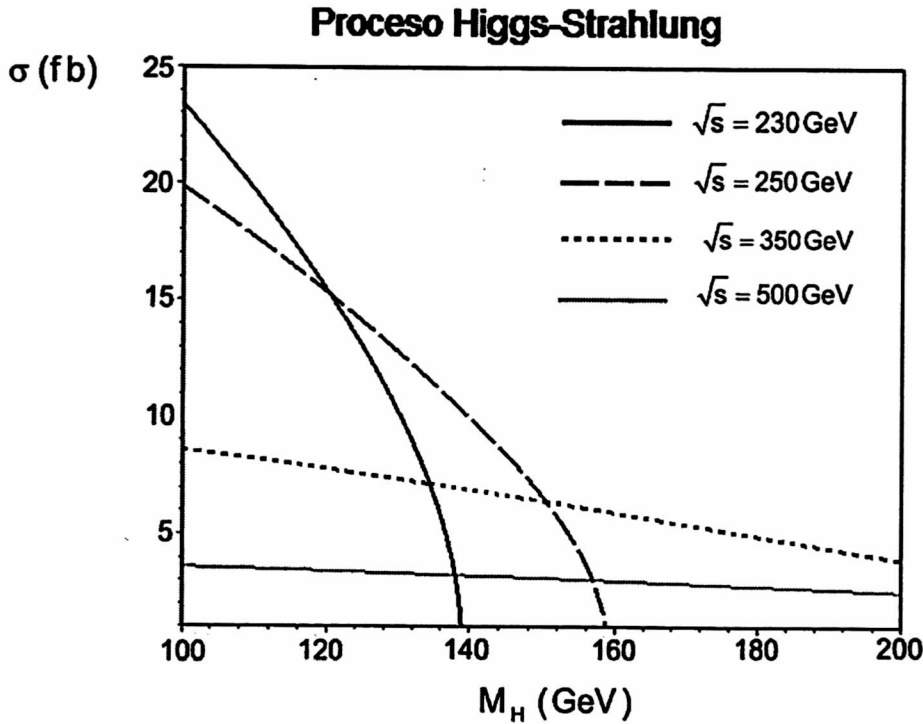


Figure 3: Plot of σ vs M_H for the Higgs-Strahlung process in the $SU(3)_L \otimes U(1)_N$ model considering C.M energy values of $\sqrt{s}=230, 250, 350$ y 500 GeV.

One can see that the cross section decreases as the Higgs mass increases, and for moderate values of the Higgs masses, the cross section is large enough for small CM energies. The last two are showing that to get the most effective section, you should choose the CM energies, close to the maxima of the graphs. It is for this reason that the Higgs-strahlung is the dominant production process for moderate values of energy.

References

- [1] J. Ellis, M. K. Gaillard and D. V. Nanopoulos, arXiv:1201.6045v1 [hep-ph] (2012).
- [2] P.W. Higgs, Phys. Rev. Lett. 13 (1964) 508; *ibid.* Phys. Rev. 145 (1966) 1156; F. Englert and R. Brout, Phys. Rev. Lett. 13 (1964) 321; G.S. Guralnik, C.R. Hagen and T. Kibble, Phys. Rev. Lett. 13 (1965) 585; T. Kibble, Phys. Rev. 155 (1967) 1554.
- [3] International Linear Collider, <http://www.linearcollider.org/> (2012).
- [4] W. Lee, C. Quigg and H. B. Thacker, Phys. Rev. D16, 1519 (1977).
- [5] C. Quigg, Gauge Theories Of Strong, Weak, And Electromagnetic Interactions, Westview Press, 1997.

- [6] V.Pleitez and M.Tonasse, Phys. Rev. D48 2353 (1993).
- [7] D. Romero, Tesis de Maestria, Universidad Nacional de Ingeniera, (2010).
- [8] F. Pisano and V. Pleitez, Phys. Rev. D46, 410 (1992).
- [9] Tonasse, M. D, Phys.Lett. B381 (1996).
- [10] N.T. Anh, N.A. Ky, H.N. Long, Int. J. Mod. Phys. A 15, 283 (2000).
- [11] Hegne Li, Tesis de Doctardo, Universite de Paris-Sud (2009).
- [12] Barate et al., LEP WG for Higgs boson searches, Phys. Lett. **B565**, 61 (2003).
- [13] Tevatron New Phenomena, Higgs Working Group, FERMILAB-PUB-09-060-E, CDF-NOTE- 9713, D-NOTE-5889, arXiv:0903.4001 (2009).
- [14] K. Nakamura *et al.* (Particle Data Group), JPG **583**, (2010).
- [15] A. Djouadi arXiv:hep-ph/0503172v2 (2005).

Noncommutative (generalized) sine-Gordon/massive Thirring correspondence, integrability and solitons

H. Blas ^a and H. L. Carrion ^b

a) Instituto de Física, Universidade Federal de Mato Grosso

Av. Fernando Correa, s/n, Coxipó

78060-900, Cuiabá - MT - Brazil

b) Escola de Ciências e Tecnologia / UFRN

Campus Universitário Lagoa Nova, CEP 59078-970 - Natal - RN - Brazil.

Abstract

Some properties of the correspondence between the non-commutative versions of the (generalized) sine-Gordon (NCGSG_{1,2}) and the massive Thirring (NCGMT_{1,2}) models are studied. Our method relies on the master Lagrangian approach to deal with dual theories. The master Lagrangians turn out to be the NC versions of the so-called affine Toda model coupled to matter fields (NCATM_{1,2}), in which the Toda field g belongs to certain subgroups of $GL(3)$, and the matter fields lie in the higher grading directions of an affine Lie algebra. Depending on the form of g one arrives at two different NC versions of the NCGSG_{1,2}/NCGMT_{1,2} correspondence. In the NCGSG_{1,2} sectors, through consistent reduction procedures, we find NC versions of some well-known models, such as the NC sine-Gordon (NCSG_{1,2}) (Lechtenfeld et al. and Grisaru-Penati proposals, respectively), NC (bosonized) Bukhvostov-Lipatov (NCbBL_{1,2}) and NC double sine-Gordon (NCDSG_{1,2}) models. The NCGMT_{1,2} models correspond to Moyal product extension of the generalized massive Thirring model. The NCGMT_{1,2} models possess constrained versions with relevant Lax pair formulations, and other sub-models such as the NC massive Thirring (NCMT_{1,2}), the NC Bukhvostov-Lipatov (NCBL_{1,2}) and constrained versions of the last models with Lax pair formulations. We have established that, except for the well known NCMT_{1,2} zero-curvature formulations, generalizations ($n_F \geq 2$, n_F = number of flavors) of the massive Thirring model allow zero-curvature formulations only for constrained versions of the models and for each one of the various constrained sub-models defined for less than n_F flavors, in the both NCGMT_{1,2} and ordinary space-time descriptions (GMT), respectively. The non-commutative solitons and kinks of the $GL(3)$ NCGSG_{1,2} models are investigated.

Keywords: Integrable hierarchies, non-commutativity, solitons, integrable field theories.

1 Introduction

Field theories in non-commutative (NC) space-times are receiving considerable attention in recent years in connection to the low-energy dynamics of D-branes in the presence of background B-field (see e.g. [1]). In particular, the NC versions of integrable systems (in two dimensions) are being considered [2]. On the other hand, conformal theories on the usual two-dimensional space-time play an important role in various aspects of modern physics, from string theory to applications in condensed matter. So, one might ask about the role played by QFTs in $(1+1)$ -dimensional non-commutative space-time. Indeed there is reason to believe that similar applications would emerge and they deserve further investigations, since it is possible to define notions of conformal invariance, Kac-Moody and Virasoro symmetries in this context [3]. Furthermore, there is some optimism regarding the following analogy with the usual known relationship: it is believed that the integrable models, defined on two-dimensional NC *Euclidean* space, would be the NC versions of statistical models in the critical points and in the off-critical integrable directions.

The sine-Gordon type and other related integrable systems have appeared frequently in diverse areas of physics, from condensed matter to string theory, in connection to such properties as soliton solutions, integrability and duality. So, the study of their properties and the search for their solutions have greatly attracted the interest of the scientific community. In condensed matter, we can mention for example the work [4] on the nonlinear dynamics of the inhomogeneous DNA double helices chain. In topics of string theory we can mention the recent works on the magnon-type solutions on the $R \times S_n$ ($n = 2, 3$) background geometry [5, 6].

Some non-commutative versions of the sine-Gordon model (NCSG) have been proposed in the literature [7, 8, 9, 10, 11, 12]. The relevant equations of motion have the general property of reproducing the ordinary sine-Gordon equation when the non-commutativity parameter is removed. The Grisaru-Penati version [7, 8] introduces a constraint which is non-trivial only in the non-commutative case. The constraint is required by integrability but it is satisfied by the one-soliton solutions. However, at the quantum level this model gives rise to particle production as was discovered by evaluating tree-level scattering amplitudes [8]. On the other hand, introducing an auxiliary field, Lechtenfeld et al. [12] proposed a novel NCSG model which seems to possess a factorisable and causal S-matrix.

Recently, in ordinary commutative space the so-called $sl(2)$ affine Toda model coupled to matter (Dirac) fields (ATM) has been shown to be a Master Lagrangian (ML) from which one can derive the sine-Gordon and massive Thirring models, describing the strong/weak phases of the model, respectively [13]-[16]. Besides, the ML approach was successfully applied in the non-commutative case to uncover related problems in $(2+1)$ dimensions regarding the duality equivalence between the Maxwell-Chern-Simons theory (MCS) and the Self-Dual (SD) model [17].

In this paper we extend some properties of the so-called $sl(3)$ generalized affine Toda model coupled to matter fields (GATM) [15] to the NC case. We define the NCATM model by replacing the products of

fields by the \star -products on the level of its effective action. The effective action associated to this model in ordinary space gives rise to equations of motion which can be derived from a zero-curvature equation plus some constraints. In fact, the ATM model is a constrained sub-model of an off-critical model related to the so-called conformal affine Toda model coupled to matter fields (CATM) which possesses a Lax pair formulation [18]. So, we expect the NCATM model defined in this way does not belong to those class of NC field theories associated to a Lax pair formulation [11]. The NC $GL(2)$ case has been considered in [9], there the master Lagrangians turn out to be the NC versions of the ATM model associated to the group $GL(2)$, in which the Toda field belongs to certain representations of either $U(1) \times U(1)$ or the complexified $U(1)_C$, such that they correspond to the Lechtenfeld et al. (NCSG₁) or Grisaru-Penati (NCSG₂) proposals for the NC versions of the sine-Gordon model, respectively. Besides, the relevant NC massive Thirring (NCMT_{1,2}) sectors are written for two (four) types of Dirac fields corresponding to the Moyal product extension of one (two) copy(ies) of the ordinary massive Thirring model. The NCSG_{1,2} models share the same one-soliton (real Toda field sector of model 2) exact solutions with their commutative counterparts, which are found without expansion in the NC parameter θ for the corresponding Toda field. Here the $GL(3)$ extension presents the above known feature regarding the appearance of two versions of the NC (generalized) sine-Gordon model (NCGSG_{1,2}) and the corresponding NC (generalized) massive Thirring models (NCGMT_{1,2}), and some new phenomena such as the appearance of the associated sub-models: three copies for each version of the NC sine-Gordon (NCSG_{1,2}) models, (bosonized) Bukhvestov-Lipatov models (NCbBL_{1,2}), double sine-Gordon models (NCDSG_{1,2}), and three copies for each version of the NC massive Thirring models (NCMT_{1,2}), Bukhvestov-Lipatov models (NCBL_{1,2}) and the constrained NCBL_{1,2} models, respectively. In addition, we have the known NC soliton solutions in the NCGSG_{1,2} sectors and the appearance of a NC kink type solutions for the NCDSG_{1,2} sub-models. Even though we have discussed the integrability properties of the NCGSG_{1,2} models only for certain integrable directions in field space, i.e. in the NCSG_{1,2} sub-models, the NC generalized massive Thirring (NCGMT_{1,2}) sectors present intriguing properties regarding integrability: the NCGMT_{1,2} models encompass a Lax pair formulation only for a sub-model with certain eqs. of motion provided that some constraints are satisfied. Moreover, we established the integrability of certain constrained versions of the NCBL_{1,2} models by providing a corresponding recipe to construct a Lax pair for each of them. The extension of the above features for the $GL(n)$ NCATM_{1,2} models are straightforward.

The study of these models become interesting since the $su(n)$ ATM theories constitute excellent laboratories to test ideas about confinement [16, 19], the role of solitons in quantum field theories [13], duality transformations interchanging solitons and particles [13, 20], as well as the reduction processes of the (two-loop) Wess-Zumino-Novikov-Witten (WZNW) theory from which the ATM models are derivable [18, 15]. Moreover, the ATM type systems may also describe some low dimensional condensed matter phenomena, such as self-trapping of electrons into solitons, see e.g. [21], tunneling in the integer quantum Hall effect [22], and, in particular, polyacetylene molecule systems in connection with fermion number fractionization [23]. It has been shown that the $su(2)$ ATM model describes the low-energy spectrum of QCD₂ (one flavor and

N colors in the fundamental and $N = 2$ in the adjoint representations, respectively)[19]. The $sl(3)$ ATM model and its related dual sub-models GSG/GMT have been used to provide a bag model like confinement mechanism for “quarks” and it has been shown that the ATM spectrum comprises of solitons as baryons and qualitons as constituent quarks in two-dimensional QCD [24]. Moreover, the $sl(3)$ GSG model has been found to describe the low energy effective action of QCD₂ with unequal ‘quark’ masses, three flavors and N colors. This model has recently been used to describe the normal and exotic baryon spectrum of QCD₂ [25].

The paper is organized as follows. In the next section we present the NC extensions of the ATM model relevant to our discussions. It deals with the choice of the group representation for the Toda field g . We introduce two types of master Lagrangians (NCATM_{1,2}), the *first* one defined for $g \in [U(1)]^3$ with the same content of matter fields as the ordinary ATM; the *second* one defined for two copies of the NCATM₁ such that in this case $g, \bar{g} \in \mathcal{H} \subset SL(3, \mathbb{C})$. In section 3 the non-commutative versions (NCGSG_{1,2}) of the generalized sine-Gordon model (GSG) are derived from the relevant master Lagrangians through reduction procedures resembling the one performed in the ordinary GATM \rightarrow GSG reduction. In section 4 we present the two NC extensions (NCGSG_{1,2}) of the GSG model, as well as their associated sub-models such as the NCSG_{1,2}, NCbBL_{1,2} and NCDSG_{1,2} models. In section 5 we ‘decouple’ on shell the theories NCGSG_{1,2} and NCGMT_{1,2}, respectively. We discuss the conditions which must satisfy the constraints in order to have a complete decoupling, in particular for the soliton solutions. In section 6 we consider the NCGMT_{1,2} models, as well as their global symmetries, associated currents and integrability properties of the constrained sub-models. In these developments the double-gauging of a U(1) symmetry in the star-localized Noether procedure to get the currents deserve a careful treatment. We discuss their associated sub-models such as the integrable NCMT_{1,2}, the non-integrable NCBL_{1,2}, and the (constrained) NCBL_{1,2} models regarded as integrable sub-models. In section 7 we present the soliton and kink type solutions as a sub-set of solutions satisfying the both GSG and NCGSG_{1,2} models simultaneously. Some discussions and possible directions of research to pursue in the future are presented in section 8. The Appendix A provides the usual GSG model as a reduced $sl(3)$ affine Toda model couple to matter. Some results of the zero-curvature formulation of the CATM model are provided in Appendix B, and the Lagrangian formulation of the ordinary ATM model is summarized in Appendix C.

2 The NC affine Toda models coupled to matter fields (NCATM_{1,2})

In this section we present the NC versions of the so-called affine Toda model coupled to matter fields (NCATM_{1,2}). The case of $GL(2)$ NCATM model has been studied at the classical level in [9] and the related NC sine-Gordon/massive Thirring correspondence has been considered at the quantum level in [10]. Even though we present detailed computations for the $GL(3)$ case it can follow directly for any $GL(n)$. Two different NC extensions of the ATM model (184) are possible as long as each of them reproduce its ordinary equations of motion in the commutative limit. The commutative Toda field g in (174) belongs to

the complexified abelian subgroup of $SL(3, \mathbb{C})$. The symmetry group $SL(3)$ of the ordinary ATM model (see Appendix B) when considered in the NC case is not closed under the Moyal product \star ; then, the NC extension requires the $GL(3)$ group. In the next steps we define two versions of the non-commutative $GL(3)$ affine Toda model coupled to matter fields (NCATM_{1,2}). Let us define the *first* NC extension (NCATM₁) as

$$\begin{aligned} S_{NCATM_1} &\equiv S[g, W^\pm, F^\pm] \\ &= I_{WZW}[g] + \int d^2x \sum_{m=1}^2 \left\{ \frac{1}{2} \langle \partial_- W_{3-m}^- \star [E_3, W_m^-] \rangle - \right. \\ &\quad \left. \frac{1}{2} \langle [E_{-3}, W_m^+] \star \partial_+ W_{3-m}^+ \rangle + \langle F_m^- \star \partial_+ W_m^+ \rangle + \right. \\ &\quad \left. \langle \partial_- W_m^- \star F_m^+ \rangle + \langle F_m^- \star g \star F_m^+ \star g^{-1} \rangle \right\}, \end{aligned} \quad (1)$$

where $F \star G = F \exp\left(\frac{\theta}{2}(\overleftarrow{\partial}_+ \overrightarrow{\partial}_- - \overleftarrow{\partial}_- \overrightarrow{\partial}_+)\right)G$ and $g \in [U(1)]^3$. In fact, we have written the NC version of the ATM model presented in the eq. (184) of Appendix C. The fields W_m^\pm, F_m^\pm , as well as the generators $E_{\pm 3}$ of the model are defined in eqs. (168)-(173). $I_{WZW}[g]$ is a NC generalization of the WZNW action for g

$$I_{WZW}[g] = \int d^2x \left[\partial_+ g \star \partial_- g^{-1} + \int_0^1 dy \hat{g}^{-1} \star \partial_y \hat{g} \star [\hat{g}^{-1} \star \partial_+ \hat{g}, \hat{g}^{-1} \star \partial_- \hat{g}]_\star \right], \quad (2)$$

where the homotopy path $\hat{g}(y)$ such that $\hat{g}(0) = 1$, $\hat{g}(1) = g$ ($[y, x_+] = [y, x_-] = 0$) has been defined. The WZW term in this case gives a non-vanishing contribution due to the non-commutativity. This is in contrast with the action in ordinary space, i.e. the WZW term in (184)-(185) vanishes for g belonging to an abelian subgroup of $SL(3, \mathbb{C})$. From (1) one can derive the set of equations of motion for the corresponding fields

$$\partial_-(g^{-1} \star \partial_+ g) = \sum_{m=1}^2 [F_m^-, g \star F_m^+ \star g^{-1}]_\star \quad (3)$$

$$\partial_+ F_m^- = [E_{-3}, \partial_+ W_{3-m}^+], \quad \partial_- F_m^+ = -[E_3, \partial_- W_{3-m}^-], \quad (4)$$

$$\partial_+ W_m^+ = -g \star F_m^+ \star g^{-1}, \quad \partial_- W_m^- = -g^{-1} \star F_m^- \star g. \quad (5)$$

Notice that these set of eqs. closely resemble their commutative counterparts (186)-(188) of Appendix C. Substituting the derivatives of W^\pm 's given in the eqs. (5) into the eqs. (4) one can get the equivalent set of equations

$$\partial_+ F_m^- = -[E_{-3}, g \star F_{3-m}^+ \star g^{-1}], \quad \partial_- F_m^+ = [E_3, g^{-1} \star F_{3-m}^- \star g]. \quad (6)$$

Notice that in the action (1) one can use simultaneously the cyclic properties of the group trace and the \star product. Then, the action (1) and the equations of motion (3)-(5) have the left-right local symmetries given by

$$g \rightarrow h_L(x_-) \star g(x_+, x_-) \star h_R(x_+), \quad (7)$$

$$F_m^+ \rightarrow h_R^{-1}(x_+) \star F_m^+(x_+, x_-) \star h_R(x_+), \quad W_m^- \rightarrow h_R^{-1}(x_+) \star W_m^-(x_+, x_-) \star h_R(x_+), \quad (8)$$

$$F_m^- \rightarrow h_L(x_-) \star F_m^-(x_+, x_-) \star h_L^{-1}(x_-), \quad W_m^+ \rightarrow h_L(x_-) \star W_m^+(x_+, x_-) \star h_L^{-1}(x_-). \quad (9)$$

The system of eqs. (3)-(5) is invariant under the above symmetries if the following conditions are supplied

$$h_R(x_+) \star E_3 h_R^{-1}(x_+) = E_3, \quad h_L^{-1}(x_-) \star E_{-3} h_L(x_-) = E_{-3}, \quad (10)$$

where $h_{L/R}(x_{\mp}) \in \mathcal{H}_0^{L/R}$, $\mathcal{H}_0^{L/R}$ being Abelian sub-groups of $GL(3)$. These symmetry transformations written in matrix form [15] are extensions of the ordinary ones to the NC case in a straightforward manner. Notice that in the ordinary space-time, in terms of the field components, the above transformations are given in the Appendix A [see eqs. (152) and (154)-(155)]; obviously, the form of the expressions given in these eqs. will change in the NC case.

Next, we define the *second* version of the $GL(3)$ NC affine Toda model coupled to matter NCATM₂ as

$$S_{NCATM_2} \equiv S[g, W^{\pm}, F^{\pm}] + S[\bar{g}, \mathcal{W}^{\pm}, \mathcal{F}^{\pm}], \quad (11)$$

where the independent fields g and \bar{g} , related to the set of matter fields $\{W^{\pm}, F^{\pm}\}$ and $\{\mathcal{W}^{\pm}, \mathcal{F}^{\pm}\}$, respectively, belong to a complexified subgroup \mathcal{H} of $GL(3)$ to be specified in the subsection 4.2. As above the action $S[\cdot, \cdot, \cdot]$ is defined as the Moyal extension of (184). The motivation to introduce a copy of the action functional with the set of fields $\bar{g}, \mathcal{W}^{\pm}, \mathcal{F}^{\pm}$ will be clarified below. Let us mention, in the mean time, that the second version of the NCATM₂ model has also been considered in [9] for the $SL(2)$ case.

The equations of motion for the NCATM₂ model (11) comprise the eqs. (3)-(5) written for $g \in \mathcal{H} \subset GL(3)$ and a set of analogous equations for the remaining fields $\bar{g}, \mathcal{F}^{\pm}$ and \mathcal{W}^{\pm} . Moreover, in addition to the symmetry transformations (7)-(9) one must consider similar expressions for $\bar{g}, \mathcal{F}^{\pm}$ and \mathcal{W}^{\pm} .

3 NC versions of the generalized sine-Gordon model (NCGSG_{1,2})

In order to derive the NC versions of the generalized sine-Gordon model (NCGSG_{1,2}) we follow the master Lagrangian approach [26, 15], starting from the NCATM_{1,2} models (1) and (11), respectively, as performed in the $GL(2)$ case [9]. So, let us consider first the equations of motion (3)-(5). We proceed by integrating the eqs. (4)

$$F^- = [E_{-3}, W_{3-m}^+] + f_m^-(x_-), \quad F^+ = -[E_3, W_{3-m}^-] - f_m^+(x_+). \quad (12)$$

with the $f^{\pm}(x_{\pm})$'s being analytic functions. Next, we replace the F^{\pm} of eqs. (12) and the $\partial_{\pm} W^{\pm}$ of (5), written in terms of W^{\pm} , into the action (1) to get

$$\begin{aligned} S'[g, W^{\pm}, f^{\pm}] &= I_{WZW}[g] + \int d^2x \sum_{m=1}^2 \left\{ \frac{1}{2} \langle [E_{-3}, W_{3-m}^+] \star g \star f_m^+ \star g^{-1} \rangle + \right. \\ &\quad \left. \frac{1}{2} \langle g^{-1} \star f_m^- \star g \star [E_3, W_{3-m}^-] \rangle + \langle g^{-1} \star f_m^- \star g \star f_m^+ \rangle \right\}. \end{aligned} \quad (13)$$

As the next step, one writes the equations of motion for the $f^{\pm}(x_{\pm})$'s and solves for them; afterwards, substitutes those expressions into the intermediate action (13) getting

$$S''[g, W^{\pm}] = I_{WZW}[g] - \frac{1}{4} \int d^2x \sum_{m=1}^2 \langle [E_{-3}, W_{3-m}^+] \star g \star [E_3, W_{3-m}^-] \star g^{-1} \rangle. \quad (14)$$

Notice that (14) has inherited from the NCATM action the local symmetries (7)-(9). Therefore, one considers the gauge fixing

$$2i\Lambda_m^- = [E_{-3}, W_{3-m}^+], \quad 2i\Lambda_m^+ = [E_3, W_{3-m}^-], \quad (15)$$

where $\Lambda^\pm \in \hat{\mathcal{G}}_{\pm 1}$ are some constant generators in the subspaces of grade ± 1 in (182)-(183).

Then for this gauge fixing the effective action (14) becomes

$$\begin{aligned} S_{NCGSG_1}[g] &\equiv S[g] \\ &= I_{WZW}[g] + \int d^2x \sum_{m=1}^2 [\langle \Lambda_m^- \star g \star \Lambda_m^+ \star g^{-1} \rangle]. \end{aligned} \quad (16)$$

Thus, we get the equation of motion for the field g as

$$\partial_-(g^{-1} \star \partial_+ g) = \sum_{m=1}^2 [\Lambda_m^-, g \star \Lambda_m^+ g^{-1}] \quad (17)$$

The action (16) for $g \in [U(1)]^3$ will define the first version of the non-commutative generalized sine-Gordon model (NCGSG₁). The second version requires a copy of the above action for the field \bar{g}

$$S_{NCGSG_2}[g] \equiv S[g] + S[\bar{g}], \quad (18)$$

where $g \in \mathcal{H} \subset SL(3)$ (\mathcal{H} will be specified below).

Thus, the actions (16) and (18) are the multi-field extensions of the NC sine-Gordon models proposed earlier by Lechtenfeld et al. and Grisaru-Penati, respectively. As we will see below, these models contain as sub-models the relevant versions of the NCSG_{1,2} model (in fact, each version contains three NCSG_{1,2} sub-models) proposed in the literature, i.e. the Lechtenfeld et al. and Grisaru-Penati proposals for the NC extension of the sine-Gordon model, respectively. Moreover, the NCGSG_{1,2} models give rise to new phenomena with interesting properties, such as the appearance of two versions of the NC Bukhvostov-Lipatov model and the NC double sine-Gordon model, respectively, as well as their NC soliton and kink type solutions.

We present below the two NCGSG_{1,2} versions related to $GL(3)$, each one involving multi-field scalar fields.

4 The Toda field g parametrizations

In this section we present the two possible parametrizations of the field g , thus obtaining the two NC versions NCGSG_{1,2} of the GSG model, and furthermore we obtain their relevant sub-models associated to them through consistent reductions.

4.1 First parametrization: $g \in [U(1)]^3 \subset GL(3, \mathbb{C})$

Let us write the field g in the representation

$$g = \begin{pmatrix} e_{\star}^{i\phi_1} & 0 & 0 \\ 0 & e_{\star}^{i\phi_2} & 0 \\ 0 & 0 & e_{\star}^{i\phi_3} \end{pmatrix} \equiv g_1 \star g_2 \star g_3, \text{ where} \quad (19)$$

$$g_1 = \begin{pmatrix} e_{\star}^{i\phi_1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad g_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e_{\star}^{i\phi_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad g_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e_{\star}^{i\phi_3} \end{pmatrix} \quad (20)$$

with ϕ_i being real fields ($i = 1, 2, 3$). As we will see below, this parametrization constitutes the $GL(3, \mathbb{C})$ extension of the Lechtenfeld et al. proposal of the non-commutative version of the sine-Gordon model (NCSG₁) [12].

For the Λ_i 's taken as

$$\begin{aligned} \Lambda_1^+ &= \Lambda_R^1 E_{\alpha_1}^0 + \Lambda_R^2 E_{\alpha}^0 + \bar{\Lambda}_R^3 E_{-\alpha_3}^1, \\ \Lambda_1^- &= \Lambda_L^3 E_{\alpha_3}^{-1} + \bar{\Lambda}_L^1 E_{-\alpha_1}^0 + \bar{\Lambda}_L^2 E_{-\alpha_2}^0, \\ \Lambda_2^+ &= \Lambda_R^3 E_{\alpha_3}^0 + \bar{\Lambda}_R^1 E_{-\alpha_1}^1 + \bar{\Lambda}_R^2 E_{-\alpha_2}^1, \\ \Lambda_2^- &= \Lambda_L^1 E_{\alpha_1}^{-1} + \Lambda_L^2 E_{\alpha_2}^{-1} + \bar{\Lambda}_L^3 E_{-\alpha_3}^0, \end{aligned} \quad (21)$$

the action (16) for g given in (19), upon using twice the Polyakov-Wiegmann identity

$$I_{WZW}(g_1 \star g_2) = I_{WZW}(g_1) + I_{WZW}(g_2) + \int dz^2 \langle g_1^{-1} \star \partial_- g_1 \star \partial_+ g_2 \star g_2^{-1} \rangle, \quad (22)$$

can be written as

$$\begin{aligned} S_{NCSG_1}[g_1, g_2, g_3] &= I_{WZW}[g_1] + I_{WZW}[g_2] + I_{WZW}[g_3] + \\ &\int d^2x \left([\Lambda_L^3 \bar{\Lambda}_R^3 e_{\star}^{i\phi_1} \star e_{\star}^{-i\phi_3} + \bar{\Lambda}_L^3 \Lambda_R^3 e_{\star}^{i\phi_3} \star e_{\star}^{-i\phi_1}] + \right. \\ &[\Lambda_L^1 \bar{\Lambda}_R^1 e_{\star}^{i\phi_1} \star e_{\star}^{-i\phi_2} + \bar{\Lambda}_L^1 \Lambda_R^1 e_{\star}^{i\phi_2} \star e_{\star}^{-i\phi_1}] + \\ &\left. [\Lambda_L^2 \bar{\Lambda}_R^2 e_{\star}^{i\phi_2} \star e_{\star}^{-i\phi_3} + \bar{\Lambda}_L^2 \Lambda_R^2 e_{\star}^{i\phi_3} \star e_{\star}^{-i\phi_2}] \right). \end{aligned} \quad (23)$$

Notice that the last term in the Polyakov-Wiegmann identity (22) vanishes when written for each pair of the fields in the parametrizations (20). Then, the relevant eqs. of motion become

$$\begin{aligned} \partial_- (e_{\star}^{-i\phi_1} \star \partial_+ e_{\star}^{i\phi_1}) &= [\Lambda_L^1 \bar{\Lambda}_R^1 e_{\star}^{i\phi_2} \star e_{\star}^{-i\phi_1} - \bar{\Lambda}_L^1 \Lambda_R^1 e_{\star}^{i\phi_1} \star e_{\star}^{-i\phi_2}] + \\ &[\Lambda_L^3 \bar{\Lambda}_R^3 e_{\star}^{i\phi_3} \star e_{\star}^{-i\phi_1} - \bar{\Lambda}_L^3 \Lambda_R^3 e_{\star}^{i\phi_1} \star e_{\star}^{-i\phi_3}] \end{aligned} \quad (24)$$

$$\begin{aligned} \partial_- (e_{\star}^{-i\phi_2} \star \partial_+ e_{\star}^{i\phi_2}) &= [\Lambda_L^2 \bar{\Lambda}_R^2 e_{\star}^{i\phi_3} \star e_{\star}^{-i\phi_2} - \bar{\Lambda}_L^2 \Lambda_R^2 e_{\star}^{i\phi_2} \star e_{\star}^{-i\phi_3}] + \\ &[\bar{\Lambda}_L^1 \Lambda_R^1 e_{\star}^{i\phi_1} \star e_{\star}^{-i\phi_2} - \Lambda_L^1 \bar{\Lambda}_R^1 e_{\star}^{i\phi_2} \star e_{\star}^{-i\phi_1}]. \end{aligned} \quad (25)$$

$$\begin{aligned} \partial_- (e_{\star}^{-i\phi_3} \star \partial_+ e_{\star}^{i\phi_3}) &= [\bar{\Lambda}_L^3 \Lambda_R^3 e_{\star}^{i\phi_1} \star e_{\star}^{-i\phi_3} - \Lambda_L^3 \bar{\Lambda}_R^3 e_{\star}^{i\phi_3} \star e_{\star}^{-i\phi_1}] + \\ &[\bar{\Lambda}_L^2 \Lambda_R^2 e_{\star}^{i\phi_2} \star e_{\star}^{-i\phi_3} - \Lambda_L^2 \bar{\Lambda}_R^2 e_{\star}^{i\phi_3} \star e_{\star}^{-i\phi_2}]. \end{aligned} \quad (26)$$

Setting

$$\Lambda_L^j \bar{\Lambda}_R^j = e^{i\delta_j} M_j / 8, \quad j = 1, 2, 3; \quad (27)$$

for M_j, δ_j some constants, we define the system of eqs. (24)-(26) as the *first version* of the non-commutative generalized $GL(3, \mathbb{C})$ sine-Gordon model (NCGSG₁). Notice that it is defined for three real scalar fields.

In the commutative limit $\theta \rightarrow 0$ the above equations can be written as

$$\partial^2 \phi_1 = M_1 \sin(\phi_2 - \phi_1 + \delta_1) + M_3 \sin(\phi_3 - \phi_1 + \delta_3); \quad (28)$$

$$\partial^2 \phi_2 = M_2 \sin(\phi_3 - \phi_2 + \delta_2) + M_1 \sin(\phi_1 - \phi_2 - \delta_1); \quad (29)$$

$$\partial^2 \phi_3 = M_2 \sin(\phi_2 - \phi_3 - \delta_2) + M_3 \sin(\phi_1 - \phi_3 - \delta_3). \quad (30)$$

From the above system of equations one gets a free scalar equation of motion

$$\partial^2 \Phi = 0, \quad \Phi \equiv \phi_1 + \phi_2 + \phi_3. \quad (31)$$

For the particular solution $\Phi \equiv 0$ of (31) and making $M_j \rightarrow -M_j, \phi_1 \rightarrow -\phi_1$, one can write the first two equations (28)-(29) as

$$\partial^2 \phi_1 = M_1 \sin(\phi_2 + \phi_1 + \delta_1) + M_3 \sin(2\phi_1 - \phi_2 + \delta_3); \quad (32)$$

$$\partial^2 \phi_2 = M_2 \sin(2\phi_2 - \phi_1 - \delta_2) + M_1 \sin(\phi_1 + \phi_2 + \delta_1). \quad (33)$$

This system of eqs. is precisely the commutative generalized sine-Gordon model (GSG) [24, 25] [the form written in (32)-(33) corresponds to eqs. (163)-(164) of Appendix A].

In the following subsections we will examine certain sub-models obtained through consistent reductions of the NCGSG₁ system (24)-(26).

4.1.1 Non-commutative sine-Gordon model (NCSG₁): Lechtenfeld et al. proposal

We show that the model (24)-(26) contains as sub-models the Lechtenfeld et al. proposal for the NCSG₁ model. So, setting $M_2 = M_3 = 0, M_1 = 8M, \phi_3 = \delta_j = 0$ and changing $\phi_2 \rightarrow -\phi_2$ we get the system of equations [12]

$$\partial_- (e_*^{-i\phi_1} \star \partial_+ e_*^{i\phi_1}) = M [e_*^{-i\phi_2} \star e_*^{-i\phi_1} - e_*^{i\phi_1} \star e_*^{i\phi_2}] \quad (34)$$

$$\partial_- (e_*^{i\phi_2} \star \partial_+ e_*^{-i\phi_2}) = M [e_*^{i\phi_1} \star e_*^{i\phi_2} - e_*^{-i\phi_2} \star e_*^{-i\phi_1}]. \quad (35)$$

In fact, there are additional two possibilities for meaningful reductions, i.e., 1) $M_1 = M_2 = 0, M_3 = 8M, \phi_2 = \delta_j = 0; \phi_3 \rightarrow -\phi_3$ and 2) $M_1 = M_3 = 0, M_2 = 8M, \phi_1 = \delta_j = 0; \phi_3 \rightarrow -\phi_3$ respectively, providing in each case a Lechtenfeld et al. NCSG₁ model.

4.1.2 Non-commutative (bosonized) Bukhvostov-Lipatov model (NCbBL₁). First version

Another reduction is possible by making $M_1 = 0$, $M_2 = M_3 = -M$, and $\phi_1 \rightarrow -\phi_1$, in the eqs. (24)-(26) followed by the substitution $\phi_3 = \phi_1 - \phi_2$. So, one gets the set of equations

$$\partial_- \left(e_{\star}^{\pm i\phi_a} \star \partial_+ e_{\star}^{\mp i\phi_a} \right) = -\frac{M}{8} \left[e_{\star}^{i(\phi_1 - \phi_2)} \star e_{\star}^{\pm i\phi_a} - e_{\star}^{\mp i\phi_a} \star e_{\star}^{-i(\phi_1 - \phi_2)} \right], \quad a = 1, 2 \quad (36)$$

$$0 = \partial_- \left[e_{\star}^{i\phi_1} \star \partial_+ e_{\star}^{-i\phi_1} + e_{\star}^{-i\phi_2} \star \partial_+ e_{\star}^{i\phi_2} + e_{\star}^{-i(\phi_1 - \phi_2)} \star \partial_+ e_{\star}^{i(\phi_1 - \phi_2)} \right], \quad (37)$$

where the upper (lower) signs in (36) correspond to the index $a = 1(2)$ for the field ϕ_a . In the commutative limit the eq. (37) becomes trivial, whereas the set of equations (36) become $\partial^2 \phi_1 = M \sin(2\phi_1 - \phi_2)$ and $\partial^2 \phi_2 = M \sin(2\phi_2 - \phi_1)$. Defining the new fields $\psi_1 = \frac{1}{2}(\phi_1 + \phi_2)$, $\psi_2 = \frac{\sqrt{3}}{2}(\phi_1 - \phi_2)$ we arrive at the model $\partial^2 \psi_1 = M \sin(\psi_1) \cos(\sqrt{3}\psi_2)$, $\partial^2 \psi_2 = M \sqrt{3} \cos(\psi_1) \sin(\sqrt{3}\psi_2)$. This system of equations is precisely the bosonized form of the so-called Bukhvostov-Lipatov model [27, 28, 29, 20]. In view of these relationships we define the model (36)-(37) as the *first version* of the non-commutative bosonized Bukhvostov-Lipatov model (NCbBL₁).

4.1.3 Non-commutative double sine-Gordon model (NCDSG₁). First version

The usual double sine-Gordon model (DSG) is defined in terms of just one scalar field ϕ and the potential terms $[\cos(\phi) + \cos(2\phi)]$ in the action. So, we would like to reduce the above model in a consistent way in order to get a sub-model defined for just one scalar field. Let us take advantage of a particular solution of the free field equation (31). So, we consider the reduction $\phi_1 = -\phi_3 = \phi$, $\phi_2 = 0$ and substitute these relations into the equations (24)-(26). Then we obtain the next two equations

$$\partial_- \left(e_{\star}^{-i\phi} \star \partial_+ e_{\star}^{i\phi} \right) = M_1 (e_{\star}^{-i\phi} - e_{\star}^{i\phi}) + M_3 (e_{\star}^{-i\phi} \star e_{\star}^{-i\phi} - e_{\star}^{i\phi} \star e_{\star}^{i\phi}), \quad (38)$$

$$\partial_- \left(e_{\star}^{i\phi} \star \partial_+ e_{\star}^{-i\phi} \right) = M_1 (e_{\star}^{i\phi} - e_{\star}^{-i\phi}) + M_3 (e_{\star}^{i\phi} \star e_{\star}^{i\phi} - e_{\star}^{-i\phi} \star e_{\star}^{-i\phi}) \quad (39)$$

plus an equation which reduces to a trivial identity (we have imposed $M_1 = M_2$, $\delta_i = 0$).

The above two equations can be written in the equivalent form

$$\partial_- \left(e_{\star}^{i\phi} \star \partial_+ e_{\star}^{-i\phi} - e_{\star}^{-i\phi} \star \partial_+ e_{\star}^{i\phi} \right) = 4iM_1 \sin_{\star} \phi + 4iM_3 \sin_{\star} 2\phi \quad (40)$$

$$\partial_- \left(e_{\star}^{-i\phi} \star \partial_+ e_{\star}^{i\phi} + e_{\star}^{i\phi} \star \partial_+ e_{\star}^{-i\phi} \right) = 0. \quad (41)$$

The system (40)-(41) constitutes the *first version* of the non-commutative double sine-Gordon model (NCDSG₁) defined for just one scalar field.

The first equation (40) contains the potential terms which is the natural generalization of the ordinary double sine-Gordon potential, whereas the other one (41) has the structure of a conservation law and it can be seen as imposing an extra condition on the system. In the commutative limit, the first equation reduces to the ordinary double sine-Gordon equation (DSG), whereas the second one becomes trivial. The equations

are in general complex and possess the \mathbb{Z}_2 symmetry of the ordinary DSG (the invariance under $\phi \rightarrow -\phi$ is easily seen in (38)-(39).

4.2 Second parametrization: $g \in \mathcal{H} \subset GL(3, \mathbb{C})$

Let us consider the parametrization

$$g = \begin{pmatrix} e_*^{\varphi_1} \star e_*^{\varphi_0} & 0 & 0 \\ 0 & e_*^{-\varphi_1+\varphi_2} \star e_*^{\varphi_0} & 0 \\ 0 & 0 & e_*^{-\varphi_2} \star e_*^{\varphi_0} \end{pmatrix} \equiv g_1 \star g_2, \quad (42)$$

with

$$g_1 = \begin{pmatrix} e_*^{\varphi_1} & 0 & 0 \\ 0 & e_*^{-\varphi_1+\varphi_2} & 0 \\ 0 & 0 & e_*^{-\varphi_2} \end{pmatrix}, \quad g_2 = e_*^{\varphi_0} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (43)$$

where the fields φ_j , $j = 0, 1, 2$ are general complex fields. The additional field \bar{g} is defined by substituting the fields φ_j above as φ_j^\dagger . The fields g and \bar{g} are formally considered to be independent fields.

This parametrization becomes the $GL(3)$ extension of the Grisaru-Penati proposal for the non-commutative version of the sine-Gordon model (NCSG₂) [7, 8].

The following equations of motion can be obtained directly from the first term $S[g]$ of the action (18) for the parametrization (42)

$$\begin{aligned} \partial_- \left(e_*^{-\varphi_0} \star e_*^{-\varphi_1} \star \partial_+ \left(e_*^{\varphi_1} \star e_*^{\varphi_0} \right) \right) &= [\Lambda_L^1 \bar{\Lambda}_R^1 e_*^{-\varphi_1+\varphi_2} \star e_*^{-\varphi_1} - \bar{\Lambda}_L^1 \Lambda_R^1 e_*^{\varphi_1} \star e_*^{\varphi_1-\varphi_2}] + \\ &[\Lambda_L^3 \bar{\Lambda}_R^3 e_*^{-\varphi_2} \star e_*^{-\varphi_1} - \bar{\Lambda}_L^3 \Lambda_R^3 e_*^{\varphi_1} \star e_*^{\varphi_2}]. \end{aligned} \quad (44)$$

$$\begin{aligned} \partial_- \left(e_*^{-\varphi_0} \star e_*^{\varphi_1-\varphi_2} \star \partial_+ \left(e_*^{-\varphi_1+\varphi_2} \star e_*^{\varphi_0} \right) \right) &= [\Lambda_L^2 \bar{\Lambda}_R^2 e_*^{-\varphi_2} \star e_*^{\varphi_1-\varphi_2} - \bar{\Lambda}_L^2 \Lambda_R^2 e_*^{-\varphi_1+\varphi_2} \star e_*^{\varphi_2}] + \\ &[\bar{\Lambda}_L^1 \Lambda_R^1 e_*^{\varphi_1} \star e_*^{\varphi_1-\varphi_2} - \Lambda_L^1 \bar{\Lambda}_R^1 e_*^{-\varphi_1+\varphi_2} \star e_*^{-\varphi_1}] \end{aligned} \quad (45)$$

$$\begin{aligned} \partial_- \left(e_*^{-\varphi_0} \star e_*^{\varphi_2} \star \partial_+ \left(e_*^{-\varphi_2} \star e_*^{\varphi_0} \right) \right) &= [\bar{\Lambda}_L^3 \Lambda_R^3 e_*^{\varphi_1} \star e_*^{\varphi_2} - \Lambda_L^3 \bar{\Lambda}_R^3 e_*^{-\varphi_2} \star e_*^{-\varphi_1}] + \\ &[\bar{\Lambda}_L^2 \Lambda_R^2 e_*^{-\varphi_1+\varphi_2} \star e_*^{\varphi_2} - \Lambda_L^2 \bar{\Lambda}_R^2 e_*^{-\varphi_2} \star e_*^{\varphi_1-\varphi_2}]. \end{aligned} \quad (46)$$

Introduce the parameters M_i, δ_i as in (27). So, we define the system of eqs. (44)-(46), supplied with the relevant eqs. of motion for the fields φ_j^\dagger derived from the second term $S[\bar{g}]$ of the action (18), as the *second version* of the non-commutative generalized $GL(3, \mathbb{C})$ sine-Gordon model (NCGSG₂), where the three scalar fields φ_j are in general complex.

Next, let us examine the commutative limit. Redefining $\varphi_a \rightarrow i \varphi_a$ (where the new φ_a 's are real), using definition (27) and taking the limit $\theta \rightarrow 0$ in the above system of equations (44)-(46) one can get

$$\partial^2 \varphi_1 = M_1 \sin(2\varphi_1 - \varphi_2 - \delta_1) + M_3 \sin(\varphi_1 + \varphi_2 - \delta_3) \quad (47)$$

$$\partial^2 \varphi_2 = M_2 \sin(2\varphi_2 - \varphi_1 - \delta_2) + M_3 \sin(\varphi_1 + \varphi_2 - \delta_3) \quad (48)$$

$$\partial^2 \varphi_0 = 0. \quad (49)$$

Thus, in (47)-(48) we recover again the equations of motion of the commutative generalized sine-Gordon model (GSG) [24, 25]. Notice that the field φ_0 decouples completely from the other fields in this limit, becoming simply a free field.

In analogy to the results of the first parametrization it is possible to get some sub-models as consistent reductions of the system (44)-(46). In the following we discuss the reductions associated to this second parametrization.

4.2.1 Non-commutative sine-Gordon model (NCSG₂): Grisaru-Penati proposal

A reduced single field model follows by setting $M_2 = M_3 = \delta_i = 0$, $\varphi_0 = \varphi_2 = 0$, $M_1 = -M$ and $\varphi_1 = i\varphi$ (φ , complex field). So, one gets the model

$$\partial_-(e^{\mp i\varphi} \partial_+ e^{\pm i\varphi}) = \pm M(e^{\pm 2i\varphi} - e^{\mp 2i\varphi}), \quad (50)$$

which is the Grisaru-Penati proposal for the NC extension of the sine-Gordon model (NCSG₂) [7, 8]. In fact, in this proposal one must consider additionally a couple of equations for φ^\dagger obtained from the second piece in the action (18). Additional reductions, each one providing a Grisaru-Penati NCSG₂ model, are achieved by setting $M_1 = M_3 = \delta_i = 0$, $\varphi_0 = \varphi_1 = 0$, $M_2 = -M$, $\varphi_2 = i\varphi$, and $M_1 = M_2 = \delta_i = 0$, $\varphi_0 = 0$, $M_3 = -M$, $\varphi_1 = \varphi_2 = i\varphi$, respectively.

4.2.2 Non-commutative (bosonized) Bukhvestov-Lipatov model (NCbBL₂). Second version

A reduction leading to a two field model follows as $M_3 = 0$, $M_1 = M_2 = -M$, $\varphi_n \rightarrow i\varphi_n$ ($n = 0, 1, 2$). So, one gets the model

$$\partial_- \left(e_{\star}^{-i\varphi_0} \star e_{\star}^{\mp i\varphi_a} \star \partial_+ (e_{\star}^{\pm i\varphi_a} \star e_{\star}^{i\varphi_0}) \right) = -\frac{M}{8} \left[e_{\star}^{-i(\varphi_1 - \varphi_2)} \star e_{\star}^{\mp i\varphi_a} - e_{\star}^{\pm i\varphi_a} \star e_{\star}^{i(\varphi_1 - \varphi_2)} \right], \quad (51)$$

$a = 1, 2$

$$0 = \partial_- \left[e_{\star}^{-i\varphi_0} \star e_{\star}^{-i\varphi_1} \star \partial_+ (e_{\star}^{i\varphi_1} \star e_{\star}^{i\varphi_0}) + e_{\star}^{-i\varphi_0} \star e_{\star}^{i\varphi_2} \star \partial_+ (e_{\star}^{-i\varphi_2} \star e_{\star}^{i\varphi_0}) + e_{\star}^{-i\varphi_0} \star e_{\star}^{i(\varphi_1 - \varphi_2)} \star \partial_+ (e_{\star}^{-i(\varphi_1 - \varphi_2)} \star e_{\star}^{i\varphi_0}) \right], \quad (52)$$

where the upper (lower) signs in (51) correspond to the index $a = 1(2)$ of the field φ_a . In the commutative limit the eq. (52) reduces to a free scalar field equation of motion $\partial^2 \varphi_0 = 0$, whereas the set of equations (51) become $\partial^2 \varphi_1 = M \sin(2\varphi_1 - \varphi_2)$ and $\partial^2 \varphi_2 = M \sin(2\varphi_2 - \varphi_1)$. Defining the new fields $\psi_1 = \frac{1}{2}(\varphi_1 + \varphi_2)$, $\psi_2 = \frac{\sqrt{3}}{2}(\varphi_1 - \varphi_2)$ we arrive at the model $\partial^2 \psi_1 = M \sin(\psi_1) \cos(\sqrt{3}\psi_2)$, $\partial^2 \psi_2 = M \sqrt{3} \cos(\psi_1) \sin(\sqrt{3}\psi_2)$. As we have seen before this is just the bosonized form of the so-called Bukhvestov-Lipatov model [27, 28, 29, 20].

In view of these relationships we define the model (51)-(52) as the *second version* of the non-commutative (bosonized) Bukhvostov-Lipatov model (NCbBL₂).

4.2.3 Non-commutative double sine-Gordon model (NCDSG₂). Second version

In order to reduce the NCGSG₂ system of equations into another version of the NC double sine-Gordon model one takes advantage of certain properties of its commutative counterpart. In fact, the above commutative model (47)-(49) possesses the symmetry $\varphi_1 \leftrightarrow \varphi_2$; $M_1 \leftrightarrow M_2$ in the GSG sector, whereas the auxiliary φ_0 field completely decouples in this limit. So, in the second parametrization case (42) we can impose the conditions $\varphi_1 = \varphi_2 \equiv -i\varphi$, $M_1 = M_2$, $\delta_i = 0$ into the system of eqs. (44)-(46) and obtain the following system of equations for complex φ

$$\partial_- \left(e_*^{-\varphi_0} \star e_*^{i\varphi} \star \partial_+ \left(e_*^{-i\varphi} \star e_*^{\varphi_0} \right) \right) = 2iM_1 \sin_* \varphi + 2iM_3 \sin_* 2\varphi \quad (53)$$

$$\partial_- \left(e_*^{-\varphi_0} \star \partial_+ e_*^{\varphi_0} \right) = 0 \quad (54)$$

$$\begin{aligned} \partial_- \left[e_*^{-\varphi_0} \star e_*^{i\varphi} \star \partial_+ \left(e_*^{-i\varphi} \star e_*^{\varphi_0} \right) + e_*^{-\varphi_0} \star e_*^{-i\varphi} \star \right. \\ \left. \partial_+ \left(e_*^{i\varphi} \star e_*^{\varphi_0} \right) \right] = 0. \end{aligned} \quad (55)$$

The system (53)-(55) constitutes the *second version* of the non-commutative double sine-Gordon model (NCDSG₂) defined for two complex scalar fields.

The first equation (53) contains the potential terms generalizing the ordinary double sine-Gordon potential. The second and third ones (54)-(55) have the structure of conservation laws and can be seen as imposing extra conditions on the system. Let us examine the commutative limit $\theta \rightarrow 0$ of the NCDSG₂ system. In this limit it reduces to the usual DSG model plus a free field φ_0 equations of motion

$$\partial_- \partial_+ \varphi = -2M_1 \sin \varphi - 2M_3 \sin 2\varphi \quad (56)$$

$$\partial_- \partial_+ \varphi_0 = 0. \quad (57)$$

Notice that in this limit the field φ_0 decouples completely from the DSG field φ .

Some comments are in order here.

1) The NC models obtained above reproduce the usual models in the commutative limit $\theta \rightarrow 0$. So, the both versions of the $GL(3, \mathbb{C})$ non-commutative generalized sine-Gordon model (NCGSG_{1,2}) reproduce the ordinary $GL(3)$ GSG model in this limit. The both versions of the non-commutative double sine-Gordon model NCDSG_{1,2} reproduce the usual DSG model in the ordinary space. Likewise, the both versions of the non-commutative bosonized Bukhvostov-Lipatov model NCbBL_{1,2} lead to the usual BL model. Notice that the GSG model in ordinary space-time also contains as sub-models the variety of theories we have uncovered above, i.e. the usual SG model, Bukhvostov-Lipatov model, and the double sine-Gordon model [20, 24].

2) Regarding the integrability of the $\text{NCGSG}_{1,2}$ models they are hardly expected to possess this property since they contain as sub-models the relevant $\text{NCDSG}_{1,2}$ and $\text{NCBL}_{1,2}$ theories. The $\text{NCDSG}_{1,2}$ models are not expected to possess this property since their commutative counterpart is not integrable. The same behavior may be expected for the $\text{NCBL}_{1,2}$ models since their commutative counterpart is not classically integrable (see [20] and refs. therein), except for some restricted region in parameters space. Nevertheless, see more on this point in subsection 6.1.2 when the relevant spinor version of the (constrained) NCBL_1 model is discussed in relation to integrability.

Related to this issue, let us mention that we have not been able to write in a zero-curvature form the eq. of motion (17) of the NCGSG_1 model (17), it mainly happens due to the presence of the summation index $m = 1, 2$ on both entries of the commutator. Actually, the eq. (17) differs from the integrable system of non-abelian affine Toda equations [30, 18].

3) The Letchenfeld et al. (34)-(35) and Grisaru-Penati (50) NC sine-Gordon models proposed in the literature appear in the context of the generalized NC sine-Gordon models as reduced sub-models of the corresponding NCGSG_1 and NCGSG_2 models, respectively. So, they are analogous to the results obtained in the commutative case in which the $GL(3)$ GSG model contains three SG sub-models as reduced models, each one associated to the positive root of the $gl(3)$ Lie algebra [24]. The group structure of the $GL(3)$ $\text{NCGSG}_{1,2}$ models allowed us to get three $\text{NCSG}_{1,2}$ sub-models, respectively, for each version, as in the commutative case.

4) In the three-field space of the $\text{NCGSG}_{1,2}$ models it is remarkable the appearance of three integrable directions as $\text{NCSG}_{1,2}$ sub-models, respectively. It suggests that there are at least three integrable directions in reduced field space of each one of the $\text{NCGSG}_{1,2}$ models. Examples of non-integrable reduced directions are provided by the relevant $\text{NCDSG}_{1,2}$ and $\text{NCBL}_{1,2}$ models. However, the existence of more integrable directions is suggested by the presence of certain integrable sub-models in the spinor sector of the $\text{NCGMT}_{1,2}$ models, i.e. the scalar duals of the corresponding $\text{NC(c)GMT}_{1,2}$ and $\text{NC(c)BL}_{1,2}$ spinor models, respectively (see section (6.1.1) and subsection (6.1.2)).

5) Finally, the role played by the SG model in the context of the generalized SG models is analogous to the one which happens with the correspondence between the $\lambda\phi^4$ model and the deformed linear $O(N)$ -sigma model, as it was first noticed in [31]. It could be interesting to study several properties of the generalized SG models, including their non-commutative counterparts, as for example by applying and improving the quantization method described in the last reference. Let us mention that the ordinary DSG model has been in the center of some controversy regarding the computation of its semi-classical spectrum [32, 33], however it seems that the point has recently been clarified in [34].

5 Decoupling of NCGSG_{1,2} and NCGMT_{1,2} models

In the commutative case some approaches have been proposed in order to recover the GSG and GMT dual models out of the ordinary $sl(n)$ ATM model [24, 13, 14, 15, 20]. Among them, the one which proceeds by decoupling the set of equations of motion of the ATM model into the corresponding dual models [13, 15] has turned out to be more suitable in the NC case [9]. This procedure is adapted to the NC case by writing a set of mappings between the fields of the model such that the eqs. (3) and (6) when rewritten using those mappings decouple the scalar and the matter fields. So, following the procedures employed in the ordinary $sl(n)$ case [15] and in the non-commutative $GL(2)$ ATM case [9] to the case at hand, let us consider the mappings

$$\sum_{n=1}^2 [F_n^-, g F_n^+ g^{-1}]_* = \sum_{n=1}^2 [\Lambda_n^-, g \Lambda_n^+ g^{-1}]_*, \quad (58)$$

$$[E_{-3}, g F_{3-m}^+ g^{-1}]_* = [E_{-3}, F_{3-m}^+]_* - \frac{k_2}{2} (L_m^+)^{-1} \left[\sum_n \hat{J}_n^-, \hat{F}_m^- \right]_* L_m^+, \quad (59)$$

$$[E_3, g^{-1} F_{3-m}^- g]_* = [E_3, F_{3-m}^-]_* - \frac{k_1}{2} (L_m^-)^{-1} \left[\sum_n \hat{J}_n^+, \hat{F}_m^+ \right]_* L_m^-, \quad (60)$$

$$F_m^\pm = \mp [E_{\pm 3}, W_{3-m}^\mp]_*, \quad (61)$$

$$[F_2^\pm, g^{\mp 1} F_1^\mp g^{\pm 1}]_* = 0, \quad (62)$$

where

$$\hat{J}_n^\mp \equiv [\hat{F}_{3-n}^\pm, \hat{W}_{3-n}^\mp]; \quad k_1, k_2 = \text{constant parameters.} \quad (63)$$

The hatted fields have the same algebraic structure as the corresponding unhatted ones except that they incorporate some parameters re-scaling the fields, those parameters will give rise to certain coupling constants between the currents of the model. Notice that the fields \hat{J}_m^\pm and the constant matrices L_m^\pm carry zero gradation and these will be defined below. The field g in the relations above, as defined in section 2, is assumed to belong to either $[U(1)]^3$, as in subsection 4.1, or $\mathcal{H} \subset GL(3, \mathbb{C})$, as in the second parametrization in subsection 4.2.

The relationships (58)-(61) when conveniently substituted into the ATM eqs. of motion (3) and (6) decouple them, respectively, into the NCGSG₁ eq. (17) and certain equations of motion incorporating only matter fields, which in matrix form become

$$[E_{-3}, \partial_+ W_{3-m}^+]_* = +[E_{-3}, [E_3, W_m^-]]_* - \frac{k_2}{2} \sum_{n=1}^2 (L_m^+)^{-1} [\hat{J}_n^-, [E_{-3}, \hat{W}_{3-m}^+]]_* L_m^+ \quad (64)$$

$$[E_3, \partial_- W_{3-m}^-]_* = -[E_3, [E_{-3}, W_m^+]]_* - \frac{k_1}{2} \sum_{n=1}^2 (L_m^-)^{-1} [\hat{J}_n^+, [E_3, \hat{W}_{3-m}^-]]_* L_m^- \quad (65)$$

We define these set of eqs. as the *first version* of the non-commutative (generalized) massive Thirring model (NCGMT₁).

The eqs. (62) are the constraints imposed in ref. [15] written in a compact form. These constraints, which are missing in the $GL(2)$ case, have been imposed in the non-trivial $GL(3)$ extension in order to be able to write a local Lagrangian for the off-critical and constrained ATM model out of the full set of equations of motion of the so-called conformal affine Toda model coupled to matter (CATM) [15, 18] (see the Appendices). Actually, the above 'decoupling' eqs. maintain the same form as their commutative analogs presented in eqs. (6.1)-(6.5) of the ref. [15]. We must clarify that the above 'decoupling' eqs. (58)-(60) do not completely decouple the scalar fields from the spinor-like fields due to the presence of the constraints (62). There are some instances of total decoupling, e.g. in the soliton sector of the commutative limit [20, 15]. Notice that we have not used the constraint equations (62) in order to get the eqs. (64)-(65). In order to be more specific in the discussions below we provide, in the following set of equations, the constraint eqs. (62) in terms of the component fields. Let us take the spinors as defined in (168)-(173) and the scalar field g presented in the first parametrization eq. (19), so one has

$$e_*^{i\phi_1} \star \psi_R^1 \star e_*^{-i\phi_2} \star \psi_L^2 = \psi_L^1 \star e_*^{i\phi_2} \star \psi_R^2 \star e_*^{-i\phi_3}, \quad (66)$$

$$e_*^{i\phi_2} \star \psi_R^2 \star e_*^{-i\phi_3} \star \tilde{\psi}_L^3 = -\psi_L^2 \star e_*^{i\phi_3} \star \tilde{\psi}_R^3 \star e_*^{-i\phi_1}, \quad (67)$$

$$e_*^{i\phi_3} \star \tilde{\psi}_R^3 \star e_*^{-i\phi_1} \star \psi_L^1 = -\tilde{\psi}_L^3 \star e_*^{i\phi_1} \star \psi_R^1 \star e_*^{-i\phi_2} \quad (68)$$

and

$$e_*^{-i\phi_3} \star \tilde{\psi}_L^2 \star e_*^{i\phi_2} \star \tilde{\psi}_R^1 = \tilde{\psi}_R^2 \star e_*^{-i\phi_2} \star \tilde{\psi}_L^1 \star e_*^{i\phi_1}, \quad (69)$$

$$e_*^{-i\phi_1} \star \psi_L^3 \star e_*^{i\phi_3} \star \tilde{\psi}_R^2 = -\psi_R^3 \star e_*^{-i\phi_3} \star \tilde{\psi}_L^2 \star e_*^{i\phi_2}, \quad (70)$$

$$e_*^{-i\phi_2} \star \tilde{\psi}_L^1 \star e_*^{i\phi_1} \star \psi_R^3 = -\tilde{\psi}_R^1 \star e_*^{-i\phi_1} \star \psi_L^3 \star e_*^{i\phi_3}, \quad (71)$$

associated to the grades (-1) and $(+1)$ of (62), respectively.

Analogously, one can write another set of equations for the second parametrization (42) of g

$$e_*^{\varphi_1} \star e_*^{\varphi_0} \star \psi_R^1 \star e_*^{-\varphi_0} \star e_*^{\varphi_1-\varphi_2} \star \psi_L^2 = \psi_L^1 \star e_*^{\varphi_2-\varphi_1} \star e_*^{\varphi_0} \star \psi_R^2 \star e_*^{-\varphi_0} \star e_*^{\varphi_2}, \quad (72)$$

$$e_*^{\varphi_2-\varphi_1} \star e_*^{\varphi_0} \star \psi_R^2 \star e_*^{-\varphi_0} \star e_*^{\varphi_2} \star \psi_L^3 = -\psi_L^2 \star e_*^{-\varphi_2} \star e_*^{\varphi_0} \star \tilde{\psi}_R^3 \star e_*^{-\varphi_0} \star e_*^{-\varphi_1}, \quad (73)$$

$$e_*^{-\varphi_2} \star e_*^{\varphi_0} \star \tilde{\psi}_R^3 \star e_*^{-\varphi_0} \star e_*^{-\varphi_1} \star \psi_L^1 = -\tilde{\psi}_L^3 \star e_*^{\varphi_1} \star e_*^{\varphi_0} \star \psi_R^1 \star e_*^{-\varphi_0} \star e_*^{\varphi_1-\varphi_2}, \quad (74)$$

and

$$e_*^{\varphi_0} \star \tilde{\psi}_R^2 \star e_*^{-\varphi_0} \star e_*^{\varphi_1-\varphi_2} \star \tilde{\psi}_L^1 \star e_*^{\varphi_1} = e_*^{\varphi_2} \star \tilde{\psi}_L^2 \star e_*^{\varphi_2-\varphi_1} \star e_*^{\varphi_0} \star \tilde{\psi}_R^1 \star e_*^{-\varphi_0}, \quad (75)$$

$$e_*^{\varphi_0} \star \psi_R^3 \star e_*^{-\varphi_0} \star e_*^{\varphi_2} \star \tilde{\psi}_L^2 \star e_*^{\varphi_2-\varphi_1} = -e_*^{-\varphi_1} \star \psi_L^3 \star e_*^{-\varphi_2} \star e_*^{\varphi_0} \star \tilde{\psi}_R^2 \star e_*^{-\varphi_0}, \quad (76)$$

$$e_*^{\varphi_0} \star \tilde{\psi}_R^1 \star e_*^{-\varphi_0} \star e_*^{\varphi_1} \star \psi_L^3 \star e_*^{-\varphi_2} = -e_*^{\varphi_1-\varphi_2} \star \tilde{\psi}_L^1 \star e_*^{\varphi_1} \star e_*^{\varphi_0} \star \psi_R^3 \star e_*^{-\varphi_0}, \quad (77)$$

associated to the grades (-1) and $(+1)$ of (62), respectively.

Even though that the full set of the 'decoupling' equations have not been used in order to write the eqs. (64)-(65), we expect that a non-commutative version of the usual (generalized) massive Thirring model (GMT₁) [15] defined for the fields W^\pm will emerge from these equations. In fact, we assume this point of view and study the properties of the system (64)-(65) in its own right. Nevertheless, we will recognize below certain relationships between the relevant sub-models of the both NCGSG_{1,2} and NCGMT_{1,2} sectors. Remarkably, these relationships will arise for certain reduced sectors obtained such that the constraints (62) become trivial, or completely decouple the spinors from the scalars in the soliton sector, which is equivalent to take the commutative limit (see below). The model (NCGMT₁) (64)-(65) is new in the literature and it is expected to correspond to the weak coupling sector of the NCATM₁ model whose strong coupling sector is described by the first version of the non-commutative generalized sine-Gordon model (NCGSG₁) presented in subsection 4.1.

In the ordinary space the GMT equations of motion can be achieved through Hamiltonian reduction procedures, such as the Faddeev-Jackiw method, as employed in [15] for first order in time Lagrangian; however, in the NC case, to our knowledge, there is no a similar procedure since the action of the NC GATM model involves higher order in time derivatives; actually, an infinite number of terms of increasing order in time derivatives. So, we have used the decoupling method and assumed the forms of the decoupling equations (58)-(63) to resemble the ones in the ordinary case [15], important guiding lines being the gradation structure and further, the locality of the Lagrangian in the NCGMT sector which will depend on the nature of the terms appearing in the eqs. of motion; e.g, notice the absence of terms bilinear in the spinors in the right hand side of the eqs. (64)-(65). In fact, the terms appearing in the above equations will give rise to usual kinetic and mass terms, and four-spinor coupling terms in the relevant action. The Lagrangian for the model (64)-(65) and a Lax pair formulation for a constrained version of it will be discussed below.

In order to recover the dual of the second version NCGSG₂ one must write similar decoupling expressions for the full set of fields $\{g, F^\pm, W^\pm\}$ and $\{\bar{g}, \mathcal{F}^\pm, \mathcal{W}^\pm\}$. Thus, following similar steps to the previous construction we expect to recover another version of the NC generalized massive Thirring model (NCGMT₂) defined for the fields $\{W^\pm, \mathcal{W}^\pm\}$. In the next section we propose two versions of the non-commutative (generalized) massive Thirring theories (NCGMT_{1,2}) by providing the relevant equations of motion and discussing their zero-curvature formulations.

6 The NC generalized massive Thirring models NCGMT_{1,2}

We will consider the fields $\psi^j, \tilde{\psi}^j$ as c-number ones [9] in order to define the NC generalization of the so-called (c-number) massive Thirring model (MT) [35, 36]. In ordinary space-time these type of classical c-number multi-field massive Thirring theories have long been considered in relation to one-dimensional Dirac model of extended particles [37]. The quantization of the two-dimensional fermion model with Thirring interaction among N different massive Fermi field species has recently been performed in the functional integral approach

[38].

The assumption for the fields to be c-number fields will allow the zero-curvature formulations of the NCGMT_{1,2} models to be constructed resembling analogous algebraic structures present in the GATM model in the context of the affine Lie algebra $SL(3)$. This means that the c-number fields $\psi^j, \tilde{\psi}^j$ will lie in certain higher grading directions of the principal gradation of the affine $SL(3)$ Lie algebra, as it is presented in the eqs. (170)-(173) of the Appendix B.

In ordinary space the field components of the MT model are considered to be either anti-commuting Grassmannian fields or some ordinary commuting fields (see [9] and refs. therein). Notice that the relevant (Grassmannian) GMT model would need a slightly different algebraic formulation from the one followed here for the c-number case.

6.1 NCGMT₁

We propose the NCGMT₁ action related to the eqs. of motion (64)-(65) for the fields W_m^\pm as

$$S[W_m^\pm] = \int dx^2 \left[\sum_{m=1}^2 \left\{ \frac{1}{2} < [E_{-3}, W_{3-m}^+] \star \partial_+ W_m^+ > - \frac{1}{2} < [E_3, W_{3-m}^-] \star \partial_- W_m^- > - \right. \right. \\ \left. \left. < [E_{-3}, W_m^+] \star [E_3, W_m^-] > \right\} - \frac{1}{2} \sum_{m,n=1}^2 < \hat{J}_m^+ \star \hat{J}_n^- > \right]. \quad (78)$$

In the last action the first two terms inside the summation provide the kinetic terms, the third one the mass terms and the last term the current-current interactions. The current-like matrices \hat{J}_m^\pm with zero gradation appearing in the eq. (63) have the same algebraic structure as the matrix-valued currents [15]

$$J_m^\pm = \pm \frac{1}{4} [[E_{\mp 3}, W_m^\pm], W_{3-m}^\pm]_\star, \quad (79)$$

except that they are defined in terms of some hatted variables \hat{W}_m^\pm which are constructed from the relevant unhatted ones W_m^\pm in eqs. (170)-(173) by making the re-scalings

$$\tilde{\psi}_L^1 \rightarrow \left(\frac{\lambda_1}{2}\right)^{1/4} \tilde{\psi}_L^1, \quad \tilde{\psi}_L^2 \rightarrow \left(\frac{\lambda_2}{2}\right)^{1/4} \tilde{\psi}_L^2, \quad \psi_L^3 \rightarrow \left(\frac{\lambda_3}{2}\right)^{1/4} \psi_L^3. \quad (80)$$

$$\frac{1}{L} \rightarrow \left(\frac{\delta_1}{2}\right)^{1/4} \psi_L^1, \quad \psi_L^2 \rightarrow \left(\frac{\delta_2}{2}\right)^{1/4} \psi_L^2, \quad \tilde{\psi}_L^3 \rightarrow \left(\frac{\delta_3}{2}\right)^{1/4} \tilde{\psi}_L^3. \quad (81)$$

$$\tilde{\psi}_R^1 \rightarrow \left(\frac{\alpha_1}{2}\right)^{1/4} \tilde{\psi}_R^1, \quad \tilde{\psi}_R^2 \rightarrow \left(\frac{\alpha_2}{2}\right)^{1/4} \tilde{\psi}_R^2, \quad \psi_R^3 \rightarrow \left(\frac{\alpha_3}{2}\right)^{1/4} \psi_R^3. \quad (82)$$

$$\frac{1}{R} \rightarrow \left(\frac{\beta_1}{2}\right)^{1/4} \psi_R^1, \quad \psi_R^2 \rightarrow \left(\frac{\beta_2}{2}\right)^{1/4} \psi_R^2, \quad \tilde{\psi}_R^3 \rightarrow \left(\frac{\beta_3}{2}\right)^{1/4} \tilde{\psi}_R^3, \quad (83)$$

where the $\lambda_j, \delta_j, \alpha_j, \beta_j$ are constant parameters. These constants are introduced with the aim of recovering some coupling constants between the currents of the model.

Actually, in matrix form we have the following relationships $\hat{W}_m^+ = L_m^+ W_m^+ (L_m^+)^{-1}$ and $\hat{W}_m^- = L_m^- W_m^- (L_m^-)^{-1}$.

The L_2^\pm , L_1^\mp matrices, respectively, take the following forms

$$\begin{pmatrix} \sqrt[12]{\frac{x_3}{x_1}} & 0 & 0 \\ 0 & \sqrt[12]{\frac{x_1}{x_2}} & 0 \\ 0 & 0 & \sqrt[12]{\frac{x_2}{x_3}} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \sqrt[12]{\frac{y_1}{y_3}} & 0 & 0 \\ 0 & \sqrt[12]{\frac{y_2}{y_1}} & 0 \\ 0 & 0 & \sqrt[12]{\frac{y_3}{y_2}} \end{pmatrix} \quad (84)$$

supplied with the replacements $x \rightarrow \lambda$ for L_2^+ , $y \rightarrow \beta$ for L_2^- , $x \rightarrow \alpha$ for L_1^- , and $y \rightarrow \delta$ for L_1^+ .

Some relationships between these parameters will emerge below mainly arising from the consideration of current-current (generalized Thirring) type interactions among the various flavor species and integrability requirement through the zero-curvature formulation of the equations of motion.

In the following we will consider the eqs. of motion (64)-(65) in term of the field components. For future convenience let us introduce the fields $A_{R,L}^i$ as

$$A_R^1 = \sqrt[4]{\frac{\alpha_1\beta_1}{4}} \psi_R^1 \star \tilde{\psi}_R^1 + \sqrt[4]{\frac{\beta_3\alpha_3}{4}} \psi_R^3 \star \tilde{\psi}_R^3 \quad (85)$$

$$A_R^2 = \sqrt[4]{\frac{\alpha_2\beta_2}{4}} \psi_R^2 \star \tilde{\psi}_R^2 - \sqrt[4]{\frac{\alpha_1\beta_1}{4}} \tilde{\psi}_R^1 \star \psi_R^1 \quad (86)$$

$$A_R^3 = \sqrt[4]{\frac{\beta_3\alpha_3}{4}} \tilde{\psi}_R^3 \star \psi_R^3 + \sqrt[4]{\frac{\alpha_2\beta_2}{4}} \tilde{\psi}_R^2 \star \psi_R^2. \quad (87)$$

and

$$A_L^1 = \sqrt[4]{\frac{\delta_1\lambda_1}{4}} \psi_L^1 \star \tilde{\psi}_L^1 + \sqrt[4]{\frac{\delta_3\lambda_3}{4}} \psi_L^3 \star \tilde{\psi}_L^3 \quad (88)$$

$$A_L^2 = \sqrt[4]{\frac{\delta_2\lambda_2}{4}} \psi_L^2 \star \tilde{\psi}_L^2 - \sqrt[4]{\frac{\delta_1\lambda_1}{4}} \tilde{\psi}_L^1 \star \psi_L^1 \quad (89)$$

$$A_L^3 = \sqrt[4]{\frac{\delta_3\lambda_3}{4}} \tilde{\psi}_L^3 \star \psi_L^3 + \sqrt[4]{\frac{\delta_2\lambda_2}{4}} \tilde{\psi}_L^2 \star \psi_L^2. \quad (90)$$

In terms of these fields the currents in (78) become

$$\hat{J}_1^- = \hat{J}_2^- = -\frac{i}{2} \begin{pmatrix} A_R^1 & 0 & 0 \\ 0 & A_R^2 & 0 \\ 0 & 0 & -A_R^3 \end{pmatrix} \quad \text{and} \quad \hat{J}_1^+ = \hat{J}_2^+ = -\frac{i}{2} \begin{pmatrix} A_L^1 & 0 & 0 \\ 0 & A_L^2 & 0 \\ 0 & 0 & -A_L^3 \end{pmatrix} \quad (91)$$

Therefore the action of the $NCGMT_1$ model (78) in terms of the Thirring field components become

$$\begin{aligned} S_{NCGMT_1} = \int dx^2 \sum_{i=1}^3 \{ & [2i \tilde{\psi}_L^i \star \partial_+ \psi_L^i + 2i \tilde{\psi}_R^i \star \partial_- \psi_R^i + im_i (\tilde{\psi}_L^i \star \psi_R^i - \psi_L^i \star \tilde{\psi}_R^i)] \\ & - 2(A_L^i \star A_R^i) \}, \end{aligned} \quad (92)$$

Next let us write the equations of motion for the field components derived from the action above. The following three equations of motion

$$\partial_+ \psi_L^3 = -\frac{1}{2} m_3 \psi_R^3 - i \sqrt[4]{\frac{\delta_3\lambda_3}{4}} \{ \psi_L^3 \star A_R^3 + A_L^1 \star \psi_L^3 \} \quad (93)$$

$$\partial_+ \tilde{\psi}_L^1 = -\frac{1}{2}m_1 \tilde{\psi}_R^1 + i\sqrt{\frac{\delta_1\lambda_1}{4}} \{\tilde{\psi}_L^1 \star A_R^1 - A_R^2 \star \tilde{\psi}_L^1\} \quad (94)$$

$$\partial_+ \tilde{\psi}_L^2 = -\frac{1}{2}m_2 \tilde{\psi}_R^2 + i\sqrt{\frac{\delta_2\lambda_2}{4}} \{\tilde{\psi}_L^2 \star A_R^2 + A_R^3 \star \tilde{\psi}_L^2\}, \quad (95)$$

will correspond to the matrix form (64) for $m = 1$.

One can obtain the equations of motion

$$\partial_+ \tilde{\psi}_L^3 = -\frac{1}{2}m_3 \tilde{\psi}_R^3 + i\sqrt{\frac{\delta_3\lambda_3}{4}} \{A_R^3 \star \tilde{\psi}_L^3 + \tilde{\psi}_L^3 \star A_R^1\} \quad (96)$$

$$\partial_+ \psi_L^1 = -\frac{1}{2}m_1 \psi_R^1 - i\sqrt{\frac{\delta_1\lambda_1}{4}} \{A_R^1 \star \psi_L^1 - \psi_L^1 \star A_R^2\} \quad (97)$$

$$\partial_+ \psi_L^2 = -\frac{1}{2}m_2 \psi_R^2 - i\sqrt{\frac{\delta_2\lambda_2}{4}} \{A_R^2 \star \psi_L^2 + \psi_L^2 \star A_R^3\}, \quad (98)$$

which in matrix form corresponds to eq. (64) for $m = 2$.

Similarly, one can obtain the equations of motion

$$\partial_- \psi_R^3 = \frac{1}{2}m_3 \psi_L^3 - i\sqrt{\frac{\alpha_3\beta_3}{4}} \{\psi_R^3 \star A_L^3 + A_L^1 \star \psi_R^3\} \quad (99)$$

$$\partial_- \tilde{\psi}_R^1 = \frac{1}{2}m_1 \tilde{\psi}_L^1 + i\sqrt{\frac{\alpha_1\beta_1}{4}} \{\tilde{\psi}_R^1 \star A_L^1 - A_L^2 \star \tilde{\psi}_R^1\} \quad (100)$$

$$\partial_- \tilde{\psi}_R^2 = \frac{1}{2}m_2 \tilde{\psi}_L^2 + i\sqrt{\frac{\alpha_2\beta_2}{4}} \{\tilde{\psi}_R^2 \star A_L^2 + A_L^3 \star \tilde{\psi}_R^2\}, \quad (101)$$

corresponding to $m = 2$ in (65).

Finally, the equations

$$\partial_- \tilde{\psi}_R^3 = \frac{1}{2}m_3 \tilde{\psi}_L^3 + i\sqrt{\frac{\alpha_3\beta_3}{4}} \{A_L^3 \star \tilde{\psi}_R^3 + \tilde{\psi}_R^3 \star A_L^1\} \quad (102)$$

$$\partial_- \psi_R^1 = \frac{1}{2}m_1 \psi_L^1 - i\sqrt{\frac{\alpha_1\beta_1}{4}} \{A_L^1 \star \psi_R^1 - \psi_R^1 \star A_L^2\} \quad (103)$$

$$\partial_- \psi_R^2 = \frac{1}{2}m_2 \psi_L^2 - i\sqrt{\frac{\alpha_2\beta_2}{4}} \{A_L^2 \star \psi_R^2 + \psi_R^2 \star A_L^3\}, \quad (104)$$

can be obtained from (65) in the case $m = 1$.

The set of equations of motions (93)-(104) are the $GL(3)$ extension of the equations of motion given before for the case $GL(2)$ NCMT₁ (see eqs. (5.11)-(5.14) of ref. [9]). In fact, the later system is contained in the $GL(3)$ extended model. For example, if one considers $\psi_L^1 = \psi_L^2 = \tilde{\psi}_L^1 = \tilde{\psi}_L^2 = 0$ in the eq. (93) then it is reproduced the equation (5.13) of reference [9] describing the single Thirring field ψ_3 provided that the parameters expression $\sqrt[4]{\frac{\delta_3\lambda_3\beta_3\alpha_3}{16}}$ corresponds to the coupling constant $\frac{\lambda}{2}$ of that reference.

The four field interaction terms in the action (92) can be re-written as a sum of Dirac type current-current terms for the various flavors ($j = 1, 2, 3$). In the constructions of the relevant currents the double-gauging of a $U(1)$ symmetry in the star-localized Noether procedure deserves a careful treatment [39, 9]. So, one has two types of currents for each flavor [9]

$$j_k^{(1)\mu} = \tilde{\psi}_k \gamma^\mu \star \psi_k, \quad (105)$$

$$j_k^{(2)\mu} = -\psi_k^T \gamma^0 \gamma^\mu \star \tilde{\psi}_k, \quad k = 1, 2, 3.. \quad (106)$$

Notice that in the commutative limit one has $j_k^{(1)\mu} = j_k^{(2)\mu}$. In order to write as a sum of current-current interaction terms it is necessary to impose the next constraints on the $\alpha_i, \beta_i, \delta_i, \lambda_i$ parameters

$$\frac{\delta_j \lambda_j}{\alpha_j \beta_j} = \kappa = \text{const.}; \quad j = 1, 2, 3. \quad (107)$$

Then the four-spinor interactions terms in (92), provided that (107) is taken into account, can be written as current-current interaction terms

$$\begin{aligned} -2 \sum_{i=1}^3 A_L^i A_R^i = & -g_{11} (j_{1\mu}^{(1)} \star j_1^{(1)\mu} + j_{1\mu}^{(2)} \star j_1^{(2)\mu}) - g_{22} (j_{2\mu}^{(1)} \star j_2^{(1)\mu} + j_{2\mu}^{(2)} \star j_2^{(2)\mu}) - \\ & g_{33} (j_{3\mu}^{(1)} \star j_3^{(1)\mu} + j_{3\mu}^{(2)} \star j_3^{(2)\mu}) + g_{12} (j_{1\mu}^{(1)} \star j_2^{(2)\mu}) - \\ & g_{23} (j_{2\mu}^{(1)} \star j_3^{(1)\mu}) - g_{13} (j_{1\mu}^{(2)} \star j_3^{(2)\mu}), \end{aligned} \quad (108)$$

where

$$g_{jj} = \frac{1}{4} \sqrt[4]{\alpha_j \beta_j \delta_j \lambda_j}, \quad g_{jk} = \frac{1}{2} \sqrt[4]{\alpha_j \beta_j \delta_k \lambda_k}, \quad (j \neq k); \quad j, k = 1, 2, 3. \quad (109)$$

These parameters g_{ij} define the coupling constants of the NC generalized Thirring model (NCGMT₁), even though that they are not mutually independent. Notice that considering the relationships (107) and (109) one has the three constraints

$$g_{ij} = 2 \sqrt{g_{ii} g_{jj}}, \quad i \neq j. \quad (110)$$

Taking into account the constraints (110) we are left with three independent coupling parameters at our disposal, so in order to study further properties such as the integrability and the zero-curvature formulations of the model one must consider the remaining three parameters, say the independent coupling parameters g_{11}, g_{22}, g_{33} . Then, substituting in the action (92) the current-current interaction terms (108) one has

$$\begin{aligned} S_{NCGMT_1} = & \int dx^2 \left\{ \sum_{i=1}^3 \left[2i \bar{\psi}_L^i \star \partial_+ \psi_L^i + 2i \bar{\psi}_R^i \star \partial_- \psi_R^i + im_i (\bar{\psi}_L^i \star \psi_R^i - \psi_L^i \star \bar{\psi}_R^i) \right] \right. \\ & - g_{11} (j_{1\mu}^{(1)} \star j_1^{(1)\mu} + j_{1\mu}^{(2)} \star j_1^{(2)\mu}) - g_{22} (j_{2\mu}^{(1)} \star j_2^{(1)\mu} + j_{2\mu}^{(2)} \star j_2^{(2)\mu}) - \\ & g_{33} (j_{3\mu}^{(1)} \star j_3^{(1)\mu} + j_{3\mu}^{(2)} \star j_3^{(2)\mu}) + g_{12} (j_{1\mu}^{(1)} \star j_2^{(2)\mu}) - \\ & \left. g_{23} (j_{2\mu}^{(1)} \star j_3^{(1)\mu}) - g_{13} (j_{1\mu}^{(2)} \star j_3^{(2)\mu}) \right\}. \end{aligned} \quad (111)$$

We define this model as the NC (generalized) massive Thirring model NCGMT₁ written in terms of the component fields. Its matrix version is understood to be the action (78) once the parameters relationships (107) are taken into account.

The two types of U(1) currents $j_{k\mu}^{(1)}, j_{k\mu}^{(2)}$ ($k=1,2,3$), respectively, satisfy the conservation equations

$$\partial_+ (\bar{\psi}_L^k \star \psi_L^k) + \partial_- (\bar{\psi}_R^k \star \psi_R^k) = 0, \quad \partial_+ (\psi_L^k \star \bar{\psi}_L^k) + \partial_- (\psi_R^k \star \bar{\psi}_R^k) = 0, \quad k = 1, 2, 3. \quad (112)$$

6.1.1 (Constrained) NC(c)GMT₁ zero-curvature formulation

The zero-curvature condition encodes integrability even in the NC extension of integrable models (see e.g. [9] and references therein), as this condition allows, for example, the construction of infinite conserved charges for them. In order to tackle this problem it is convenient to consider the matrix form of the equations of motion of the $GL(3)$ NC Thirring model (64)-(65) and intend to write them as originating from a zero-curvature condition. So, taking into account the gradation structure of the model let us consider the following Lax pair

$$A_- = E_{-3} + a[E_{-3}, W_1^+]_* + b[E_{-3}, W_2^+]_* + g_1[[E_{-3}, \hat{W}_1^+], \hat{W}_2^+]_* + g_2[[E_{-3}, \hat{W}_2^+], \hat{W}_1^+]_* \quad (113)$$

$$A_+ = -E_{+3} + b[E_{+3}, W_1^-]_* + a[E_{+3}, W_2^-]_* + \tilde{g}_1[[E_{+3}, \hat{W}_1^-], \hat{W}_2^-]_* + \tilde{g}_2[[E_{+3}, \hat{W}_2^-], \hat{W}_1^-]_* \quad (114)$$

where $a, b, g_1, g_2, \tilde{g}_1, \tilde{g}_2$ are some parameters to be determined below. Notice that the potentials A_{\pm} lie in the directions of the affine Lie algebra generators of grade $\mathcal{G}_{0,1,2,3}$ and $\mathcal{G}_{0,-1,-2,-3}$, respectively.

These matrix valued fields must be replaced into the zero-curvature equation

$$[\partial_+ + A_+, \partial_- + A_-]_* = 0, \quad (115)$$

We will use the following relationships which can easily be established

$$4\hat{J}_1^- = 4\hat{J}_2^- = -[[E_{+3}, \hat{W}_1^-], \hat{W}_2^-]_* = -[[E_{+3}, \hat{W}_2^-], \hat{W}_1^-]_*, \quad (116)$$

$$4\hat{J}_1^+ = 4\hat{J}_2^+ = [[E_{-3}, \hat{W}_1^+], \hat{W}_2^+]_* = [[E_{-3}, \hat{W}_2^+], \hat{W}_1^+]_* \quad (117)$$

So, the Lax pair can be rewritten as

$$A_- = E_{-3} + a[E_{-3}, W_1^+]_* + b[E_{-3}, W_2^+]_* + k_1 \hat{J}_1^+ \quad (118)$$

$$A_+ = -E_{+3} + b[E_{+3}, W_1^-]_* + a[E_{+3}, W_2^-]_* + k_2 \hat{J}_1^-, \quad (119)$$

where we have introduced the new parameters $k_{1,2}$ such that $\tilde{g}_1 + \tilde{g}_2 = -\frac{k_2}{4}$, and $g_1 + g_2 = \frac{k_1}{4}$

In order to get the relevant equations of motion (64)-(65) it is useful to take into consideration the gradation structure of the various terms. So, the terms of gradation (-1) in (115), taking into account (116), become

$$[E_{-3}, \partial_+ W_2^+]_* = +[E_{-3}, [E_3, W_1^-]]_* - k_2 (L_2^+)^{-1} [\hat{J}_1^-, [E_{-3}, \hat{W}_2^+]]_* L_2^+ + [F_1^+, F_2^-]_* \quad (120)$$

The equation (120) has the same structure as the equation of motion (64) (for $m = 1$) provided that we set $L_2^+ = L_1^+$, and impose the constraint

$$[F_1^+, F_2^-]_* = 0. \quad (121)$$

Next, looking for the gradation (+1) terms in (115) and using (117) we may get the equation

$$\left[E_3, \partial_- W_2^- \right]_* = -[E_3, [E_{-3}, W_1^+]]_* - k_1 (L_2^-)^{-1} [\hat{J}_1^+, [E_3, \hat{W}_2^-]]_* L_2^- + [F_2^+, F_1^-]_* \quad (122)$$

In a similar way, identifying $L_2^- = L_1^-$, and imposing the constraint

$$[F_2^+, F_1^-]_* = 0, \quad (123)$$

one notices that the equation (122) is equal to the equation of motion (65) (for $m = 1$).

Following the process we can write for the (± 2) gradations and conclude that in order to obtain the two equations of motion in (64)-(65) for $m = 2$, it is required the same conditions $L_2^\pm = L_1^\pm$ as above, without any new constraint.

We notice that the conditions $L_2^\pm = L_1^\pm$ which are related to the equations of motion for the gradations (± 1), (± 2) provide the following constraints between the initial parameters $(\alpha_i, \beta_i, \lambda_i, \delta_i)$

$$\alpha_i \beta_i = r_1; \quad \lambda_i \delta_i = r_2, \quad i = 1, 2, 3; \quad r_1, r_2 = \text{constants}. \quad (124)$$

In fact, these constraints are consistent with the parameters relationships (107) established above; however, eqs. (124) incorporate additional constant parameters r_1, r_2 such that $\kappa = r_2/r_1$. Additional relationships between the parameters arise by requiring that the above matrix equations derived from the zero-curvature equation to be consistent with the eqs. of motion (93)-(104). So, together with the relationships (124), it is required

$$\alpha_1 \alpha_2 \alpha_3 = \beta_1 \beta_2 \beta_3 \equiv r_1^{3/2}; \quad \lambda_1 \lambda_2 \lambda_3 = \delta_1 \delta_2 \delta_3 \equiv r_2^{3/2}, \quad k_1 = 2^{3/4} r_1^{1/8}, \quad k_2 = 2^{3/4} r_2^{1/8} \quad (125)$$

So, the set of current-current coupling constants g_{ij} in (111), which in the last section have been assumed to be equivalent to three independent parameters, in view of the additional relationships (125) they reduce to only one independent parameter g defined by

$$g_{12} = g_{23} = g_{13} = \frac{1}{2}g; \quad g_{ii} = \frac{1}{4}g, \quad i = 1, 2, 3; \quad g \equiv (r_1 r_2)^{1/4}. \quad (126)$$

Finally, for the zero gradation term there appears the following equation

$$k_1 \partial_+ \hat{J}_1^+ - k_2 \partial_- \hat{J}_1^- - ab[F_2^+, F_2^-] - ab[F_1^+, F_1^-] + k_1 k_2 [\hat{J}_1^-, \hat{J}_1^+] = 0. \quad (127)$$

We require this equation to be consistent with the full equations of motion (93)-(104) and the constraints (121) and (123). These constraints in terms of the fundamental fields become

$${}^1_R * \psi_L^2 = \psi_L^1 * \psi_R^2, \quad \psi_R^2 * {}^3_L = -\psi_L^2 * {}^3_R, \quad {}^3_L * \psi_R^1 = -{}^3_R * \psi_L^1 \quad (128)$$

and

$${}^3_R * {}^2_L = -\psi_L^3 * {}^2_R, \quad {}^1_L * \psi_R^3 = -{}^1_R * \psi_L^3, \quad {}^2_R * {}^1_L = {}^2_L * {}^1_R, \quad (129)$$

respectively.

In order to establish specific relationships between the parameters a, b and r_1, r_2 let us write (127) in terms of the fundamental fields

$$i(k_1 \partial_+ A_L^1 - k_2 \partial_- A_R^1) = -\frac{k_1 k_2}{2} (A_R^1 \star A_L^1 - A_L^1 \star A_R^1) - 2ab \{ im_1 (\sqrt[4]{\frac{\beta_1 \lambda_1}{4}} \psi_R^1 \star \tilde{\psi}_L^1 + \sqrt[4]{\frac{\alpha_1 \delta_1}{4}} \psi_L^1 \star \tilde{\psi}_R^1) + im_3 (\sqrt[4]{\frac{\alpha_3 \delta_3}{4}} \psi_R^3 \star \tilde{\psi}_L^3 + \sqrt[4]{\frac{\beta_3 \lambda_3}{4}} \psi_L^3 \star \tilde{\psi}_R^3) \} \quad (130)$$

$$i(k_1 \partial_+ A_L^2 - k_2 \partial_- A_R^2) \star = -\frac{k_1 k_2}{2} (A_R^2 \star A_L^2 - A_L^2 \star A_R^2) - 2ab \{ im_2 (\sqrt[4]{\frac{\beta_2 \lambda_2}{4}} \psi_R^2 \star \tilde{\psi}_L^2 + \sqrt[4]{\frac{\alpha_2 \delta_2}{4}} \psi_L^2 \star \tilde{\psi}_R^2) - im_1 (\sqrt[4]{\frac{\delta_2 \alpha_2}{4}} \tilde{\psi}_R^1 \star \psi_L^1 + \sqrt[4]{\frac{\beta_1 \lambda_1}{4}} \tilde{\psi}_L^1 \star \psi_R^1) \} \quad (131)$$

$$i(k_1 \partial_+ A_L^3 - k_2 \partial_- A_R^3) \star = \frac{k_1 k_2}{2} (A_R^3 \star A_L^3 - A_L^3 \star A_R^3) - 2ab \{ im_3 (\sqrt[4]{\frac{\beta_3 \lambda_3}{4}} \tilde{\psi}_R^3 \star \psi_L^3 + \sqrt[4]{\frac{\delta_3 \alpha_3}{4}} \tilde{\psi}_L^3 \star \psi_R^3) + im_2 (\sqrt[4]{\frac{\delta_2 \alpha_2}{4}} \tilde{\psi}_R^2 \star \psi_L^2 + \sqrt[4]{\frac{\beta_2 \lambda_2}{4}} \tilde{\psi}_L^2 \star \psi_R^2) \}. \quad (132)$$

Substituting the fields $A_{R,L}^j$, $j = 1, 2, 3$ in the form (85)-(90) into the eqs. (130)-(132) and taking into account the set of equations of motion (93)-(104) one gets the following relationships

$$2ab = \sqrt{\frac{g}{2}}; \quad (2)^{1/4} = r_1^{1/8} + r_2^{1/8}. \quad (133)$$

Therefore, we have established a zero-curvature formulation of a constrained version of the NCGMT₁ model. From this point forward this constrained model will be dubbed as NC(c)GMT₁.

Notice that the set of equations (130)-(132) contain the relevant eq. associated to the $SL(2)$ NC massive Thirring model written for its relevant zero gradation sector analogous to (127). So, for example, if one reduces the eq. (132) to get an equation for just one field, say ψ^3 , one has

$$i \left[k_1 \sqrt[4]{\frac{r_2}{4}} \partial_+ (\tilde{\psi}_L^3 \star \psi_L^3) - k_2 \sqrt[4]{\frac{r_1}{4}} \partial_- (\tilde{\psi}_R^3 \star \psi_R^3) \right] = -2iabm_3 \left(\sqrt[4]{\frac{\beta_3 \lambda_3}{4}} \tilde{\psi}_R^3 \star \psi_L^3 + \sqrt[4]{\frac{\delta_3 \alpha_3}{4}} \tilde{\psi}_L^3 \star \psi_R^3 \right) + \frac{k_1 k_2}{2} \sqrt[4]{\frac{r_1 r_2}{16}} (\tilde{\psi}_R^3 \star \psi_R^3 \star \tilde{\psi}_L^3 \star \psi_L^3 - \tilde{\psi}_L^3 \star \psi_L^3 \star \tilde{\psi}_R^3 \star \psi_R^3). \quad (134)$$

Now, taking into account $\alpha_3 = \beta_3 = \delta_3 = \lambda_3$ [$r_1 = r_2 \equiv r$] and the identifications $\psi^3 \rightarrow i r^{1/16} \psi$, $r^{1/2} \rightarrow \lambda$, $[m_3 r^{1/8}] \rightarrow m_\psi$ we arrive at the equation $\partial_- (\tilde{\psi}_R \star \psi_R) - \partial_+ (\tilde{\psi}_L \star \psi_L) = m_\psi (\tilde{\psi}_R \star \psi_L + \tilde{\psi}_L \star \psi_R) - i\lambda (\tilde{\psi}_R \star \psi_R \star \tilde{\psi}_L \star \psi_L - \tilde{\psi}_L \star \psi_L \star \tilde{\psi}_R \star \psi_R)$, which is the eq. (5.18) of the ref. [9].

6.1.2 NCGMT₁ sub-models

In the following we discuss some reduced models associated to the action (111) and its equations of motion (93)-(104).

NC massive Thirring (NCMT₁) models

The reduction of the NCGMT₁ model equations of motion (93)-(104) to a model with just one spinor field, say the components $\psi_{R,L}^1, \tilde{\psi}_{R,L}^1$ (consider the reduction $\psi_{R,L}^{2,3} = \tilde{\psi}_{R,L}^{2,3} = 0$) reproduces the NCMT₁ model which has been presented in [9, 10]. Notice that in this case the constraints (128) and (129), as well as the decoupling equations (62) [or in components (66)-(71)] become trivial. Let us emphasize that the full decoupling eqs. are satisfied by a subset of soliton solutions of the field equations of the $GL(2)$ NCATM₁ model such that the two sectors NCSG₁/NCMT₁ completely decouple [9]. Reducing in this way it is clear the appearance of three copies of the NCMT₁ model associated to the spinors ψ^1, ψ^2 and ψ^3 , respectively.

NC Bukhvostov-Lipatov (NCBL₁) model

Consider a reduced model with two fields, say $\psi_{R,L}^{1,2}, \tilde{\psi}_{R,L}^{1,2}$, achieved through the reduction $\psi_{R,L}^3 = \tilde{\psi}_{R,L}^3 = 0$. So, the Lagrangian (111) becomes

$$S_{NCTM} = \int dx^2 \left\{ \sum_{i=1}^{i=2} \left[2i\tilde{\psi}_L^i \star \partial_+ \psi_L^i + 2i\tilde{\psi}_R^i \star \partial_- \psi_R^i + im_i(\tilde{\psi}_L^i \star \psi_R^i - \psi_L^i \star \tilde{\psi}_R^i) \right] \right. \\ \left. - g_{11} (j_{1\mu}^{(1)} \star j_1^{(1)\mu} + j_{1\mu}^{(2)} \star j_1^{(2)\mu}) - g_{22} (j_{2\mu}^{(1)} \star j_2^{(1)\mu} + j_{2\mu}^{(2)} \star j_2^{(2)\mu}) \right. \\ \left. + g_{12} (j_{1\mu}^{(1)} \star j_2^{(2)\mu}) \right\}. \quad (135)$$

Remember that in ordinary space there is no distinction between the type of $j_i^{(1)}$ and $j_i^{(2)}$ currents for each flavor i ; so, the model (135) when written in ordinary space-time is known in the literature as the Bukhvostov-Lipatov model (BL) [29]. It has been claimed the classical integrability of the model in two special cases $g_{12} = 0$ ($2 \times$ MT model) and $g_{11} = g_{22} = 0$ (BL model) [in both cases consider $m_1 = m_2$] (see [27] and refs. therein). The quantum integrability of the BL model has been discussed in [28]. In view of the above discussion we define the model (135) as the first version of the NC Bukhvostov-Lipatov model (NCBL₁). Actually, there are additionally two reduction processes to arrive at NCBL₁ models, i.e. by setting $\psi^1 = 0$ and $\psi^2 = 0$ in (111), respectively.

(Constrained) NC Bukhvostov-Lipatov (NC(c)BL₁) and Lax pair formulation

Let us discuss a constrained version of the model (135). In view of the developments above one can establish the zero-curvature formulation of a constrained model associated to the model (135) by setting $\tilde{\psi}_{L,R}^3 = \psi_{L,R}^3 = 0$ in the matrices $W_{1,2}^\pm$ of the Lax pair eqs. (113)-(114), provided the constraints (128) and (129) given in the form $\psi_R^1 \star \psi_L^2 = \psi_L^1 \star \psi_R^2$ and $\tilde{\psi}_R^2 \star \tilde{\psi}_L^1 = \tilde{\psi}_L^2 \star \tilde{\psi}_R^1$, are considered. So, we claim that the model (135) is classically integrable provided that the above constraints are taken into account. In this way, provided that for version 2 one writes a copy of the model and their relevant constraints, one defines the (constrained) NC(c)BL_{1,2} models amenable to a Lax pair formulation.

In connection to this development, let us mention that a version of the BL model for Grassmanian fields

in usual space-time has also been recently shown to be associated to a Lax pair formulation provided some constraints are imposed [40].

In Fig. 1 we have outlined the various relationships. Notice that we have the two versions of NCATM_{1,2} and their strong/weak sectors described by the models NCGSG_{1,2} and NCGMT_{1,2}, respectively, as well as the relevant sub-models. We have emphasized the field contents in each stage of the reductions.

Some comments are in order here.

1. The action (111) (or its matrix form (78)) defines a three species NC generalized massive Thirring model. We have tried to write its eqs. of motion (64)-(65) [or in components (93)-(104)] as deriving from a zero-curvature formulation. We have proposed a Lax pair reproducing the same equations of motion provided that the constraints (121) and (123) [or in components (128) and (129)] are imposed. This fact suggests that the NCGMT₁ model (78) becomes integrable only for a sub-model defined by the eqs. of motion (93)-(104) provided the constraints (121) and (123) are satisfied [41]. So, one expects that a careful introduction of the constraints through certain Lagrange multipliers into the action will provide the Lagrangian formulation of an integrable sub-model of the NCGMT₁ theory.

2. Regarding the action related to the full zero-curvature equations of motion without constraints, determined by the set of eqs. (120) and (122), and the relevant eqs. in (64)-(65) written for $m = 2$, it is interesting to notice that the quadratic terms in the spinors present in the first couple of eqs. of motion (120) and (122) make it difficult to believe that one can find a local Lagrangian for the theory. Obviously, in that case we could not have a generalized massive Thirring model with a local Lagrangian involving bilinear (kinetic and mass terms) and usual current-current terms. This fact is intimately related to the presence of the eqs. (62) [or in components (66)-(71)] in the set of decoupling eqs. (58)-(63). In the commutative case the equations of type (62) have been incorporated in order to write a local Lagrangian for the GATM model in ref. [15]. Notice that the original theory (without constraints) allows a zero-curvature formulation; in fact, its Lax pair is just the one of the so-called conformal affine Toda model coupled to matter fields [18]. However, it does not possess a local Lagrangian formulation in terms of the fields of the model; namely, the Toda and the spinor (Dirac) fields.

3. Notice that in Fig. 1 we have emphasized the duality relationship NCGSG₁ \leftrightarrow NCGMT₁ since in this case the symmetry $U(1) \times U(1) \times U(1)$ of the NCGSG₁ model is implemented in the star-localized Noether procedure to get the three $U(1)$ currents of the NCGMT₁ sector. Regarding the relationships between the sub-models of the both sectors NCGSG₁ and NCGMT₁, it is clear the appearance of the duality NCGSG₁ \leftrightarrow NCMT₁ which has been discussed in the literature [9, 10]. This duality has also been discussed in the context of noncommutative bosonization of the massive Thirring model [42]. In addition, it is expected the duality relationship NCbBL₁ \leftrightarrow NCBL₁, since in the ordinary space-time the former is the bosonized version of the later model [27, 28]. Regarding this type of duality relationships between the remaining models a more careful investigation is needed, e.g. we have not been able to describe neither the spinor model corresponding to the NCDSG₁ model, nor the scalar sectors of the (constrained) NC(c)GMT₁ and

NC(c)BL₁ models, respectively.

6.2 NCGMT₂

As mentioned in the last paragraph of section 5 we expect that another NCGMT₂ version, with twice the number of fields of the NCGMT₁ theory, will appear when one performs a similar decoupling procedure for the extended system with $\{F_m^\pm, W_m^\pm\}$ and $\{\mathcal{F}_m^\pm, \mathcal{W}_m^\pm\}$ fields. In fact, a copy of the NCGMT₁ action (78), as well as the relevant zero-curvature equation of motion can be written for the fields $\{\mathcal{F}_m^\pm, \mathcal{W}_m^\pm\}$. Following similar steps one can construct a copy for each one of the sub-models presented above. Since it involves a direct generalization we will not present more details; however, see a corresponding construction for the $GL(2)$ case in ref. [9]. In this way one can get the NCGMT₂ model which is expected to be related to the NCGSG₂ model. Similarly to the NCGMT₁ case, one can expect that only a sub-model of NCGMT₂ will possess a zero-curvature formulation provided that a set of constraints similar to the eqs. (121) and (123), and a copy of them written for the fields \mathcal{F}_m^\pm ($m = 1, 2$) are considered.

7 Non-commutative solitons and kinks

It is a well known fact that the one-soliton solutions of certain models solve their NC counterparts. This feature holds for the SG model and its NCSG_{1,2} counterparts [9]. In the multi-field models, this feature means that the GSG model and its NCGSG_{1,2} extensions have a common subset of solutions, in particular the one-soliton and kink type solutions as we will see below. Of course the additional constraints, in the form of conservation laws which we have described before, e.g. the eqs. (41) and (54)-(55), respectively in the two versions of NCDSG models, must also be verified for the common subset of solutions. In fact, as we have noticed before they become trivial equations in the commutative limit.

The properties mentioned above reside on a simple observation: it is known that if $f(x_0, x_1)$ and $g(x_0, x_1)$ depend only on the combination $(x_1 - vx_0)$, then the product $f \star g$ coincides with the ordinary product $f \cdot g$ [11, 43]. Therefore, all the \star products in the NCGSG₁ system (24)-(26) reduce to the ordinary ones, so for these types of functions one has: NCGSG₁ \rightarrow GSG model; the GSG model was defined in (32)-(33) [see also eqs. (163)-(164)]. In the following we record the solutions with this property, i.e. the one-soliton solutions of the NCGSG₁ model and the kink type solution of the NCDSG₁ sub-model. Actually, the same analysis can be done for the NCGSG₂ case.

7.1 Solitons and kinks

Next we write the 1-soliton and 1-kink type solutions associated to the fields $\phi_{1,2}$ of the NCGSG₁ model, which in accordance to the discussion above reduce to the GSG system of eqs. (32)-(33). We will see that these solitons are, in fact, associated to the various sine-Gordon models obtained as sub-models of the GSG

theory, and the kink type solution corresponds to the double sine-Gordon sub-model [24].

1. Taking $\phi_1 = -\phi_2$ and $M_3 = M_2$, $\delta_i = 0$ in (32)-(33) one has

$$\phi_1 = 4 \arctan\{d \exp[\gamma_1(x - vt)]\}. \quad (136)$$

2. For $\phi_1 = \phi_2$ and $M_2 = M_3$, $M_1 = 0$ one has

$$\phi_1 = 4 \arctan\{d \exp[\gamma_2(x - vt)]\}. \quad (137)$$

Another SG model is given by setting $\phi_1 = \phi_2$ and $M_2 = M_3 = 0$ in (32)-(33) which leads to another soliton solution.

3. The kink solution is associated to the reduced double sine-Gordon model obtained by taking $\phi_1 = \phi_2 \equiv \phi$ and $M_3 = M_2$, $M_j \neq 0$. So, one has

$$\phi := 4 \arctan[d_K \sinh[\gamma_K(x - vt)]], \quad (138)$$

which is the usual DSG kink solution [44].

The $\gamma_{1,2}, \gamma_K, d, d_K, v$ above are some constant parameters.

8 Conclusions and discussions

Some properties of the NC extensions of the GATM model and their weak-strong phases described by the NCGMT_{1,2} and NCGSG_{1,2} models, respectively, have been considered. The Fig. 1 summarizes the relationships we have established, as well as the field contents in each sub-model.

In the $\theta \rightarrow 0$ limit we have the following correspondences: NCATM_{1,2} \rightarrow GATM; NCGSG_{1,2} (the real sector of model 2) \rightarrow GSG(plus a free scalar in the case of model 2); NCbBL_{1,2} \rightarrow bBL; NCDSG_{1,2} (the real sector of model 2) \rightarrow DSG(plus a free scalar in the case of model 2); NCGMT_{1,2} \rightarrow GMT (two copies in case of model 2); NCBL_{1,2} \rightarrow BL (two copies in case of model 2). In addition, the constrained versions NC(c)GMT_{1,2} and NC(c)BL_{1,2} give rise, in this limit, to the relevant (constrained) GMT and BL models, respectively, in ordinary space. To our knowledge, these are novel spinor integrable models.

The NCGMT_{1,2} Lagrangians describe three flavor massive spinors (case 2 considers twice the number of spinors) with current-current interactions among themselves. In the process of constructing the Noether currents one recognizes the $[U(1)]^3$ symmetry in both NCGMT_{1,2} models (in fact, as a subgroup of $[U(1)_C]^3$ in the model 2). We have provided the zero-curvature formulation of certain sub-models of the NCGMT_{1,2}. In fact, in order to write the eqs. of motion (93)-(104) as a zero-curvature equation for a suitable Lax pair one needs to impose the constraints (128)-(129), defining in this way the NC(c)GMT_{1,2} models. Likewise, the (constrained) NC(c)BL_{1,2} models possess certain Lax pairs.

The generalized sine-Gordon model, the usual SG model, the Bukhvostov-Lipatov model and the double sine-Gordon theory appear in the commutative limit of the both versions of the NCGSG_{1,2} models. We have

concluded that the $\text{NCGSG}_{1,2}$ models possess the same soliton and kink type solutions as their commutative counterparts. The appearance of the non-integrable double sine-Gordon model as a sub-model of the GSG model suggests that even the $\text{NCGSG}_{1,2}$ models are non-integrable theories for the arbitrary set of values of the parameter space, since they possess as sub-models the corresponding $\text{NCDSG}_{1,2}$ models. However, the $\text{NCGSG}_{1,2}$ models possess certain integrable directions in field space, as remarkable examples one has the $\text{NCSG}_{1,2}$ sub-models. In view of the presence of the (constrained) $\text{NC(c)GMT}_{1,2}$ and $\text{NC(c)BL}_{1,2}$ models with corresponding zero-curvature formulations, it is expected the existence of other integrable directions in the scalar sector, which we have not pursued further in the present work.

Actually, the procedures presented so far can directly be extended to the NCATM model for the affine Lie algebra $sl(n)$. Therefore one can conclude that, except for the usual MT model, a multi-flavor generalization ($n_F \geq 2$, n_F = number of flavors) of the massive Thirring model allows certain zero-curvature formulations only for its various constrained sub-models, in the both NC and ordinary space-time descriptions.

Except for the $\text{NCSG}_{1,2}$ models, which must correspond to the $\text{NCMT}_{1,2}$ models, whose Lax pair formulations have already been provided in the literature, we have not been able to find the Lax pair formulations of the $\text{NCGSG}_{1,2}$ remaining sub-models. The relevant scalar field models, and their Lax pair formulations, which must be the counterparts of the (constrained) $\text{NC(c)GMT}_{1,2}$ and $\text{NC(c)BL}_{1,2}$ models are missing; if such Lax pairs exist they are expected to contain certain nonlocal expressions of the fields of the $\text{NCGSG}_{1,2}$ models. These points deserve a careful consideration in future research.

Various aspects of the models studied above deserve attention in future research, e.g. the NC solitons and kinks of the $\text{NCATM}_{1,2}$ models and their relations with the confinement mechanism studied in ordinary space [24], the bosonization of the $\text{NCGMT}_{1,2}$ and their sub-models, the NC zero-curvature formulation of the bosonic sector of the $\text{NC(c)GMT}_{1,2}$ and $\text{NC(c)BL}_{1,2}$ models, as discussed above. In particular, in the bosonization process of the $\text{NCGMT}_{1,2}$ models, initiated in [10] for the $\text{NCMT}_{1,2}$ case, we believe that a careful understanding of the star-localized NC Noether symmetries, as well as the classical soliton spectrum would be desirable. In view of the rich spectra and relationships present in the above models it could be interesting to apply and improve some quantization methods, such as the one proposed in [31], in order to compute the soliton and kink masses quantum corrections. Another direction of research constitutes the NC zero-curvature formulations of the $\text{NCGMT}_{1,2}$ type models defined for Grassmannian fields.

Acknowledgments

We thank G. Takacs and R. Schiappa for communications and valuable comments about their works, and Prof. J. M. Guilarte for useful discussions about GSG models. HLC thanks IF-USP (São Paulo) for hospitality. HLC has been supported by FAPESP in the initial stage of the work and HB has been partially supported by CNPq.

A GSG as a reduced affine Toda model coupled to matter

We provide the algebraic construction of the $sl(3, \mathbb{C})$ conformal affine Toda model coupled to matter fields (CATM) following refs. [15, 18]. The reduction process to arrive at the classical GSG model closely follows the ref. [24]. The $sl(3, \mathbb{C})$ CATM model is a two-dimensional field theory involving four scalar fields and six Dirac spinors. The interactions among the fields are as follows: 1) in the scalars equations of motion there are the coupling of bilinears in the spinors to exponentials of the scalars. 2) Some of the equations of motion for the spinors have certain bilinear terms in the spinors themselves. That fact makes it difficult to find a local Lagrangian for the theory. Nevertheless, the model presents a lot of symmetries. It is conformally invariant, possesses local gauge symmetries as well as vector and axial conserved currents bilinear in the spinors. One of the most remarkable properties of the model is that it presents an equivalence between a $U(1)$ vector conserved current, bilinear in the spinors, and a topological currents depending only on the first derivative of some scalars. This property allow us to implement a bag model like confinement mechanism resembling what one expects to happen in QCD. The model possesses a zero-curvature representation based on the $\hat{sl}(3, \mathbb{C})$ affine Kac Moody algebra. It constitutes a particular example of the so-called conformal affine Toda models coupled to matter fields which has been introduced in [18]. The corresponding model associated to $\hat{sl}(2, \mathbb{C})$ has been studied in [16] where it was shown, using bosonization techniques, that the equivalence between the currents holds true at the quantum level and so the confinement mechanism does take place in the quantum theory.

The off-critical affine Toda model coupled to matter (ATM) is defined by gauge fixing the conformal symmetry [14] and imposing certain constraints in order to write a local Lagrangian for the model [15]. These treatments of the $sl(3, \mathbb{C})$ ATM model used the symplectic and on-shell decoupling methods to unravel the classical generalized sine-Gordon (GSG) and generalized massive Thirring (GMT) dual theories describing the strong/weak coupling sectors of the ATM model [20, 15, 14]. As mentioned above the ATM model describes some scalars coupled to spinor (Dirac) fields in which the system of equations of motion has a local gauge symmetry. Conveniently gauge fixing the local symmetry by setting some spinor bilinears to constants we are able to decouple the scalar (Toda) fields from the spinors, the final result is a direct construction of the classical generalized sine-Gordon model (GSG) involving only the scalar fields. In the spinor sector we are left with a system of equations in which the Dirac fields couple to the GSG fields. Another instance in which the quantum version of the generalized sine-Gordon theory arises is in the process of bosonization of the generalized massive Thirring model (GMT), which is a multi-flavor extension of the usual massive Thirring model such that, apart from the usual current-current self-interaction for each flavor, it presents current-current interactions terms among the various $U(1)$ flavor currents [45].

The zero-curvature condition (165) supplied with the potentials (166) gives the following equations of motion for the CATM model [18]

$$\frac{\partial^2 \phi_a}{4i e^\eta} = m_1 [e^{\eta - i\theta_a} \tilde{\psi}_R^l \psi_L^l + e^{i\theta_a} \tilde{\psi}_L^l \psi_R^l] + m_3 [e^{-i\theta_3} \tilde{\psi}_R^3 \psi_L^3 + e^{\eta + i\theta_3} \tilde{\psi}_L^3 \psi_R^3]; \quad a = 1, 2 \quad (139)$$

$$-\frac{\partial^2 \tilde{\nu}}{4} = im_1 e^{2\eta - \theta_1} \tilde{\psi}_R^1 \psi_L^1 + im_2 e^{2\eta - \theta_2} \tilde{\psi}_R^2 \psi_L^2 + im_3 e^{\eta - \theta_3} \tilde{\psi}_R^3 \psi_L^3 + m^2 e^{3\eta}, \quad (140)$$

$$-2\partial_+ \psi_L^1 = m_1 e^{\eta + i\theta_1} \psi_R^1, \quad -2\partial_+ \psi_L^2 = m_2 e^{\eta + i\theta_2} \psi_R^2, \quad (141)$$

$$2\partial_- \psi_R^1 = m_1 e^{2\eta - i\theta_1} \psi_L^1 + 2i \left(\frac{m_2 m_3}{im_1} \right)^{1/2} e^\eta (-\psi_R^3 \tilde{\psi}_L^2 e^{i\theta_2} - \tilde{\psi}_R^2 \psi_L^3 e^{-i\theta_3}), \quad (142)$$

$$2\partial_- \psi_R^2 = m_2 e^{2\eta - i\theta_2} \psi_L^2 + 2i \left(\frac{m_1 m_3}{im_2} \right)^{1/2} e^\eta (\psi_R^3 \tilde{\psi}_L^1 e^{i\theta_1} + \tilde{\psi}_R^1 \psi_L^3 e^{-i\theta_3}), \quad (143)$$

$$-2\partial_+ \psi_L^3 = m_3 e^{2\eta + i\theta_3} \psi_R^3 + 2i \left(\frac{m_1 m_2}{im_3} \right)^{1/2} e^\eta (-\psi_L^1 \tilde{\psi}_R^2 e^{i\theta_2} + \psi_L^2 \tilde{\psi}_R^1 e^{i\theta_1}), \quad (144)$$

$$2\partial_- \psi_R^3 = m_3 e^{\eta - i\theta_3} \psi_L^3, \quad 2\partial_- \tilde{\psi}_R^1 = m_1 e^{\eta + i\theta_1} \tilde{\psi}_L^1, \quad (145)$$

$$-2\partial_+ \tilde{\psi}_L^1 = m_1 e^{2\eta - i\theta_1} \tilde{\psi}_R^1 + 2i \left(\frac{m_2 m_3}{im_1} \right)^{1/2} e^\eta (-\psi_L^2 \tilde{\psi}_R^3 e^{-i\theta_3} - \tilde{\psi}_L^3 \psi_R^2 e^{i\theta_2}), \quad (146)$$

$$-2\partial_+ \tilde{\psi}_L^2 = m_2 e^{2\eta - i\theta_2} \tilde{\psi}_R^2 + 2i \left(\frac{m_1 m_3}{im_2} \right)^{1/2} e^\eta (\psi_L^1 \tilde{\psi}_R^3 e^{-i\theta_3} + \tilde{\psi}_L^3 \psi_R^1 e^{i\theta_1}), \quad (147)$$

$$2\partial_- \tilde{\psi}_R^2 = m_2 e^{\eta + i\theta_2} \tilde{\psi}_L^2, \quad -2\partial_+ \tilde{\psi}_L^3 = m_3 e^{\eta - i\theta_3} \tilde{\psi}_R^3, \quad (148)$$

$$2\partial_- \tilde{\psi}_R^3 = m_3 e^{2\eta + i\theta_3} \tilde{\psi}_L^3 + 2i \left(\frac{m_1 m_2}{im_3} \right)^{1/2} e^\eta (\tilde{\psi}_R^1 \tilde{\psi}_L^2 e^{i\theta_2} - \tilde{\psi}_R^2 \tilde{\psi}_L^1 e^{i\theta_1}), \quad (149)$$

$$\partial^2 \eta = 0, \quad (150)$$

where $\theta_1 \equiv 2\phi_1 - \phi_2$, $\theta_2 \equiv 2\phi_2 - \phi_1$, $\theta_3 \equiv \phi_1 + \phi_2$. Therefore, one has

$$\theta_3 = \theta_1 + \theta_2 \quad (151)$$

The ϕ fields are considered to be in general complex fields. In order to define the classical generalized sine-Gordon model we will consider these fields to be real.

Apart from the *conformal invariance* the above equations exhibit the $(U(1)_L)^2 \otimes (U(1)_R)^2$ *left-right local gauge symmetry*

$$\phi_a \rightarrow \phi_a + \xi_+^a(x_+) + \xi_-^a(x_-), \quad a = 1, 2 \quad (152)$$

$$\tilde{\nu} \rightarrow \tilde{\nu}; \quad \eta \rightarrow \eta \quad (153)$$

$$\psi^i \rightarrow e^{i(1+\gamma_5)\Xi_+^i(x_+) + i(1-\gamma_5)\Xi_-^i(x_-)} \psi^i, \quad (154)$$

$$\tilde{\psi}^i \rightarrow e^{-i(1+\gamma_5)(\Xi_+^i(x_+) - i(1-\gamma_5)(\Xi_-^i(x_-))} \tilde{\psi}^i, \quad i = 1, 2, 3; \quad (155)$$

$$\Xi_\pm^1 \equiv \pm \xi_\pm^2 \mp 2\xi_\pm^1, \quad \Xi_\pm^2 \equiv \pm \xi_\pm^1 \mp 2\xi_\pm^2, \quad \Xi_\pm^3 \equiv \Xi_\pm^1 + \Xi_\pm^2.$$

One can get global symmetries for $\xi_\pm^a = \mp \xi_\mp^a = \text{constants}$. For a model defined by a Lagrangian these would imply the presence of two vector and two chiral conserved currents. However, it was found only half of such currents [46]. This is a consequence of the lack of a Lagrangian description for the $sl(3)^{(1)}$ CATM in terms of the B and F^\pm fields (however see Appendix C for a local Lagrangian description of an off-critical and constrained sub-model). So, the vector current

$$J^\mu = \sum_{j=1}^3 m_j \bar{\psi}^j \gamma^\mu \psi^j \quad (156)$$

and the chiral current

$$J^{5\mu} = \sum_{j=1}^3 m_j \bar{\psi}^j \gamma^\mu \gamma_5 \psi^j + 2\partial_\mu (m_1 \phi_1 + m_2 \phi_2) \quad (157)$$

are conserved

$$\partial_\mu J^\mu = 0, \quad \partial_\mu J^{5\mu} = 0. \quad (158)$$

The conformal symmetry is gauge fixed by setting [14]

$$\eta = \text{const.} \quad (159)$$

The off-critical ATM model obtained in this way exhibits the vector and topological currents equivalence [18, 14]

$$\sum_{j=1}^3 m_j \bar{\psi}^j \gamma^\mu \psi^j \equiv \epsilon^{\mu\nu} \partial_\nu (m_1 \phi_1 + m_2 \phi_2), \quad m_3 = m_1 + m_2, \quad m_i > 0. \quad (160)$$

In the next steps we implement the reduction process to get the GSG model through a gauge fixing of the ATM theory [24]. The local symmetries (152)-(155) can be gauge fixed through

$$i^{-j} \psi^j = iA_j = \text{const.}; \quad {}^{-j} \gamma_5 \psi^j = 0. \quad (161)$$

From the gauge fixing (161) one can write the following bilinears

$$\tilde{\psi}_R^j \psi_L^j + \tilde{\psi}_L^j \psi_R^j = 0, \quad j = 1, 2, 3; \quad (162)$$

so, the eqs. (161) effectively comprises three gauge fixing conditions.

It can be directly verified that the gauge fixing (161) preserves the currents conservation laws (158), i.e. from the equations of motion (139)-(150) and the gauge fixing (161) together with (159) it is possible to obtain the currents conservation laws (158).

Taking into account the constraints (161) in the scalar sector, eqs. (139), we arrive at the following system of equations (set $\eta = 0$)

$$\partial^2 \phi_1 = M^1 \sin(2\phi_1 - \phi_2) + M^3 \sin(\phi_1 + \phi_2), \quad (163)$$

$$\partial^2 \phi_2 = M^2 \sin(2\phi_2 - \phi_1) + M^3 \sin(\phi_1 + \phi_2), \quad M^i \equiv 4A_i m_i, \quad i = 1, 2, 3. \quad (164)$$

The system of equations above considered for real fields $\phi_{1,2}$ as well as for real parameters M^i defines the *generalized sine-Gordon model* (GSG).

B The zero-curvature formulation of the $\hat{sl}(3)$ CATM model

We summarize the zero-curvature formulation of the $\hat{sl}(3)$ CATM model [18, 46]. Consider the zero-curvature condition

$$\partial_+ A_- - \partial_- A_+ + [A_+, A_-] = 0. \quad (165)$$

The potentials take the form

$$A_+ = -BF^+B^{-1}, \quad A_- = -\partial_- BB^{-1} + F^-, \quad (166)$$

with

$$F^+ = F_1^+ + F_2^+, \quad F^- = F_1^- + F_2^-, \quad (167)$$

where B and F_i^\pm contain the fields of the model. Let us define

$$F_m^\pm = \mp [E_{\pm 3}, W_{3-m}^\mp] \quad (168)$$

$$E_{\pm 3} = \frac{1}{6}[(2m_1 + m_2)H_1^{\pm 1} + (2m_2 + m_1)H_2^{\pm 1}], \quad m_3 = m_1 + m_2 \quad (169)$$

$$W_1^- = -\sqrt{\frac{4i}{m_3}}\psi_R^3 E_{\alpha 3}^{-1} + \sqrt{\frac{4i}{m_1}}\tilde{\psi}_R^1 E_{-\alpha 1}^0 + \sqrt{\frac{4i}{m_2}}\tilde{\psi}_R^2 E_{-\alpha 2}^0 \quad (170)$$

$$W_1^+ = \sqrt{\frac{4i}{m_1}}\psi_L^1 E_{\alpha 1}^0 + \sqrt{\frac{4i}{m_2}}\psi_L^2 E_{\alpha 2}^0 - \sqrt{\frac{4i}{m_3}}\tilde{\psi}_L^3 E_{-\alpha 3}^1 \quad (171)$$

$$W_2^- = -\sqrt{\frac{4i}{m_1}}\psi_R^1 E_{\alpha 1}^{-1} - \sqrt{\frac{4i}{m_2}}\psi_R^2 E_{\alpha 2}^{-1} + \sqrt{\frac{4i}{m_3}}\tilde{\psi}_R^3 E_{-\alpha 3}^0 \quad (172)$$

$$W_2^+ = \sqrt{\frac{4i}{m_3}}\psi_L^3 E_{\alpha 3}^0 - \sqrt{\frac{4i}{m_1}}\tilde{\psi}_L^1 E_{-\alpha 1}^1 - \sqrt{\frac{4i}{m_2}}\tilde{\psi}_L^2 E_{-\alpha 2}^1 \quad (173)$$

$$B = e^{i\theta_1 H_1^0 + i\theta_2 H_2^0} e^{\tilde{\nu}C} e^{\eta Q_{ppal}} \equiv g e^{\tilde{\nu}C} e^{\eta Q_{ppal}}. \quad (174)$$

$E_{\alpha i}^n, H_1^n, H_2^n$ and C ($i = 1, 2, 3; n = 0, \pm 1$) are some generators of $sl(3)^{(1)}$; Q_{ppal} being the principal gradation operator. The commutation relations for an affine Lie algebra in the Chevalley basis are

$$[H_a^m, H_b^n] = mC \frac{2}{\alpha_a^2} K_{ab} \delta_{m+n,0} \quad (175)$$

$$[H_a^m, E_{\pm\alpha}^n] = \pm K_{\alpha a} E_{\pm\alpha}^{m+n} \quad (176)$$

$$[E_\alpha^m, E_{-\alpha}^n] = \sum_{a=1}^r l_a^\alpha H_a^{m+n} + \frac{2}{\alpha^2} mC \delta_{m+n,0} \quad (177)$$

$$[E_\alpha^m, E_\beta^n] = \varepsilon(\alpha, \beta) E_{\alpha+\beta}^{m+n}; \quad \text{if } \alpha + \beta \text{ is a root} \quad (178)$$

$$[D, E_\alpha^n] = nE_\alpha^n, \quad [D, H_a^n] = nH_a^n. \quad (179)$$

where $K_{\alpha a} = 2\alpha \cdot \alpha_a / \alpha_a^2 = n_b^\alpha K_{ba}$, with n_a^α and l_a^α being the integers in the expansions $\alpha = n_a^\alpha \alpha_a$ and $\alpha / \alpha^2 = l_a^\alpha \alpha_a / \alpha_a^2$, and $\varepsilon(\alpha, \beta)$ the relevant structure constants.

Take $K_{11} = K_{22} = 2$ and $K_{12} = K_{21} = -1$ as the Cartan matrix elements of the simple Lie algebra $sl(3)$. Denoting by α_1 and α_2 the simple roots and the highest one by $\psi (= \alpha_1 + \alpha_2)$, one has $l_a^\psi = 1$ ($a = 1, 2$), and $K_{\psi 1} = K_{\psi 2} = 1$. Take $\varepsilon(\alpha, \beta) = -\varepsilon(-\alpha, -\beta)$, $\varepsilon_{1,2} \equiv \varepsilon(\alpha_1, \alpha_2) = 1$, $\varepsilon_{-1,3} \equiv \varepsilon(-\alpha_1, \psi) = 1$ and $\varepsilon_{-2,3} \equiv \varepsilon(-\alpha_2, \psi) = -1$.

One has $Q_{ppal} \equiv \sum_{a=1}^2 s_a \lambda_a^\vee \cdot H + 3D$, where λ_a^\vee are the fundamental co-weights of $sl(3)$, and the principal gradation vector is $s = (1, 1, 1)$ [47]. This gradation decomposes $\widehat{sl_3(\mathbb{C})}$ into the following subspaces

$$\hat{\mathcal{G}}_0 = \mathbb{C}H_1 \oplus \mathbb{C}H_2 \oplus \mathbb{C}C \oplus \mathbb{C}D = \mathbb{C}H_1 \oplus \mathbb{C}H_2 \oplus \mathbb{C}C \oplus \mathbb{C}Q_{ppal}, \quad (180)$$

and

$$\hat{\mathcal{G}}_{3m} = \mathbb{C} H_1^m \oplus \mathbb{C} H_2^m, \quad m \neq 0, \quad (181)$$

$$\hat{\mathcal{G}}_{3m+1} = \mathbb{C} E_{\alpha_1}^m \oplus \mathbb{C} E_{\alpha_2}^m \oplus \mathbb{C} E_{-\alpha_3}^{m+1}, \quad (182)$$

$$\hat{\mathcal{G}}_{3m+2} = \mathbb{C} E_{-\alpha_1}^{m+1} \oplus \mathbb{C} E_{-\alpha_2}^{m+1} \oplus \mathbb{C} E_{\alpha_3}^m. \quad (183)$$

C The off-critical and constrained $sl(3)$ ATM model

The off-critical and constrained $sl(3)$ affine Toda model coupled to matter fields (ATM) is defined by the action [15]

$$\begin{aligned} \frac{1}{k} I_{\text{ATM}}^{(3)} = & I_{\text{WZNW}}[g] + \int_M d^2x \left\{ \sum_{m=1}^2 \left[\langle F_m^-, g F_m^+ g^{-1} \rangle \right. \right. \\ & - \frac{1}{2} \langle E_{-3}, [W_m^+, \partial_+ W_{3-m}^+] \rangle + \langle F_m^-, \partial_+ W_m^+ \rangle \\ & \left. \left. + \frac{1}{2} \langle [W_m^-, \partial_- W_{3-m}^-], E_3 \rangle + \langle \partial_- W_m^-, F_m^+ \rangle \right] \right\}, \end{aligned} \quad (184)$$

where

$$I_{\text{WZNW}}[g] = \frac{1}{8} \int_M d^2x \text{Tr}(\partial_\mu g \partial^\mu g^{-1}) + \frac{1}{12} \int_D d^3x \epsilon^{ijk} \text{Tr}(g^{-1} \partial_i g g^{-1} \partial_j g g^{-1} \partial_k g), \quad (185)$$

is the Wess-Zumino-Novikov-Witten (WZNW) action for the matrix scalar field of the model. The first term inside the summation of (184) defines the form of the interactions and the remaining terms are the kinetic terms for the matrix fields associated to the spinors. The equations of motion derived from this action

$$\partial_-(g^{-1} \partial_+ g) = \sum_{m=1}^2 [F_m^-, g F_m^+ g^{-1}] \quad (186)$$

$$\partial_+ F_m^- = [E_{-3}, \partial_+ W_{3-m}^+], \quad \partial_- F_m^+ = -[E_3, \partial_- W_{3-m}^-], \quad (187)$$

$$\partial_+ W_m^+ = -g F_m^+ g^{-1}, \quad \partial_- W_m^- = -g^{-1} F_m^- g, \quad (188)$$

are equivalent to the above CATM equations of motion (139)-(150) provided the following constraints

$$\eta = 0 \quad (189)$$

$$[F_2^\pm, g^{\mp 1} F_1^\mp g^{\pm 1}] = 0, \quad (190)$$

are imposed. The first constraint defines an off-critical model, whereas the second ones allow a local Lagrangian description of the model. Let us emphasize that the constraints (190) amount to drop all the terms with spinor bilinears on the right hand side of the set of equations (142)-(144), (146)-(147) and (149), respectively. These constraints were introduced in refs. [20, 15] since they are trivially satisfied by the soliton type solutions of the full CATM model.

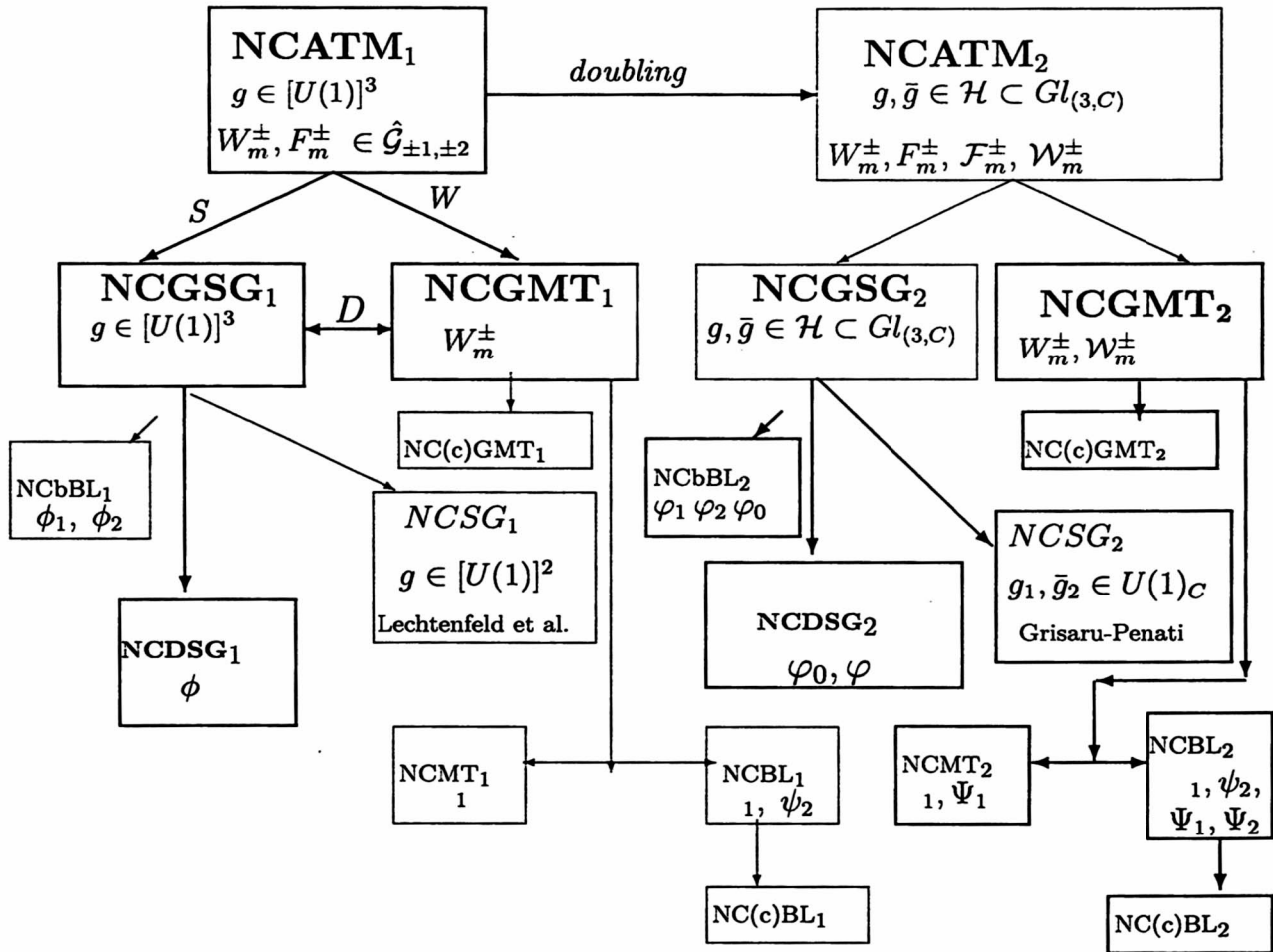


Fig. 1: $\text{NCATM}_{1,2}$: dual sectors, sub-models and field contents.

Duality: **S**= strong sector; **W**= weak sector; **D**= *S* - *W* duality. A *Lax pair* is available for $\text{NCSG}_{1,2}/\text{NCMT}_{1,2}$, $\text{NC(c)GMT}_{1,2}$ and $\text{NC(c)BL}_{1,2}$, respectively.

Dual sectors of the models $\text{NC(c)GMT}_{1,2}$, $\text{NC(c)BL}_{1,2}$ and $\text{NCDSG}_{1,2}$ are missing in the table above and deserve future investigations.

References

- [1] N. Seiberg and E. Witten, *JHEP* **9909** (1999) 032;
 A. Connes, M. R. Douglas and A. Schwarz, *JHEP* **9802** (1998) 003;
 M. R. Douglas and C. M. Hull, *JHEP* **9802** (1998) 008;
 N. R. F. Braga, H. L. Carrion and C. F. L. Godinho, *Phys. Lett.* **638B** (2006) 272; [arXiv:hep-th/0602212].
- [2] M. Hamanaka and K. Toda, *Phys. Lett.* **316A** (2003) 77.
- [3] F. Lizzi, S. Vaidya and P. Vitale, *Phys. Rev.* **D73** (2006) 125020.
 F. Lizzi and P. Vitale, *Mod. Phys. Lett.* **A23** (2008) 3307.
 S. Vaidya, *Phys. Lett.* **655B** (2007) 294.
- [4] M. Daniel and V. Vasumathi, *Physica* **D231** (2007) 10.
- [5] D. M. Hofman and J. M. Maldacena, *J. Physics: Math. and Gen.* **A39** (2006) 13095; [arXiv:hep-th/0604135].
- [6] K. Okamura and R. Suzuki, *Phys. Rev.* **D75** (2007) 046001; [arXiv:hep-th/0609026].
- [7] M. T. Grisaru and S. Penati, *Nucl. Phys.* **B655** (2003) 250.
- [8] M.T. Grisaru, L. Mazzanti, S. Penati and L. Tamassia, *JHEP* **0404** (2004) 057.
- [9] H. Blas, H. L. Carrion and M. Rojas, *JHEP* **0503** (2005) 037.
 H. Blas and H. L. Carrion, *International J. of Pure and Appl. Math.* **50** (2009) 213, see also [arXiv:hep-th/08093793].
- [10] H. Blas, *JHEP* **0506** (2005) 022.
- [11] I. Cabrera-Carnero, M. Moriconi, *Nucl. Phys.* **B673** (2003) 437.
- [12] O. Lechtenfeld, L. Mazzanti, S. Penati, A. D. Popov, L. Tamassia, *Nucl. Phys.* **B705** (2005) 477.
- [13] H. Blas, *Nucl. Phys.* **B596** (2001) 471; see also [arXiv:hep-th/0005037].
- [14] H. Blas and B.M. Pimentel, *Annals Phys.* **282** (2000) 67.
- [15] H. Blas, *JHEP* **0311** (2003) 054.
 H. Blas, "Generalized sine-Gordon and massive Thirring models", A chapter in "New Developments in Soliton Research, p. 123-147" Editor: L.V. Chen (Nova Science Publishers, Hauppauge NY, ISBN: 1-59554-561-8; 2006), see also [arXiv:hep-th/0407020].
- [16] H. Blas and L.A. Ferreira, *Nucl. Phys.* **B571** (2000) 607.

- [17] S. Ghosh, *Phys. Lett.* **558B** (2003) 245;
M.Bota and P. Minces, *Eur. Phys. J.* **C34** (2004) 393.
- [18] L.A. Ferreira, J.-L. Gervais, J.Sánchez Guillen and M.V.Saveliev, *Nucl. Phys.* **B470** (1996) 236.
- [19] H. Blas, *Phys. Rev.* **D66** (2002) 127701; see also [arXiv:hep-th/0005130].
- [20] J. Acosta, H. Blas, *J. Math. Phys.* **43** (2002) 1916, see also [arXiv:hep-th/0407020].
- [21] S. Brazovskii, *J. Phys. IV* **10** (2000) 169; also in [arXiv:cond-mat/0006355];
A.J. Heeger, S. Kivelson, J.R. Schrieffer and W. -P. Wu, *Rev. Mod. Phys.* **60** (1988) 782.
- [22] D.G. Barci and L. Moriconi, *Nucl. Phys.* **B438** (1995) 522.
- [23] R. Jackiw and C. Rebbi, *Phys. Rev.* **D13** (1976) 3398;
J. Goldstone and F. Wilczek, *Phys. Rev. Lett.* **47** (1981) 986;
J.A. Mignaco and M.A. Rego Monteiro, *Phys. Rev.* **D31** (1985) 3251.
- [24] H. Blas and H.L. Carrion, *JHEP* **0701** (2007) 027.
H. Blas and H.L. Carrion, “Solitons as baryons and qualitons as constituent quarks in two-dimensional QCD”, A chapter of the book “**The Physics of Quarks: New Research**, p. 197-225”, Editors: N.L. Watson and T.M. Grant (Nova Science Publishers, Hauppauge NY, ISBN:978-1-60456-802-8 **2009**), see also [arXiv:hep-th/08030536].
- [25] H. Blas, *JHEP* **0703** (2007) 055.
H. Blas, “Generalized sine-Gordon model and bayons in two-dimensional QCD”, A chapter in “**High Energy Physics Research Advances**, p. 79-119” Editor: T.P. Harrison et al, (Nova Science Publishers, Inc. Hauppauge NY, ISBN:978-1-60456-304-7, **2008**), see also [arXiv:hep-th/07093667].
- [26] S. Deser and R. Jackiw, *Phys. Lett.* **139B** (1984) 371;
S. E. Hjelmeland and U. Lindstrom, “Duality for the Nonspecialist”, [arXiv:hep-th/9705122].
- [27] M. Ameduri, C. J. Efthimiou and B. Gerganov, *Mod. Phys. Lett.* **A14** (1999) 2341.
- [28] H. Saleur, *J. Physics: Math. and Gen.* **A32** (1999) L207.
- [29] A.P. Bukhvostov and L.N. Lipatov, *Nucl. Phys.* **B180** (1981) 116.
- [30] L.A. Ferreira, J.L. Miramontes and J. Sánchez Guillén, *Nucl. Phys.* **B449** (1995) 631,
- [31] J. M. Guilarte, A. Alonso-Izquierdo, W. Garcia Fuertes, M. de la Torre Mayado and M. J. Senosiain, “Quantum fluctuations around low-dimensional topological defects,” PoS **ISFTG2009**, 013 (2009); [arXiv:0909.2107 [hep-th]].

- [32] G. Mussardo, V. Riva and G. Sotkov, *Nucl. Phys.* **B687** (2004) 189.
D. Controzzi and G. Mussardo, *Phys. Rev. Lett.* **92** (2004) 021601.
G. Mussardo, *Nucl. Phys.* **B779** (2007) 101.
- [33] G. Takacs and F. Wagner, *Nucl. Phys.* **B741** (2006) 353.
Z. Bajnok, L. Palla, G. Takacs and F. Wagner, *Nucl. Phys.* **B601** (2001) 503.
- [34] G. Takacs, *Nucl. Phys.* **B825** (2010) 466.
- [35] S. Orfanidis, *Phys. Rev.* **D14** (1976) 472.
- [36] P. Garbaczewski, *J. Math. Phys.* **24** (1983) 1806;
T. Bhattacharyya, *J. Math. Phys.* **46** (2005) 012301.
- [37] A.F. Rañada, *Phys. Rev.* **D33** (1986) 1714;
A. Alvarez and A.F. Rañada, *Phys. Rev.* **D38** (1988) 3330.
- [38] L. V. Belvedere and A. F. Rodrigues, *J. Phys. A: Math. Theor.* **42** (2009) 015401.
- [39] Y. Liao and K. Sibold, *Phys. Lett.* **586B** (2004) 420.
- [40] J. F. Gomes, D. M. Schmidt and A. H. Zimerman, *Nucl. Phys.* **B821** (2009) 553; [arXiv:0901.4040 [hep-th]].
- [41] H. Blas, H.L. Carrion and B.M. Cerna, "Some comments on the integrability of the noncommutative generalized massive Thirring model", PoS(ISFTG)042; see also arXiv:0910.0541[hep-th].
- [42] C. Nunez, K. Olsen and R. Schiappa, *JHEP* **0007** (2000) 030.
- [43] A. Dimakis and F. Mueller-Hoissen, "A noncommutative version of the nonlinear Schrödinger equation," [arXiv:hep-th/0007015].
- [44] D. K. Campbell, M. Peyrard and P. Sodano, *Physica* **D19** (1986) 165.
- [45] H. Blas, *Eur. Phys. J.* **C37** (2004) 251, see also [arXiv:hep-th/0409269] .
H. Blas, "Bosonization, soliton-particle duality and Mandelstam-Halpern operators", A chapter in "Trends in Boson Research, p. 79-108", Editor: A.V. Ling (Nova Science Publishers, Hauppauge NY, ISBN:1-59454-521-9. 2006), see also [arXiv:hep-th/0409269].
- [46] A.G. Bueno, L.A. Ferreira and A.V. Razumov, *Nucl. Phys.* **B626** (2002) 463.
- [47] V.G. Kac, Infinite dimensional Lie algebras, 3rd edition (Cambridge University Press, Cambridge, 1990).

The variable mass Thirring/sine Gordon duality and continuous topological configurations

H. Blas and J.M. Jaramillo

Instituto de Física
Universidade Federal de Mato Grosso
Av. Fernando Correa, s/n, Coxipó
78060-900, Cuiabá - MT - Brazil

Abstract

We consider a two-dimensional integrable and conformally invariant field theory possessing two Dirac spinors and three scalar fields. There are two types of interactions, the ones which couple bilinear terms in the spinors to exponentials of the scalars and the self-interaction of one of the scalar fields. Its integrability properties are based on the $sl(2)$ affine Kac-Moody algebra. The off-critical sub-model is related to a variety of fermion-soliton systems studied in the literature in which the shape of the static soliton is prescribed. In this paper we uncover the weak and strong sectors of the submodel as being the (variable mass) Thirring and sine-Gordon models, respectively. However, certain topological configurations in between are identified in which the both scalar and spinor type fields interact to greater or lesser extent. The spectrum of the Thirring sector is obtained.

1 Introduction

Many aspects of soliton-fermion systems in two dimensions, such as the quantum corrections, have been studied in the literature; however, some important features at the zero order are yet to be investigated. In this context, the $sl(2)$ conformal affine Toda model coupled to matter fields (CATM) is a two dimensional integrable and conformally invariant field theory with spinor and scalar fields, which presents topological solitons in the pseudoscalar field. It has two Dirac spinors coupled to a pseudoscalar field φ and a scalar field η which plays the role of a conformal symmetry ‘gauge’ connection. By choosing a particular constant solution $\eta = \text{const.}$ of the equations of motion the conformal symmetry is broken, thus obtaining a massive theory, the so-called affine Toda model coupled to matter (Dirac) fields (ATM) involving the Dirac spinors and the pseudoscalar field φ only. The integrability of the theory is established using a zero curvature formulation of its equations of motion based on the $\widehat{sl}(2)$ affine Kac-Moody algebra. The general solution as well as explicit one and two soliton solutions have been obtained in [1, 2]. Besides the conformal and local gauge symmetries, the model presents chiral symmetry and some discrete symmetries. However, one of its main properties is that for certain solutions there is an equivalence between the $U(1)$ Noether current, involving the spinors only, and the topological current associated to the pseudoscalar soliton. That fact was established in [1] at the classical level, and in [2] at the quantum mechanical level. It implies that the density of the $U(1)$ charge has to be concentrated in the regions where the pseudoscalar field has non vanishing space derivative. The one-soliton of the theory is of the sine-Gordon type with fractional topological charge, and therefore the charge density is concentrated inside the soliton. The CATM model constitutes an excellent laboratory to test ideas about confinement [2, 3], the role of solitons in quantum field theories [2], duality transformations interchanging solitons and particles [4].

The off-critical affine Toda model coupled to matter (Dirac) fields has been considered in many physical situations related to a fermion-soliton system. Some of which consider the shape of the static soliton as prescribed. It appears in the study of the dynamical Peierls energy gap generation in one-dimensional charge-density wave systems [5], quantum field-theory description of tunneling in the integer quantum Hall effect [6], discussions of fractional charge or topological charges induced in the ground state of a Fermi system by its coupling to a chiral field [7, 8], the spectrum and string tension of the low-energy effective Lagrangian of QCD_2 [9]. It is also closely related to the integrable variable mass sine-Gordon model presented in [10]. Recently, an exact numerical solutions have been presented for a related version of the model [11] and the Casimir energy of that system has been computed, such that a static pseudoscalar field is prescribed with two adjustable parameters [12].

In [2] the general soliton solutions of the ATM model have been obtained using two techniques: the dressing transformations and the Hirota’s method. The basic idea was to look for vacuum configurations where the Lax potentials lie in an abelian (up to central terms) subalgebra of $\widehat{sl}(2)$. The solitons are obtained by performing the dressing transformations from those vacuum configurations. Such procedure led quite naturally to the definition of tau-functions, and then the Hirota’s method was easily implemented too. The authors discussed the conditions for the solutions to be real, and evaluate the topological charges. The interactions of the solitons was studied by calculating their time delays. It has been shown that it

is attractive, and the time delays have shown to be the same as those for the sine-Gordon solitons. An interesting aspect of the solutions is that when the two Dirac spinors of the theory are related by a reality condition, then either the soliton or the anti-soliton disappears from the spectrum. It has been interpreted as indicative of the existence of a duality involving the solitons and the spinor particles.

We show that the pseudoscalar can represent topological solitons with infinitely many different shapes such that the topological charge can take any real value depending on a particular gauge fixing condition. This is a consequence of the fact that the ATM model does not possess a particle conjugation symmetry [7]. An important property of the model under consideration is that it possesses special type of scalar and spinor particles. That is envisaged by mapping the system to either, the sine-Gordon model or the massive Thirring model through different gauge fixing procedures. In that scenario the solitons of the sine-Gordon model are interpreted as the spinor particles of the Thirring theory. Thus, the ATM theory is related to the sine-Gordon and the massive Thirring models and its N-soliton solutions can be constructed using the well known solutions of these models. The role played by the local gauge symmetries and the physical soliton spectrum on the Noether and topological currents equivalence at the classical level deserve clarification. The first thing to be established is if such equivalence is maintained by the gauge invariant fields. That issue can be established exactly by implementing a continuous transformation between the strong and weak coupling sectors of the model. This is reminiscent of the so-called smooth bosonization technique interpolating between the sine-Gordon and the massive Thirring theories through convenient gauge fixings [13].

In the intermediate region of field space in which the spinor and scalar fields interact to greater or lesser extent the system possesses a rich physical properties. This intermediate model defines a fermion-soliton system which has been widely studied in the literature.

The paper is organized as follows: in section 2 we summarize the main properties of the model, following [2]. The submodels are presented: in section 3 the variable mass Thirring model and in section 4 the sine-Gordon model. In addition, in subsection 3.1 it is found the bound states and the energies in the massive Thirring sector of the model. Finally, section 5 discusses the main results.

2 The model

Consider the two-dimensional field theory defined by the Lagrangian¹

$$\mathcal{L} = -\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi + i\bar{\psi}\gamma^\mu\partial_\mu\psi - m_\psi\bar{\psi}e^{\eta+i\sqrt{8/k}\varphi\gamma_5}\psi + \frac{1}{2}\partial_\mu\nu\partial^\mu\nu - \frac{k}{8}m_\psi^2e^{2\eta} \quad (1)$$

where φ , η and ν are real scalar fields, and ψ is a Dirac spinor. It is defined $\bar{\psi} \equiv \tilde{\psi}^T \gamma_0$, with $\tilde{\psi}$ being a second Dirac spinor. In many applications of [2] it has been considered the reality condition

$$\tilde{\psi} = e_\psi \psi^* \quad (2)$$

where e_ψ is a real constant. It has been shown that the sign of e_ψ plays an important role in determining the spectrum of soliton solutions. The theory (1) in [1] is defined for a general complex fields, i.e. a complex

¹Lagrangian (1) follows if one makes $\varphi \rightarrow \sqrt{\frac{2}{k}}\varphi$, $\nu \rightarrow \frac{1}{k}\nu$, $\psi \rightarrow \sqrt{\frac{1}{k}}\psi$, $\tilde{\psi} \rightarrow \sqrt{\frac{1}{k}}\tilde{\psi}$ in the model of ref. [2]. The CATM model is related to the Wess-Zumino-Novikov-Witten model, and $k = \frac{n}{2\pi}$, $n \in \mathbb{Z}$.

Lagrangian. In order to define a suitable physical Lagrangian one can follow the prescription to restrict the model to a subspace of well-behaved classical solutions [4]. For example, for $e_\psi = -1$ one can drop an overall minus sign from the Lagrangian in order to get usual kinetic terms for the spinor and φ scalar fields in such a way that one defines a submodel with a Hamiltonian bounded from below. In fact, the one and two soliton solutions for φ and the corresponding spinor solutions satisfy the reality conditions (2).

One of the main conclusions in ref. [2] is that without the reality condition (2) the theory (1) has two Dirac spinors and it also has the soliton and anti-soliton solutions associated to φ , since both signs of the soliton charge are admissible. However, if one imposes the reality condition (2) the theory (1) loses one Dirac spinor and also one-soliton (two-soliton) solution, since for a given choice of e_ψ only one topological charge $Q_{\text{topol}} = \text{sign}(e_\psi) \frac{1}{2} [Q_{\text{topol}} = \text{sign}(e_\psi)]$ corresponding to one-soliton (two-soliton) or one-anti-soliton (two-anti-soliton), is permitted. That is indicative of the existence of some sort of duality between solitons and antisolitons associated to φ and the spinor particles ψ , $\tilde{\psi}$, respectively. So, one can have (without (2)) soliton and anti-soliton solutions with φ real. However, a charge zero solution corresponding to the scattering of a soliton and anti-soliton for φ real does not exist. Such solution exists, however, for φ complex but asymptotically real.

The corresponding equations of motion are²

$$\partial^2 \varphi = im_\psi \sqrt{\frac{8}{k}} \bar{\psi} \gamma_5 e^{\eta + i\sqrt{\frac{8}{k}} \varphi} \gamma_5 \psi, \quad (3)$$

$$\partial^2 \nu = -2m_\psi \bar{\psi} e^{\eta + i\sqrt{\frac{8}{k}} \varphi} \gamma_5 \psi - \frac{k}{2} m_\psi^2 e^{2\eta}, \quad (4)$$

$$\partial^2 \eta = 0, \quad (5)$$

$$i\gamma^\mu \partial_\mu \psi = m_\psi e^{\eta + i\sqrt{\frac{8}{k}} \varphi} \gamma_5 \psi, \quad (6)$$

$$i\gamma^\mu \partial_\mu \tilde{\psi} = m_\psi e^{\eta - i\sqrt{\frac{8}{k}} \varphi} \gamma_5 \tilde{\psi}, \quad (7)$$

The integrability properties, the construction of the general solution including the solitonic ones and many other properties were discussed in [1, 2]. We want to discuss some special features of the theory at the classical level as well as the soliton solutions. In particular, we will identify the gauge invariant scalar and spinor fields associated to soliton solutions. We start by reviewing the symmetries of (1) and the eqs. of motion (3)-(7).

Left-right local symmetries. The Lagrangian (1) is invariant³ under the local $U(1)_L \otimes U(1)_R$ transformations

$$\varphi \rightarrow \varphi + \sqrt{\frac{k}{2}} \xi_+(x_+) + \sqrt{\frac{k}{2}} \xi_-(x_-); \quad \nu \rightarrow \nu; \quad \eta \rightarrow \eta \quad (8)$$

and

$$\psi \rightarrow e^{-i[(1+\gamma_5)\xi_+(x_+) - (1-\gamma_5)\xi_-(x_-)]} \psi; \quad \tilde{\psi} \rightarrow e^{i[(1+\gamma_5)\xi_+(x_+) - (1-\gamma_5)\xi_-(x_-)]} \tilde{\psi} \quad (9)$$

$U(1)$ global symmetry. Notice that, by taking $\xi_+(x_+) = -\xi_-(x_-) = -\frac{1}{2}\theta$, with $\theta = \text{const.}$, one gets a

²We take $\gamma_0 = -i \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $\gamma_1 = -i \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\gamma_5 = \gamma_0 \gamma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

³Notice that the eqs. of motion are invariant under these transformations, whereas the Lagrangian (1) is invariant up to a surface term.

global $U(1)$ transformation

$$\varphi \rightarrow \varphi; \quad \nu \rightarrow \nu; \quad \eta \rightarrow \eta; \quad \psi \rightarrow e^{i\theta} \psi; \quad \tilde{\psi} \rightarrow e^{-i\theta} \tilde{\psi} \quad (10)$$

The corresponding Noether current is given by

$$J^\mu = \frac{1}{k} \bar{\psi} \gamma^\mu \psi, \quad \partial_\mu J^\mu = 0. \quad (11)$$

Chiral symmetry. In addition, if one takes $\xi_+(x_+) = \xi_-(x_-) = -\frac{1}{2} \alpha$, with $\alpha = \text{const.}$, one gets the global chiral symmetry

$$\psi \rightarrow e^{i\gamma_5 \alpha} \psi; \quad \tilde{\psi} \rightarrow e^{-i\gamma_5 \alpha} \tilde{\psi}; \quad \varphi \rightarrow \varphi - \alpha; \quad \nu \rightarrow \nu; \quad \eta \rightarrow \eta \quad (12)$$

with the corresponding Noether current

$$J_5^\mu = \frac{1}{k} \bar{\psi} \gamma_5 \gamma^\mu \psi + \frac{1}{2} \sqrt{\frac{2}{k}} \partial^\mu \varphi; \quad \partial_\mu J_5^\mu = 0 \quad (13)$$

Topological charge. One can shift the φ field as $\varphi \rightarrow \varphi + n\pi \sqrt{\frac{k}{2}}$, keeping all the other fields unchanged, that the Lagrangian is left invariant. That means that the theory possesses an infinite number of vacua, and the topological charge

$$Q_{\text{topol.}} \equiv \int dx j^0, \quad j^\mu = \frac{1}{2\pi} \sqrt{\frac{2}{k}} \epsilon^{\mu\nu} \partial_\nu \varphi \quad (14)$$

can assume non trivial values. This model possesses fractional topological charges as we will see below.

CP-like symmetry. In addition, the Lagrangian (1) is invariant under the transformation

$$x_+ \leftrightarrow x_-; \quad \psi \leftrightarrow i\epsilon\gamma_0 \tilde{\psi}; \quad \tilde{\psi} \leftrightarrow -i\epsilon\gamma_0 \psi; \quad \varphi \leftrightarrow \varphi; \quad \eta \leftrightarrow \eta; \quad \nu \leftrightarrow \nu \quad (15)$$

where $\epsilon = \pm 1$. The reality condition (2) breaks such CP symmetry, for any real value of the constant e_ψ ⁴. In addition, the theory (1) is not invariant under spatial parity. In general, solitons are transformed into anti-solitons by parity transformations; so, it is very unlikely that the theory has the both type of solitons associated to φ for a given choice of e_ψ in (2). However, the both type of solitons arise for a submodel defined for a gauge invariant field, which turns out to be the sine-Gordon field, see below.

Conformal symmetry. The model (1) is invariant under the conformal transformations⁵

$$x_+ \rightarrow \hat{x}_+ = f(x_+), \quad x_- \rightarrow \hat{x}_- = g(x_-), \quad (16)$$

with f and g being analytic functions; and with the fields transforming as

$$\begin{aligned} \varphi(x_+, x_-) &\rightarrow \hat{\varphi}(\hat{x}_+, \hat{x}_-) = \varphi(x_+, x_-), \\ e^{-\nu(x_+, x_-)} &\rightarrow e^{-\hat{\nu}(\hat{x}_+, \hat{x}_-)} = (f')^\delta (g')^\delta e^{-\nu(x_+, x_-)}, \\ e^{-\eta(x_+, x_-)} &\rightarrow e^{-\hat{\eta}(\hat{x}_+, \hat{x}_-)} = (f')^{\frac{1}{2}} (g')^{\frac{1}{2}} e^{-\eta(x_+, x_-)}, \\ \psi(x_+, x_-) &\rightarrow \hat{\psi}(\hat{x}_+, \hat{x}_-) = e^{\frac{1}{2}(1+\gamma_5) \log(f') - \frac{1}{2} + \frac{1}{2}(1-\gamma_5) \log(g')}^{-\frac{1}{2}} \psi(x_+, x_-), \end{aligned} \quad (17)$$

⁴The CP symmetry is preserved if one takes $e_\psi = \pm i$.

⁵We are using $x_\pm = t \pm x$, and so, $\partial_\pm = \frac{1}{2}(\partial_t \pm \partial_x)$, and $\partial^2 = \partial_t^2 - \partial_x^2 = 4\partial_+ \partial_-$.

where the conformal weight δ , associated to $e^{-\nu}$, is arbitrary, and $\tilde{\psi}$ transforms in the same way as ψ .

Associated to the conformal symmetry (16) there are two chiral currents

$$\mathcal{J} = -i\frac{1}{k}\tilde{\psi}^T(1+\gamma_5)\psi + \sqrt{\frac{2}{k}}\partial_+\varphi + \partial_+\eta, \quad \bar{\mathcal{J}} = i\frac{1}{k}\tilde{\psi}^T(1-\gamma_5)\psi + \sqrt{\frac{2}{k}}\partial_-\varphi + \partial_-\eta \quad (18)$$

satisfying

$$\partial_-\mathcal{J} = 0; \quad \partial_+\bar{\mathcal{J}} = 0. \quad (19)$$

Observe, from (17), that the currents \mathcal{J} and $\bar{\mathcal{J}}$ have conformal weights (1,0) and (0,1) respectively. Under the conformal transformations (16), the chiral currents transform as

$$\mathcal{J}(x_+) \rightarrow [\ln f'(x_+)]^{-1}(\mathcal{J}(x_+) - [\ln f'(x_+)]'), \quad (20)$$

$$\bar{\mathcal{J}}(x_-) \rightarrow [\ln g'(x_-)]^{-1}(\bar{\mathcal{J}}(x_-) - [\ln g'(x_-)]') \quad (21)$$

Then, given a solution of the model, one can always map it, under a conformal transformation, into a solution where

$$\mathcal{J} = 0; \quad \bar{\mathcal{J}} = 0. \quad (22)$$

Such a procedure amounts to gauging away the free field η [4]. In other words, (22) are constraints implementing a Hamiltonian reduction. So, for every regular solution $\eta \equiv \bar{\eta}$ the CATM defined on a space-time (x_+, x_-) will correspond to an off-critical submodel, the so-called affine Toda model coupled to matter (ATM). For the particular solution $\bar{\eta} = 0$ the CATM and the ATM are defined on the same space-time. The Lagrangian for the set of equations of motion corresponding to the Toda and Dirac fields becomes

$$\mathcal{L} = -\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi + i\bar{\psi}\gamma^\mu\partial_\mu\psi - m_\psi e^{\bar{\eta}}\bar{\psi}e^{i\sqrt{8/k}\varphi}\gamma_5\psi. \quad (23)$$

This Lagrangian is not conformal invariant and defines the ATM model. As presented in [1] the soliton masses of the theory have a topological character due to the spontaneous breakdown of the conformal symmetry; in fact, the behavior of the space derivative of the field ν at $x = \pm\infty$ provide the soliton masses in the framework of the dressing method implemented on the orbit of the vacuum $\bar{\eta} = \varphi = \psi = \tilde{\psi} = 0$, $\nu = -\frac{1}{8}m_\psi^2 x_+ x_-$.

The structure of the vacuum of the system (23) is rather complicated due to the presence of the spinor fields and the left-right local symmetry (8)-(9) which remains as a symmetry of the ATM model. Of course, the vacuum solution is not unique, and it was remained an open problem to identify which one provides the absolute minimum of the energy. In the next sections we will address this point and provide the possible values for $\bar{\eta}$ at the classical level.

Let us mention that the field η plays the role of a Higgs field, since it not only spontaneously breaks the conformal symmetry, but also because its vacuum expectation value sets the mass scale of the theory [1]. The vacuum solution $\bar{\eta} = 0$ was used in [1, 2] to perform the dressing transformation in order to obtain soliton

solutions, which are in the orbit of a vacuum solution. It can be also noticed that these transformations do not excite the field η and the solitonic solutions are solutions of the gauge fixed model.

One can easily check that the constraints (22), once we have set $\eta = \eta_0 = \text{const.}$, are equivalent to

$$\frac{1}{2\pi} \sqrt{\frac{2}{k}} \epsilon^{\mu\nu} \partial_\nu \varphi = \frac{1}{k\pi} \bar{\psi} \gamma^\mu \psi. \quad (24)$$

Therefore, in the reduced model, the Noether current (11) is proportional to the topological current (14). That fact has profound consequences in the properties of such theory. For instance, it implies (taking $\tilde{\psi}$ to be the complex conjugate of ψ) that the charge density $\psi^\dagger \psi$ is proportional to the space derivative of φ . Consequently, the Dirac field is confined to live in regions where the field φ is not constant. The best example of that is the one-soliton solution of (1) which was calculated in [1] and it is given by

$$\varphi = 2\sqrt{\frac{k}{2}} \arctan \left(\exp \left(2m_\psi (x - x_0 - vt) / \sqrt{1 - v^2} \right) \right) \quad (25)$$

$$\psi = \sqrt{k} e^{i\theta} \sqrt{m_\psi} e^{m_\psi (x - x_0 - vt) / \sqrt{1 - v^2}} \begin{pmatrix} \left(\frac{1-v}{1+v} \right)^{1/4} \frac{1}{1 + i e^{2m_\psi (x - x_0 - vt) / \sqrt{1 - v^2}}} \\ - \left(\frac{1+v}{1-v} \right)^{1/4} \frac{1}{1 - i e^{2m_\psi (x - x_0 - vt) / \sqrt{1 - v^2}}} \end{pmatrix} \quad (26)$$

$$\nu = -\frac{k}{2} \log \left(1 + \exp \left(4m_\psi (x - x_0 - vt) / \sqrt{1 - v^2} \right) \right) - \frac{k}{8} m_\psi^2 x_+ x_- \quad (27)$$

$$\eta = 0 \quad (28)$$

and the solution for $\tilde{\psi}$ is the complex conjugate of ψ . Notice that, from (14), one has $Q_{\text{topol.}} = \frac{1}{2}$ for the solution (28) which is in accordance with the topological charge definition in [1]. In fact, this is a fractional charge and it differs from the one presented in [2] where the charge has been defined as twice that of the ref. [1]. In Fig. 1 this solution is plotted in the static case.

Notice that the solution for φ is of the sine-Gordon type soliton, and therefore $\partial_x \varphi$ is non-vanishing only in a region of size of the order of m_ψ^{-1} . In addition, the solution for ψ resembles the massive Thirring model type soliton. One can check that (28) satisfies (24), and so is a solution of the reduced model. However, one must emphasize that these solutions are not truly sine-Gordon/massive Thirring solitons, e.g. its topological charge is $+1/2$ whereas the 1-soliton of the usual sine-Gordon model possesses integer charge $+1$.

We point out that the condition (24) together with the equations of motion for the Dirac spinors (6)-(7) imply the equation of motion for φ , namely (3). Therefore in the reduced model, defined by the constraints (22), one can replace a second order differential equation, i.e. (3), by two first order equations, i.e. (24).

On the other hand, in the so-called smooth bosonization method [13] the bosonic and fermionic formulations of the same theory is understood as two different gauge fixings of a larger gauge-invariant action containing the both fermionic and bosonic fields. Equivalence of the both sectors of the theory then amount to the usual gauge-fixing independence of the S-matrix elements in gauge theories. This method has been used to derive the bosonization rules for certain models. It implies that the standard bosonization and fermionization prescriptions interpolate two extreme field representations of a “larger” theory which in a completely smooth way can be brought to almost any desired form in between. The equivalence between certain bosonic and fermionic theories (e.g. sine-Gordon/massive Thirring) then has profound consequences in the properties of the original gauge invariant theory, extending to a continuum of theories in which fermions

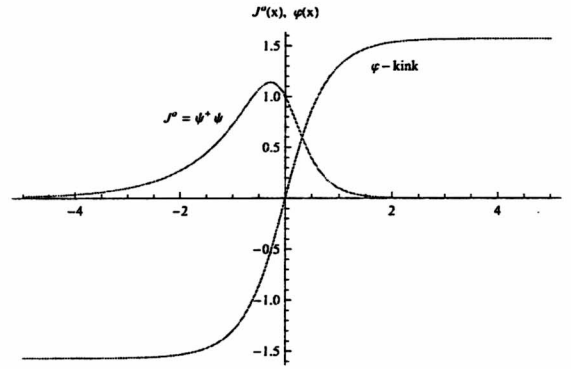


Figure 1: The graphs of the charge density $[\psi^\dagger \psi]$ and $\sqrt{\frac{2}{k}}[\varphi + \pi/2]$ -kink corresponding to the parameters $k = 1$, $m_\psi = 1$, $x_0 = 0$, $v = 0$ for time $t = 0$.

and bosons appear interactively to a greater or lesser extent. This is the main idea of the so-called smooth bosonization and we will follow in the next sections.

3 Gauge invariant spinors: variable mass Thirring

Let us consider the transformation

$$\psi_L = e^{\frac{i}{2}\sqrt{\frac{8}{k}}(\varphi-\gamma)}\chi_L, \quad \psi_R = e^{-\frac{i}{2}\sqrt{\frac{8}{k}}(\varphi+\gamma)}\chi_R, \quad (29)$$

$$\tilde{\psi}_L = e^{-\frac{i}{2}\sqrt{\frac{8}{k}}(\varphi-\gamma)}\tilde{\chi}_L, \quad \tilde{\psi}_R = e^{\frac{i}{2}\sqrt{\frac{8}{k}}(\varphi+\gamma)}\tilde{\chi}_R. \quad (30)$$

Then, one can rewrite the Lagrangian (1) as

$$\begin{aligned} \mathcal{L} = & -2\partial_+\varphi\partial_-\varphi + 2i\tilde{\chi}_R\partial_-\chi_R + \sqrt{\frac{8}{k}}\chi_R\tilde{\chi}_R(\partial_-\varphi + \partial_-\gamma) + 2i\tilde{\chi}_L\partial_+\chi_L - \\ & \sqrt{\frac{8}{k}}\chi_L\tilde{\chi}_L(\partial_+\varphi - \partial_+\gamma) + 2\partial_+\nu\partial_-\eta - \frac{k}{8}m_\psi^2e^{2\eta} - im_\psi e^\eta(\tilde{\chi}_R\chi_L - \tilde{\chi}_L\chi_R). \end{aligned} \quad (31)$$

The following eqs. of motion arise

$$4\partial_-\partial_+\varphi = \sqrt{\frac{8}{k}}\partial_-(\tilde{\chi}_R\chi_R) - \sqrt{\frac{8}{k}}\partial_+(\tilde{\chi}_L\chi_L), \quad (32)$$

$$\partial^2\nu = -\frac{km_\psi^2}{2}e^{2\eta} - 2im_\psi e^\eta(\tilde{\chi}_R\chi_L - \tilde{\chi}_L\chi_R), \quad (33)$$

$$2i\partial_-\chi_R = -\sqrt{\frac{8}{k}}\chi_R(\partial_-\varphi + \partial_-\gamma) + im_\psi e^\eta\chi_L, \quad (34)$$

$$-2i\partial_-\tilde{\chi}_R = -\sqrt{\frac{8}{k}}\tilde{\chi}_R(\partial_-\varphi + \partial_-\gamma) - im_\psi e^\eta\tilde{\chi}_L, \quad (35)$$

$$2i\partial_+\chi_L = +\sqrt{\frac{8}{k}}\chi_L(\partial_+\varphi - \partial_+\gamma) - im_\psi e^\eta\chi_R, \quad (36)$$

$$-2i\partial_+\tilde{\chi}_L = +\sqrt{\frac{8}{k}}\tilde{\chi}_L(\partial_+\varphi - \partial_+\gamma) + im_\psi e^\eta\tilde{\chi}_R, \quad (37)$$

$$\partial^2\eta = 0. \quad (38)$$

From the set of eqs. (34)-(37) one can get⁶

$$\partial_+(\tilde{\chi}_L\chi_L) + \partial_-(\tilde{\chi}_R\chi_R) = 0, \quad (39)$$

which together with (32) implies the next relationships

$$\partial_+\varphi = \frac{1}{2}\sqrt{\frac{8}{k}}(\chi_R\tilde{\chi}_R) + \partial_+\sigma, \quad (40)$$

$$\partial_-\varphi = -\frac{1}{2}\sqrt{\frac{8}{k}}(\chi_L\tilde{\chi}_L) - \partial_-\sigma, \quad (41)$$

where σ is a free scalar field satisfying $\partial^2\sigma = 0$. Notice that the $U(1)$ current conservation equation (39) is precisely the compatibility condition of the system of eqs. (40)-(41). As noticed before, the ATM second order eq. (3) allows to be replaced by a system of first order eqs. (40)-(41) also in the new spinor parametrization. However, now it needs the introduction of an additional free field σ . In order to decouple the spinors from the scalars fields φ and s one can choose

$$\gamma \equiv \sigma, \quad (42)$$

i.e. s being a free field, and then substitute the derivatives of the φ , s fields appearing in (40)-(41) into the right hand sides of the eqs. (34)-(37). The outcome will be the set of eqs.

$$2i\partial_-\chi_R - g\chi_R\chi_L\chi_L^* - im_\psi e^\eta\chi_L = 0, \quad (43)$$

$$2i\partial_+\chi_L - g\chi_L\chi_R\chi_R^* + im_\psi e^\eta\chi_R = 0, \quad g \equiv \frac{4e_\psi}{k} \quad (44)$$

plus two eqs. arising by taking the complex conjugates of the above. This is the massive Thirring type model coupled to the η field. In fact, for every regular solution of the free field η these eqs. correspond to a MT model with variable mass. The case $\eta = \eta_0$ defines the usual MT model with fermion mass $m_\psi e^{\eta_0}$ and coupling constant $g = \frac{4e_\psi}{k}$. Notice that the χ field ($\tilde{\chi}_{R,L} = e_\psi\chi_{R,L}^*$) inherits the relationship (2) and it is responsible for the appearance of the parameter e_ψ in the eqs (43)-(44).

Finally, let us rewrite the currents equivalence (24) in terms of the scalar field φ and the spinor χ ; so, taking into account (29)-(30) one can write the eq. (24) as

$$\frac{1}{2\pi}\sqrt{\frac{8}{k}}\epsilon^{\mu\nu}\partial_\nu\varphi = \frac{2e_\psi}{\pi k}\tilde{\chi}\gamma^\mu\chi, \quad (45)$$

where $\tilde{\chi} \equiv \chi^\dagger\gamma^0$.

3.1 1-soliton solution of the MT model and Hamiltonian eigenvalues

For later purpose let us write the 1-soliton solution

$$\chi_R = \frac{\tau_R}{\tau_0}, \quad \chi_L = \frac{\tau_L}{\tau_1}, \quad \tilde{\chi}_R = \frac{\tilde{\tau}_R}{\tau_1}, \quad \tilde{\chi}_L = \frac{\tilde{\tau}_L}{\tau_0}, \quad (46)$$

⁶Alternatively, one can obtain the eq. (39) as an equation of motion for the lagrangian (31) associated to the field γ .

where the tau functions become

$$\tau_0 = 1 - i \frac{a_+ a_-}{4} e^{2\Gamma(x_+, x_-)}, \quad \tau_1 = 1 + i \frac{a_+ a_-}{4} e^{2\Gamma(x_+, x_-)}, \quad (47)$$

$$\tau_R = \sqrt{\frac{m_\psi k z}{e_\psi}} e^{\frac{\eta_0}{2} \frac{a_+}{2}} e^{\Gamma(x_+, x_-)}, \quad \tau_L = -\sqrt{\frac{m_\psi k}{z e_\psi}} e^{\frac{\eta_0}{2} \frac{a_+}{2}} e^{\Gamma(x_+, x_-)}, \quad (48)$$

$$\tilde{\tau}_R = \sqrt{\frac{m_\psi k z}{e_\psi}} e^{\frac{\eta_0}{2} \frac{a_-}{2}} e^{\Gamma(x_+, x_-)}, \quad \tilde{\tau}_L = -\sqrt{\frac{m_\psi k}{z e_\psi}} e^{\frac{\eta_0}{2} \frac{a_-}{2}} e^{\Gamma(x_+, x_-)}, \quad (49)$$

$$\Gamma(x_+, x_-) = \frac{m_\psi}{2} e^{\eta_0} \left(z x_+ - \frac{1}{z} x_- \right) \quad (50)$$

$$= \frac{m_\psi}{2} e^{\eta_0} (z + 1/z) (x - vt), \quad v \equiv \frac{1 - z^2}{1 + z^2}. \quad (51)$$

The complex parameters a_+ , a_- must satisfy $[a_+ a_-] \in \mathbb{R}$.

It is convenient to introduce the static version of the system (43)-(44) as an eigenvalue problem

$$i\partial_x \chi_R + \frac{4e_\psi}{k} \chi_R \chi_L \chi_L^* + i m_\psi e^{\eta_0} \chi_L = E_\kappa \chi_R, \quad (52)$$

$$-i\partial_x \chi_L + \frac{4e_\psi}{k} \chi_L \chi_R \chi_R^* - i m_\psi e^{\eta_0} \chi_R = E_\kappa \chi_L. \quad (53)$$

We will use the eigenvalues and eigenfunctions of the Hamiltonian associated to the MT model in our discussions of the ATM system below. So, let us supply a family of bound state solutions for the system (52)-(53). The following modified tau functions

$$\tau_0(\kappa) = 1 - i \hat{a}_+ \hat{a}_- \left(\frac{E_\kappa^2 + \kappa^2}{\kappa^2} \right) e^{i\delta} e^{2\kappa x}, \quad \tau_1(\kappa) = 1 + i \hat{a}_+^* \hat{a}_-^* \left(\frac{E_\kappa^2 + \kappa^2}{\kappa^2} \right) e^{-i\delta} e^{2\kappa x}, \quad (54)$$

$$\tau_R(\kappa) = \sqrt{\frac{i m_\psi k e^{\eta_0}}{e_\psi}} e^{i\delta/2} \hat{a}_+ e^{\kappa x}, \quad \tau_L(\kappa) = \sqrt{\frac{-i m_\psi k e^{\eta_0}}{e_\psi}} e^{-i\delta/2} \hat{a}_+ e^{\kappa x}, \quad (55)$$

$$\tilde{\tau}_R(\kappa) = \sqrt{\frac{i m_\psi k e^{\eta_0}}{e_\psi}} e^{-i\delta/2} \hat{a}_- e^{\kappa x}, \quad \tilde{\tau}_L(\kappa) = -\sqrt{\frac{-i m_\psi k e^{\eta_0}}{e_\psi}} e^{i\delta/2} \hat{a}_- e^{\kappa x}, \quad (56)$$

$$E_\kappa = \pm \sqrt{m_\psi^2 e^{2\eta_0} - \kappa^2}, \quad \delta \equiv \arctan \left[\frac{\kappa}{E_\kappa} \right], \quad (57)$$

replaced into the relationships (46) provide the relevant static solutions of (52)-(53) with eigenvalues E_κ . Notice that in the limit $E_\kappa \rightarrow 0$ one has $\kappa \rightarrow m_\psi e^{\eta_0}$, $\delta \rightarrow \pi/2$, and the tau functions (54)-(56) become the static version of (47)-(49) provided that $z = 1$, $\hat{a}_+ = \frac{a_+}{2} e^{-i\pi/2}$, $\hat{a}_- = \frac{a_-}{2}$. Therefore, the static soliton solution provided by (47)-(49) is an eigenfunction with zero ($E_\kappa = 0$) eigenvalue.

In the next sections we follow, at the classical level, related ideas considered in the smooth bosonization. The off-critical model (23) still possesses the left-right local symmetry (8)-(9); so, in order to study its classical soliton spectrum we consider certain gauge fixings such that the solutions are easy to obtain. So, we will pursue certain gauge invariant fields in the both sectors of the ATM theory (23), i.e. in the purely scalar or spinor formulations, respectively. In the process we will have the opportunity to visualize certain intermediate models in which the scalars and spinors appear interactively to certain extent.

4 Gauge invariant field θ : sine-Gordon model

In order to identify a gauge invariant scalar we introduce a linear system associated to the spinor components $\chi^T = (\chi_R, \chi_L)$. In order to accomplish this construction one notices that the Dirac equations (34)-(37) expresses $\partial_+\chi_L$ and $\partial_-\chi_R$ in terms of χ_R, χ_L , however the derivatives $\partial_+\chi_R$ and $\partial_-\chi_L$ are missing. Let us assume that the all derivatives of χ to be written in terms of χ itself as

$$\partial_+\chi = M\chi, \quad \partial_-\chi = N\chi \quad (58)$$

where the 2×2 matrices M, N are defined by

$$M = \begin{pmatrix} A & B \\ -m & \partial_+\phi_2 \end{pmatrix}, N = \begin{pmatrix} \partial_-\phi_1 & m \\ C & D \end{pmatrix}, \quad m \equiv \frac{m_\psi}{2} e^\eta, \quad (59)$$

$$\phi_1 \equiv \frac{i}{2} \sqrt{\frac{8}{k}} (\varphi + \gamma), \quad \phi_2 \equiv -\frac{i}{2} \sqrt{\frac{8}{k}} (\varphi - \gamma), \quad (60)$$

with A, B, C, D being in general complex fields. Then, one has eleven real fields defining our matrices M, N , i.e. the set $\{A, B, C, D, i\phi_1, i\phi_2, m\}$. This field space will be reduced below to a set of gauge invariant fields. The integrability condition for the system (58) becomes

$$\partial_-M - \partial_+N + [M, N] = 0 \quad (61)$$

which provides

$$\partial_+\partial_-\phi_1 - \partial_-A - m^2 - BC = 0 \quad (62)$$

$$\partial_-\partial_+\phi_2 - \partial_+D - BC - m^2 = 0 \quad (63)$$

$$\partial_+C + m\partial_-\phi_1 + AC - C\partial_+\phi_2 - mD + \partial_-m = 0 \quad (64)$$

$$\partial_-B + mA + BD - B\partial_-\phi_1 - m\partial_+\phi_2 - \partial_-m = 0 \quad (65)$$

From the eqs. (62)-(63) one gets the relationship

$$\partial_-(\partial_+\phi_1 - A) = \partial_+(\partial_-\phi_2 - D) \quad (66)$$

Hence, there exists a potential Θ such that A and D can be written as

$$D = \partial_-(\phi_2 - \Theta), \quad A = \partial_+(\phi_1 - \Theta) \quad (67)$$

Since the fields $\phi_{1,2}$ are purely imaginary, the last relationships amount to write $A^R = \partial_+\Theta^R$, $D^R = \partial_-\Theta^R$ which implies $\partial_-A^R - \partial_+D^R = 0$. The superscripts R, I label the real and imaginary parts of the fields. Similarly, one has $iA^I = \partial_+(\phi_1 - i\Theta^I)$, $iD^I = \partial_-(\phi_2 - i\Theta^I)$ implying $i(\partial_-A^I - \partial_+D^I) + \partial_+\partial_-(\phi_2 - \phi_1) = 0$. Since the real and imaginary components of the fields A and D , respectively, are related to each other one has that the number of fields for M, N are reduced to nine.

Taking into account (67) the eqs. (62)-(63) reduce to the single equation

$$\partial_-\partial_+\Theta - m^2 - BC = 0, \quad (68)$$

whereas, the eqs. (64)-(65) become

$$\partial_- B - B\partial_- \Lambda_2 = m\partial_+ \Lambda_1 + \partial_+ m \quad (69)$$

$$\partial_+ C - C\partial_+ \Lambda_1 = -m\partial_- \Lambda_2 - \partial_- m, \quad (70)$$

$$\Lambda_1 \equiv \Theta + (\phi_2 - \phi_1), \quad \Lambda_2 \equiv \Theta - (\phi_2 - \phi_1). \quad (71)$$

So, the initial system of eqs reduce to (68) and (69)-(70). Since B, C, Θ are in general complex fields one can consider an equivalent system of six real differential equations. However one has at our disposal nine fields, i.e. six real fields related to B, C, Θ , and three real fields associated to the real scalar m and the purely imaginary fields $\phi_{1,2}$. Then, one is left with three gauge freedoms for M and N , since the nine real field components are related by six real differential equations.

Let us parametrize the fields B, C as

$$B = f^+(x_+)e^{\Lambda_2} + m\rho_1 e^{-\Lambda_1} \quad (72)$$

$$C = g^-(x_-)e^{\Lambda_1} + m\rho_2 e^{-\Lambda_2} \quad (73)$$

where f^+, g^- and $\rho_{1,2}$ are complex fields. Substituting these parametrizations into (69)-(70) one has

$$e^{\Lambda_2} \partial_- (m\rho_1 e^{-\Lambda_1 - \Lambda_2}) = m\partial_+ \Lambda_1 + \partial_+ m; \quad e^{\Lambda_1} \partial_+ (m\rho_2 e^{-\Lambda_1 - \Lambda_2}) = -m\partial_- \Lambda_2 - \partial_- m. \quad (74)$$

Notice that the expressions for B, C in (72)-(73) satisfy, respectively, the relevant homogeneous equations arising in (69)-(70) in the case $m = 0$.

The remaining three gauge freedoms allow us to choose the conditions

$$\partial_- \rho_1 = 0, \quad \partial_+ \rho_2 = 0, \quad m = m_0 = \text{const.} \quad (75)$$

then one has $\rho_1 \equiv \rho_1^+(x_+)$, $\rho_2 \equiv \rho_2^-(x_-)$, i.e, still remains one gauge freedom since one has the two arbitrary chiral fields ρ_1^+, ρ_2^- . The remaining gauge freedom is fixed by choosing

$$\Theta^R = 0, \quad \Theta^I \equiv \theta, \quad (76)$$

where θ is a real field and the real component of the field Θ is considered to be vanishing. Taking into account the parametrizations (72)-(73) and the above gauge conditions the eqs. (68) and (69)-(70) can be written respectively as

$$i\partial_+ \partial_- \theta - \left[f^+ g^- e^{2i\theta} + \rho_1^+ \rho_2^- m_0^2 e^{-2i\theta} \right] - [m_0^2 + m_0 f^+ \rho_2^- + m_0 g^- \rho_1^+] = 0 \quad (77)$$

$$2i\rho_1^+ \partial_- \theta = -\partial_+ e^{\Lambda_1}, \quad (78)$$

$$2i\rho_2^- \partial_+ \theta = \partial_- e^{\Lambda_2} \quad (79)$$

Notice that from (71) one has $2\Theta = \Lambda_1 + \Lambda_2$, then, due to the gauge fixing (76), in the eqs. (78)-(79) we must have

$$\Lambda_2^I = 2\theta - \Lambda_1^I, \quad \Lambda_2^R = \Lambda_1^R = 0. \quad (80)$$

Therefore the system (77)-(79) is equivalent to six real equations for the fields $\theta, \Lambda_1^I, \rho_1^+, \rho_2^-, f^+, g^-$.

4.1 sine-Gordon

The eq. (77) contains a sine-Gordon equation as a sub-model. In order to see this let us notice that the products $(\rho_1^+ \rho_2^-)$ and $(f^+ g^-)$ appearing in the terms containing the θ field in (77) are irrelevant since they can be absorbed into a redefinition of the function θ and a change of the coordinates x_+ , x_- . So, let us assume

$$f^+ = -m_0/a, \quad g^- = am_0, \quad \rho_1^+ = b, \quad \rho_2^- = 1/b, \quad a, b = \text{const.} \quad (81)$$

Then (77) becomes

$$\partial_+ \partial_- \theta + 4m_0^2 \sin(\theta - \theta_0) \cos(\theta + \theta_0) = 0, \quad \sin(2\theta_0) \equiv (1 + r - \frac{1}{r})/2, \quad r = ab. \quad (82)$$

In fact, (82) reduces to the sine-Gordon model

$$\partial_+ \partial_- \theta + \frac{2m_0^2}{\sqrt{2k}} \sin \sqrt{8k} \theta = 0, \quad (83)$$

provided that $\theta_0 = n \frac{\pi}{2}$, $n \in \mathbb{Z}$. The parameter k has been introduced by redefining $\theta \rightarrow \sqrt{2k} \theta$ according to the definitions (60), (71) and the gauge fixing (76). In fact, the usual kinetic term for φ in the lagrangian would reproduce the term $\sqrt{\frac{k}{2}} \partial_+ \theta \partial_- \theta$. The sine-Gordon coupling constant β defined by $\sin \beta \theta$ ($\beta \equiv \sqrt{8k}$) in (83) and the massive Thirring coupling g in (44) are related by

$$\beta^2 = \frac{32e\psi}{g}. \quad (84)$$

This is a manifestation of the strong/weak duality between the SG and MT models. Since the quantum equivalence between the models is established for $\beta^2 = 4\pi/(1 + g/\pi)$ we can consider (84) as a classical limit of this relationship.

Once the θ field is found as a solution of (82) the remaining field Λ_1^f can be obtained from (78)-(79) taking into account (80) and the parametrization (81). The duality relationship in the case with $m(x)$ deserves a careful treatment, a research in this direction is in progress.

5 Discussion

We have considered a Fermi field chirally coupled to a pseudoscalar field which evolves into a soliton. We have provided the both weak/strong sectors of the system as being related to the well known massive Thirring/sine-Gordon correspondence. Moreover, we have found the bound states of the fermion sector. In the intermediate region of field space in which the spinor and scalar fields interact to greater or lesser extent the system possesses a rich physical properties. This intermediate fermion-soliton system has recently been considered in the literature (for recent results see [18]).

Acknowledgements

JMJM was supported by a Fapemat grant. The authors would like to thank PRONEX-CNPq-FAPEMAT for partial financial support.

References

- [1] L.A. Ferreira, J-L. Gervais, J.Sánchez Guillen and M.V.Saveliev, *Nucl. Phys.* **B470** (1996) 236-288.
- [2] H. Blas and L.A. Ferreira, *Nucl. Phys.* **B571** (2000) 607.
- [3] H. Blas, *Nucl. Phys.* **B596** (2001) 471.
- [4] H. Blas and B.M. Pimentel. *Ann. of Phys.* **282** (2000) 67.
- [5] L.V. Belvedere, R.L.P.G. do Amaral, A.F. de Queiroz, *Phys. Lett.* **209A** (2001) 177.
- [6] D.G. Barci, L. Moriconi, *Nucl. Phys.* **B438** (1995) 522.
- [7] J. Goldstone and F. Wilzeck, *Phys. Rev. Lett.* **47** (1981) 986.
- [8] M. Stone and F. Gaitan, *Ann. of Phys.* **178** (1987) 89.
- [9] H. Blas, *Phys. Rev.* **D66** (2002) 127701.
- [10] A. Kundu, *Phys. Rev. Lett.* **99** (2007) 154101.
- [11] L. Shahkarami and S.S. Gousheh, *JHEP* **06** (2011) 116.
- [12] L. Shahkarami, A. Mohammadi and S.S. Gousheh, *JHEP* **11** (2011) 140.
- [13] P. H. Damgaard, H. B. Nielsen and R. Sollacher, *Nucl. Phys.* **B385** (1992) 227.
- [14] D.G. Barci and L. Moriconi, *Nucl. Phys.* **B438** (1995) 522.
- [15] S.S. Gousheh and R. López-Mobilial, *Nucl. Phys.* **B428** (1994) 189.
- [16] R. MacKenzie and F Wilczek, *Phys. Rev.* **D30** (1984) 2194.
- [17] R. Rajaraman, *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory*, Elsevier, North Holland (1989).
- [18] H. Blas, *An integrable fermion-soliton system: bound states and analytical solutions*, to appear.

Breaking the Λ CDM degeneracy using CMB data

A. Bernui^a

^a Instituto de Ciências Exatas
Universidade Federal de Itajubá
Av. BPS 1303, bairro Pinheirinho, Itajubá
MG - Brasil

Abstract

Recent measurements of the cosmic microwave background radiation (CMB) from the WMAP satellite led to formulate a successful concordance cosmological model, termed Λ CDM. This model satisfactorily explains the origin and structure of the CMB temperature fluctuations, from small to large angular scales, and moreover it accurately fits –with only six parameters– the CMB angular power spectrum. Despite of their triumphs in describing the observed WMAP data, we notice that some Λ CDM cosmological parameters can attain, due to their error bars, slightly different values and this degree of freedom could produce a significant impact in our understanding of the primordial universe. We are talking about the degeneracy problem, that is cosmological models with parameters that are a little bit different from those given by the Λ CDM model but fits equally well the angular power spectrum of the CMB data. Our interest here is to investigate the Gaussian statistical property, at large angular scales, in two sets of Monte Carlo CMB maps produced by seeding them with slightly different Λ CDM angular power spectra.

1 Introduction

Recent data releases from the Wilkinson Microwave Anisotropy Probe (WMAP) [1] confirm the validity of the standard cosmological model, that is, the Λ CDM. This model is obtained through the best-fitting of the cosmic microwave background radiation (CMB) angular power spectrum, as a result it provides the best values for the cosmological parameters, which of course are not sharp values but values within small intervals due to the error and uncertainty limits. For this, it is interesting to investigate synthetic CMB maps obtained from angular power spectra generated by very similar (but not equal) cosmological parameters that fit the WMAP data as well as the Λ CDM model. One such parameter is the spectral index of primordial fluctuations n_s , which is fundamental for our understanding of the physics of the primordial universe. The main difficulty to establish the correct spectral index value is that it determines the form of the angular power spectra at the largest scales (i.e. low multipoles), precisely the region where the cosmic variance uncertainty is large and dominates other effects.

The importance to determine, with the highest degree of accuracy, the value of the spectral index n_s is because it is related to physical processes of the primordial universe (remember that the CMB radiation is -with present techniques- the oldest cosmological observable). And because such processes probably left imprints in the statistical attributes of the observed CMB, it deserves a meticulous analysis the relationship between the spectral index and the Gaussian properties of the CMB temperature field. For instance, the value $n_s = 1$ means that the CMB temperature fluctuations correspond to a statistically Gaussian random field, therefore small variations from this value should be correlated with small deviations of Gaussianity in the data.

For these reasons, besides the determination of the spectral index n_s by fitting the CMB angular power spectrum, the released data from the WMAP [1] is nowadays under rigorous scrutiny for the possible deviations from Gaussianity [2] and from statistical isotropy [3] in the CMB data, properties that could be, in some sense, related. Moreover, a detection or non-detection of primordial non-Gaussianity in the CMB data is crucial to discriminate among inflationary models, and to test alternative scenarios of the early universe [4].

We shall investigate the effects caused on the statistical properties of two sets of Monte Carlo CMB maps produced by seeding them with slightly different angular power spectra, each one obtained using the same cosmological parameters as the Λ CDM model except that we consider different spectral indexes values: $n_s = 0.96$ and 1.00 , respectively. To detect the presence of Gaussian deviations in these CMB maps we use two recently published non-Gaussian indicators [5], based on the application of skewness and kurtosis in large-angle patches of CMB temperature maps, which allow us to construct the **S**- and **K**-maps (for details see section 3) from which we calculate their angular power spectra (APS). After that, we compare the APS of the two sets of **S**-maps obtained by applying our skewness tool to both sets of Monte Carlo CMB maps. Similar procedures are done with the two sets of **K**-maps. Our results show that, in the mean, the Gaussianity property of these sets of Monte Carlo maps is different, and this fact seems to be crucial when one has to quantify the confidence level of some data analysis result. That is to say, the statistical-significance evaluation of a result concerning a WMAP CMB map is model dependent.

This work is organized as follows. In Section 2 we explain how to produce the sets of simulated Monte Carlo CMB maps based on the Λ CDM angular power spectrum, and also the ILC WMAP map, used in the analyses. In Section 3 we describe our method to analyze the statistical properties of the CMB maps. Finally, in Section 4 we summarize our results and formulate our conclusions and perspectives for future works.

2 Λ CDM Angular Power Spectra and Monte Carlo CMB data

The degeneracy problem in the concordance cosmological model consists on having two or more APS similar to that one of the Λ CDM model obtained with slightly different cosmological parameter value(s), but all of them consistent with current data because of the error bars.

Here we shall study two Λ CDM APS with the spectral indexes attaining slightly different values: $n_s = 0.96, 1.00$. For this, we use the CMBFAST tool [7] to generate the corresponding APS seeds that are then used in the SYNFAST algorithm [8] to produce two sets of 1000 random realizations each. As we show in Fig. 1, left panel, the Λ CDM APS corresponding to $n_s = 0.96$ is very similar to the APS obtained with the same cosmological parameters as the concordance model, except that $n_s = 1.00$. Moreover, we also observe in the right panel that both APS fits suitably the APS from WMAP data. The comparison of the statistical properties between these two sets of Monte Carlo CMB maps constitute the target of our analysis.

It is well known that foregrounds introduce non-Gaussianities in the CMB data. For these the WMAP science team have performed substantial efforts to minimize the presence of foregrounds in the measured CMB data. The recently released seven-year foreground-reduced ILC map represents well a CMB map that is consistent with Gaussianity, after the application of the KQ75-7yr mask [1]. We find interesting to use this WMAP map to test the mean statistical properties of these two sets of Monte Carlo CMB maps.

In the next section we shall describe our method to analyze the statistical features of the CMB maps, both Monte Carlo and WMAP ones. The comparison of the statistical properties between these two sets of Monte Carlo CMB maps, together with the analysis of the ILC map, will be performed in the last section.

3 The Method For Statistical Analysis

In this section we briefly present how to construct two non-Gaussianity indicators. For a detailed discussion of these indicators we refer the readers to ref. [5].

Consider a discrete set of points $\{j = 1, \dots, N_c\}$ uniformly distributed on the sphere S^2 as the center of spherical caps of a given aperture γ , and calculate for each cap j the skewness and kurtosis values

$$S_j \equiv \frac{1}{N_p \sigma_j^3} \sum_{i=1}^{N_p} (T_i - \bar{T}_j)^3, \quad K_j \equiv \frac{1}{N_p \sigma_j^4} \sum_{i=1}^{N_p} (T_i - \bar{T}_j)^4 - 3,$$

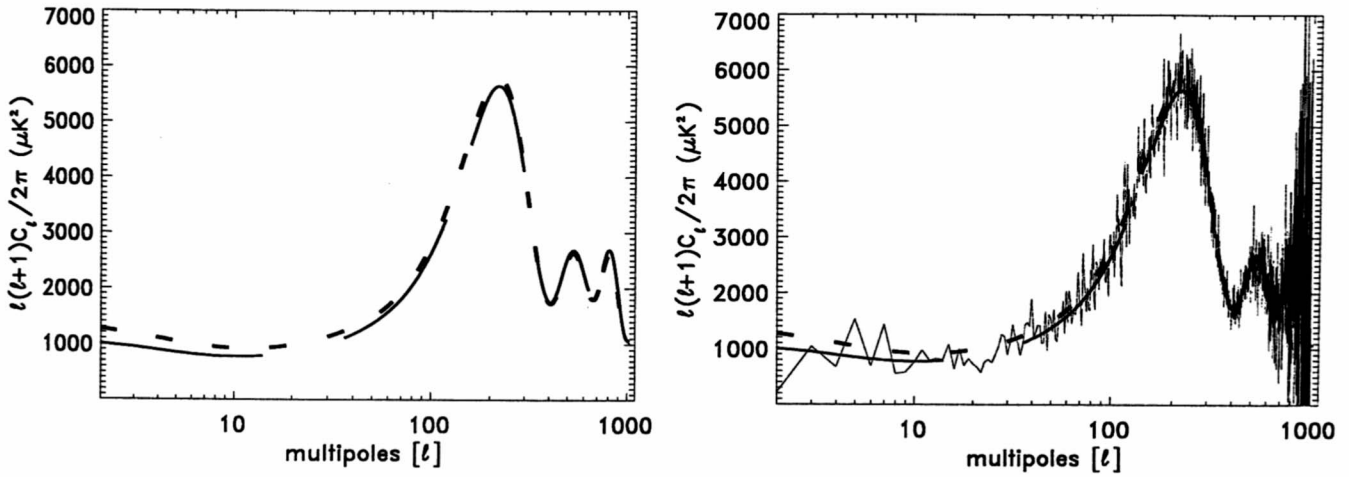


Figure 1: (Left panel) Angular power spectra for Λ CDM models with slightly different values for the spectral index: $n_s = 0.96$ (dotted curve) and $n_s = 1.00$ (dashed curve). (Right panel) The same angular power spectra appearing in the left panel plus the angular power spectrum from WMAP-7yr data (continuous noisy curve).

where N_p is the number of pixels in the j^{th} cap, T_i is the temperature at the i^{th} pixel, \bar{T}_j is the CMB mean temperature on the j^{th} cap, and σ_j is the standard deviation for each j . The sets of values $\{S_j, j = 1, \dots, N_c\}$ and $\{K_j, j = 1, \dots, N_c\}$ can then be viewed as a measure of the non-Gaussianity in the direction (θ_j, ϕ_j) (which corresponds to the center of the j^{th} cap). Patching together the $\{S_j\}$ and $\{K_j\}$ values, we obtain the functions $S = S(\theta, \phi)$ and $K = K(\theta, \phi)$ defined on S^2 . Throughout this work the sky map representation of the functions $S = S(\theta, \phi)$ and $K = K(\theta, \phi)$ are named, respectively, **S**-map and **K**-map. Clearly, one can expand each of these functions in their spherical harmonics and calculate their corresponding angular power spectrum. For the skewness function $S = S(\theta, \phi) = \{S_j, j = 1, \dots, N_c\}$, for example, one has

$$S(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} b_{\ell m} Y_{\ell m}(\theta, \phi) \implies S_{\ell} = \frac{1}{2\ell+1} \sum_m |b_{\ell m}|^2,$$

where S_{ℓ} is the corresponding angular power spectrum. Similar expressions hold for the $K = K(\theta, \phi)$.

Thus, to obtain quantitative information for each multipole component ℓ about the non-Gaussianity at large-angles (low ℓ) from the **S**- and **K**-maps, which were obtained from the CMB temperature maps, we calculate their power spectra S_{ℓ} and K_{ℓ} . Finally, to compare the statistical properties of the spectra $\{S_{\ell}\}$ and $\{K_{\ell}\}$, obtained from the analyses of the different Monte Carlo CMB sets, we compare their corresponding mean \bar{S}_{ℓ} and \bar{K}_{ℓ} and 2-sigma values (i.e., 95% CL). Here we concentrate on the large angular scales, that is, for $\ell = 1 - 10$.

4 Results and Conclusions

In this section we report the results of Gaussianity analyses performed by applying our indicators, discussed in the previous section, to the Monte Carlo and WMAP data described in section 2.

In Fig. 2 we show the APS of the **S**–maps and **K**–maps calculated from the Monte Carlo CMB maps, which were produced by the APS with slightly different spectral indexes, namely $n_s = 0.96$ and 1.00 . As expected, the similarity in the CMB APS implies in the similarity of the angular power spectra obtained from the **S**– and **K**–maps. We also exhibit in these plots the APS from the **S**– and **K**–maps obtained from the ILC-7yr WMAP map (with the corresponding mask to ensure Gaussianity properties, see section 2).

Our main conclusion is that, as observed in Fig. 2, Monte Carlos generated through APS with a different spectral index n_s have different statistical properties and therefore they show different confidence level when used to evaluate the statistical significance of some result obtained from WMAP data. For instance, the confidence level of the APS $\{S_\ell\}$ and $\{K_\ell\}$ for the WMAP ILC-7yr foreground-reduced map reveal that it is more likely non-Gaussian ($n_s = 0.96$) than Gaussian ($n_s = 1.00$). In order to quantify the possible correlation between data from WMAP and data from the sets of Monte Carlo maps we use the reduced χ^2 -goodness-of-fit test which gives an overall measure between two data sets by quantifying how related they are. For this we shall calculate the value $\chi^2/dof \equiv \chi^2_{red}$, where dof stands for degree of freedom, between pairs of data sets regarding the APS values S_ℓ and K_ℓ , for $\ell = 1 - 10$, calculated from the ILC map versus the mean-APS \bar{S}_ℓ and \bar{K}_ℓ obtained from the two sets of Monte Carlo CMB maps, those produced with $n_s = 0.96$ and $n_s = 1.00$ (see Fig. 2).

Our results are: $\chi^2_{red} = 1.8(1.3)$ between the APS of the **S**–map (**K**–map) from the ILC versus the mean APS of the **S**–maps (**K**–maps) obtained from Monte Carlo CMB maps with $n_s = 0.96$. Similarly, we obtain $\chi^2_{red} = 4.7(4.3)$ between the APS of the **S**–map (**K**–map) from the ILC versus the mean APS of the **S**–maps (**K**–maps) obtained from Monte Carlo CMB maps with $n_s = 1.00$. Remember that a χ^2 value near 1 means a high correlation between analyzed data, while larger values means poorer correlations. Thus, our results indicate that is more probable that (the $\{S_\ell\}$ and $\{K_\ell\}$ APS from) the WMAP ILC map plus the KQ75-7yr mask corresponds to the case $n_s = 0.96$, with a p –value of $\sim 15\%$, than to the Gaussian case $n_s = 1.00$, with a p –value of less than 0.1% .

Summarizing, our APS analyses of **S**– and **K**–maps produced from the two sets of Monte Carlo CMB maps (seeded by Λ CDM spectra with $n_s = 0.96, 1.00$, respectively) show that, in the average, these data contain different amounts of non-Gaussianity. Furthermore, Monte Carlos seeded on the Λ CDM concordance model spectrum ($n_s = 0.96$) can not be considered as the absolute reference for *Gaussian* properties in CMB maps. Additionally, our results show that, in the average, the Gaussianity property of these sets of Monte Carlo maps is different, and this fact seems to be crucial when one has to quantify the confidence level of some result. That is to say, *the statistical-significance evaluation of any result concerning WMAP data is model dependent*.

It would be interesting to test the Gaussian properties of these two sets of Monte Carlo CMB maps using some statistical tools available in the literature.

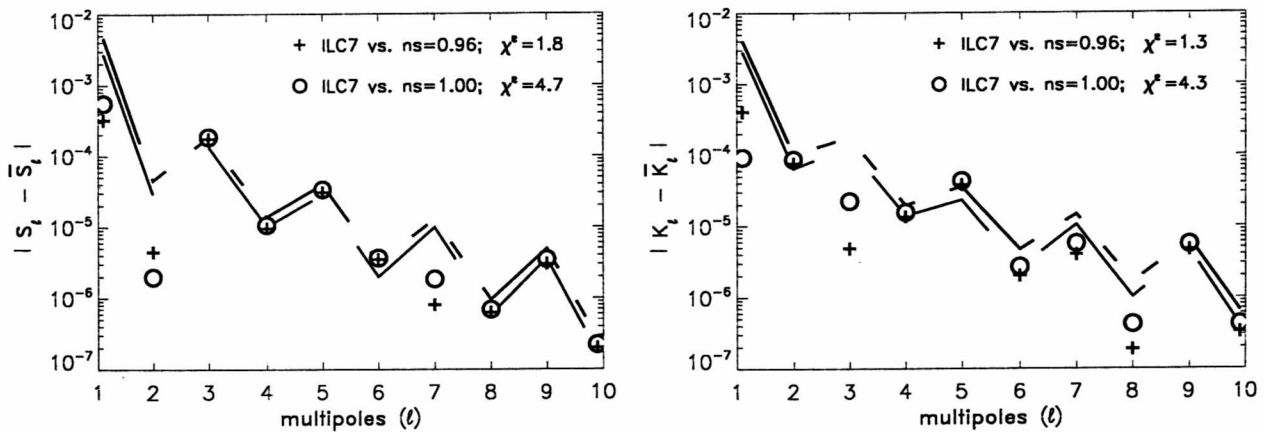


Figure 2: Angular Power Spectra for the **S**– and **K**–maps, left and right panel, respectively. The dot-dashed (dashed) line corresponds to the 2σ level, that is, 95% confidence level of the **S**– and **K**–map’s power spectra, maps obtained by applying our statistical tools to 1 000 Monte Carlo CMB maps produced from a Λ CDM model with spectral index $n_s = 0.96$ ($n_s = 1.00$).

5 Acknowledgments

We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA) [1]. Some of the results in this paper were derived using the HEALPix package [8]. This work was partially supported by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) – Brasil, under grant APQ-01893-10. We also thank the fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Brasil.

References

- [1] N. Jarosik *et al.*, arXiv:1001.4744 [astro-ph.CO];
D. Larson *et al.*, arXiv:1001.4635 [astro-ph.CO];
<http://lambda.gsfc.nasa.gov/>
- [2] M. Cruz *et al.*, *New Astron. Rev.* **50**, 880 (2006); A. Bernui, C. Tsallis and T. Villela, *Europhys. Lett.* **78**, 19001 (2007); *Phys. Lett. A* **356**, 426 (2006); P. D. Naselsky *et al.*, arXiv:0712.1118 [astro-ph]; C. Monteserín *et al.*, *Mon. Not. R. Astron. Soc.* **387**, 209 (2008); B. Lew, *JCAP* **08**, 017 (2008); Y. Wiaux *et al.*, *Mon. Not. Roy. Astron. Soc.* **385**, 939 (2008); J. D. McEwen *et al.*, *Mon. Not. R. Astron. Soc.* **388**, 659 (2008); M. Kawasaki *et al.*, *JCAP* **11**, 019 (2008); C. Raeth *et al.*, arXiv:0810.3805 [astro-ph]; G. Rossmannith, C. Raeth, A. J. Banday and G. Morfill, arXiv:0905.2854 [astro-ph.CO]; P. Vielva and J. L. Sanz, *Mon. Not. R. Astron. Soc.* **397**, 837 (2009); M. Kawasaki, K. Nakayama and F. Takahashi, *JCAP* **01**, 026 (2009); M. Kawasaki *et al.*, *JCAP* **01**,

- 042 (2009); D. Pietrobon *et al.*, *Mon. Not. Roy. Astron. Soc.* **396**, 1682 (2009); M. Cruz, E. Martínez-González and P. Vielva, arXiv:0901.1986 [astro-ph]; P. Cabella *et al.*, arXiv:0910.4362 [astro-ph.CO]; A. Bernui and M. J. Rebouças, *Int. J. Mod. Phys. A* **24**, 1664 (2009); *Int. J. Mod. Phys. D* **19**, 1411 (2010); D. Pietrobon *et al.*, *Mon. Not. Roy. Astron. Soc.* **402**, L34 (2010); Y. Ayaita, M. Weber and C. Wetterich, *Phys. Rev. D* **81**, 023507 (2010); C. Raeth *et al.*, arXiv:1012.2985 [astro-ph.CO]; A. P. S. Yadav and B. D. Wandelt, arXiv:1006.0275 [astro-ph.CO]; C. Raeth, G. Rossmanith, G. Morfill, A. J. Banday and K. M. Górski, arXiv:1005.2481 [astro-ph.CO]; N. Bartolo, S. Matarrese, O. Pantano and A. Riotto, *Class. Quant. Grav.* **27**, 124009 (2010); N. Bartolo, S. Matarrese and A. Riotto, arXiv:1001.3957 [astro-ph.CO]; A. Bernui, M. J. Rebouças and A. F. F. Teixeira, *Int. J. Mod. Phys. D* **19**, 1405 (2010); R. Saha, arXiv:1105.6298 [astro-ph.CO].
- [3] J. D. McEwen, M. P. Hobson, A. N. Lasenby and D. J. Mortlock, *Mon. Not. Roy. Astron. Soc.* **371**, L50 (2006); R. J. Adler, J. D. Bjorken and J. M. Overduin, gr-qc/0602102; L. R. Abramo *et al.*, *Phys. Rev. D* **74**, 063506 (2006); A. Bernui *et al.*, *Astron. Astrophys.* **454**, 409 (2006); P. Vielva, Y. Wiaux, E. Martínez-González and P. Vandergheynst, *Mon. Not. Roy. Astron. Soc.* **381**, 932 (2007); A. Gruppuso *et al.*, arXiv:1006.1979 [astro-ph.CO]; A. Gruppuso, *Phys. Rev. D* **76**, 083010 (2007); T. Ghosh, A. Hajian and T. Souradeep, *Phys. Rev. D* **75**, 083007 (2007); A. Bernui, B. Mota, M. J. Rebouças and R. Tavakol, *Astron. Astrophys.* **464**, 479 (2007); *Int. J. Mod. Phys. D* **16**, 411 (2007); C. Gordon, *Astrophys. J.* **656**, 636 (2007); B. Lew, *JCAP* **09**, 023 (2008); P. K. Samal, R. Saha, P. Jain and J. P. Ralston, *Mon. Not. R. Astron. Soc.* **385**, 1718 (2008); T. Koivisto and D. F. Mota, *JCAP* **06**, 018 (2008); C. Monteserin *et al.*, *Mon. Not. R. Astron. Soc.* **387**, 209 (2008); T. Kahniashvili, G. Lavrelashvili and B. Ratra, *Phys. Rev. D* **78**, 063012 (2008); A. Bernui and W. S. Hipólito-Ricaldi, *Mon. Not. R. Astron. Soc.* **389**, 1453 (2008); P. Vielva and J. L. Sanz, arXiv:0812.1756 [astro-ph]; A. Bernui, *Phys. Rev. D* **78**, 063531 (2008); P. K. Samal, R. Saha, P. Jain and J. P. Ralston, *Mon. Not. R. Astron. Soc.* **396**, 511 (2009); C. M. Hirata, *JCAP* **09**, 011 (2009); C. Caprini, F. Finelli, D. Paoletti and A. Riotto, *JCAP* **06**, 021 (2009); J. Kim and P. Naselsky, *JCAP* **07**, 041 (2009); I. Y. Areféva, N. V. Bulatov, L. V. Joukovskaya and S. Y. Vernov, *Phys. Rev. D* **80**, 083532 (2009); *Phys. Rev. D* **80**, 123010 (2009); Y. Shtanov and H. Pyatkovska, *Phys. Rev. D* **80**, 023521 (2009); G. P. Holder, K. M. Nollett and A. van Engelen, arXiv:0907.3919 [astro-ph.CO]; L. R. Abramo, A. Bernui and T. S. Pereira, *JCAP* **12**, 013 (2009); A. L. Erickcek, C. M. Hirata and M. Kamionkowski, *Phys. Rev. D* **80**, 083507 (2009); J. F. Donoghue, K. Dutta and A. Ross, *Phys. Rev. D* **80**, 023526 (2009); N. Joshi, S. Jhingan, T. Souradeep and A. Hajian, arXiv:0912.3217 [astro-ph.CO]; T. R. Seshadri and K. Subramanian, *Phys. Rev. Lett.* **103**, 081303 (2009); T.-P. Li, H. Liu, L.-M. Song, S.-L. Xiong and J.-Y. Nie, arXiv:0905.0075 [astro-ph.CO]; M. Frommert and T. A. Ensslin, arXiv:0908.0453 [astro-ph.CO]; L. R. Abramo and T. S. Pereira, arXiv:1002.3173 [astro-ph.CO]; H.-C. Kim and M. Minamitsuji, arXiv:1002.1361 [gr-qc]; A. Gruppuso and K. M. Górski, *JCAP* **03**, 019 (2010); D. G. Yamazaki, K. Ichiki, T. Kajino and G. J. Mathews, *Phys. Rev. D* **81**, 023008 (2010); A. R. Pullen and C. M. Hirata, arXiv:1003.0673 [astro-ph.CO]; S.-Y.

- Zhong and X. Wu, *Phys. Rev. D* **81**, 104037 (2010); S.-Y. Zhong, X. Wu, S.-Q. Liu, X.-F. Deng, *Phys. Rev. D* **82**, 124040 (2010); M. Cruz, P. Vielva, E. Martínez-González and R. B. Barreiro, arXiv:1005.1264 [astro-ph.CO]; F. Paci *et al.*, *Mon. Not. Roy. Astron. Soc.* **407**, 399 (2010); T. Kahniashvili, A. Brandenburg, A. G. Tevzadze and B. Ratra, *Phys. Rev. D* **81**, 123002 (2010); D. Pietrobon, A. Balbi, P. Cabella and K. M. Górski, *Astrophys. J.* **723**, 1 (2010); R.-G. Cai, B. Hu and H.-B. Zhang, *JCAP* **08**, 025 (2010); F. R. Urban and A. R. Zhitnitsky, arXiv:1011.2425 [astro-ph.CO]; P. K. Aluri, P. K. Samal, P. Jain and J. P. Ralston, arXiv:1007.1827 [astro-ph.CO]; C. J. Copi, D. Huterer, D. J. Schwarz and G. D. Starkman, *Adv. Astron.* **2010**, 847541 (2010); I. Ya. Areféva, N. V. Bulatov and S. Yu. Vernov, *Theor. Math. Phys.* **163**, 788 (2010); N. Bartolo, M. Fasiello, S. Matarrese and A. Riotto, *JCAP* **12**, 026 (2010). H.-C. Kim and M. Minamitsuji, *Phys. Rev. D* **81**, 083517 (2010); Erratum-ibid. *Phys. Rev. D* **82**, 109904 (2010); *JCAP* **03**, 038 (2011); H. Wei, *JCAP* **04**, 022 (2011); N. Afshordi, A. Slosar and Y. Wang, *JCAP* **01**, 019 (2011); J.-Q. Xia *et al.*, arXiv:1104.5015 [astro-ph.CO];
- [4] N. Bartolo *et al.*, *Phys. Rep.* **402**, 103 (2004); E. Komatsu *et al.*, arXiv:0902.4759.
- [5] A. Bernui and M. J. Rebouças, *Phys. Rev. D* **79**, 063528 (2009); *Phys. Rev. D* **81**, 063533 (2010);
- [6] F. Elsner and B. D. Wandelt, *Astrophys. J. Suppl. Ser.* **184**, 264 (2009).
- [7] http://lambda.gsfc.nasa.gov/toolbox/tb_cmbfast_ov.cfm
http://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm
- [8] K. M. Górski *et al.*, *Astrophys. J.* **622**, 759 (2005).

Generación de soluciones exactas de las Ecuaciones de Einstein con ayuda de euclidones

Páuyac Huamán, J. Aníbal

1. Introducción

- La Teoría General de la Relatividad (Einstein, 1915) es una teoría del campo gravitatorio.
- Para conocer la estructura del espacio-tiempo tetra-dimensional hay que resolver las ecuaciones de Einstein.
- Un área de investigación es la búsqueda de soluciones exactas de las ecuaciones de Einstein axialmente simétricas.
- Dos clases de soluciones axialmente simétricas:
 - estáticas
 - estacionarias.
- Algunas soluciones que han encontrado una interpretación física definida: Schwarzschild, Kerr .

2. Ecuaciones fundamentales

Ecuaciones de Einstein:

$$R_{ik} = \frac{8\pi k}{c^4} \left(T_{ik} - \frac{1}{2} g_{ik} T \right)$$

donde R_{ik} es el tensor de Ricci
 T_{ik} es el tensor de energía-impulso
 g_{ik} es el tensor métrico
 k Es la constante de gravitación universal
 $T = T_i^i$

Métrica:

$$ds^2 = g_{ik} dx^i dx^k \quad (2)$$

La magnitud fundamental que caracteriza la curvatura de un espacio cuatridimensional es el tensor de Riemann

$$R_{iklm} = g_{n\rho} (\Gamma_{kl}^n \Gamma_{im}^\rho - \Gamma_{km}^n \Gamma_{il}^\rho) + \frac{1}{2} (g_{im,kl} + g_{kl,im} - g_{il,km} - g_{km,il})$$

donde Γ_{kl}^i se llaman Símbolos de Christoffel

$$\Gamma_{kl}^i = \frac{1}{2} g^{im} (g_{mk,l} + g_{ml,k} - g_{kl,m})$$

El tensor de Ricci

$$R_{\sigma\nu} = R^\rho_{\sigma\rho\nu} = \partial_\rho \Gamma_{\nu\sigma}^\rho - \partial_\nu \Gamma_{\rho\sigma}^\rho + \Gamma_{\rho\lambda}^\rho \Gamma_{\nu\sigma}^\lambda - \Gamma_{\nu\lambda}^\rho \Gamma_{\rho\sigma}^\lambda$$

En el vacío (en el exterior de un cuerpo):

$$R_{ik} = 0$$

Soluciones axialmente simétricas estáticas

Cuadrado del intervalo métrico, en coordenadas canónicas de Weyl

$$ds^2 = f^{-1} [e^{2\gamma} (d\rho^2 + dz^2) + \rho^2 d\varphi^2] - f dt^2$$

Ecuaciones del campo

$$\begin{aligned} f\Delta f &= \nabla f \nabla f \\ 2\gamma_{,z} &= \rho f^{-2} f_{,\rho} f_{,z} \\ 4\gamma_{,\rho} &= \rho f^{-2} (f_{,\rho}^2 - f_{,z}^2) \end{aligned}$$

Cambio

$$f = e^{2\Psi}$$

Entonces

$$\Delta\Psi = 0$$

Método de Waylen

$$\Psi(\rho, z) = \frac{1}{\pi} \int_0^\pi \left\{ C_1(u) + C_2(u) \ln \left[\frac{\rho}{\rho_0} \sin^2 \theta \right] \right\} d\theta, \quad u = z + i\rho \cos \theta$$

$$C_1(u) = -\frac{m}{z + i\rho \cos \theta}, \quad C_2 = 0$$

$$\Psi = -\frac{m}{\sqrt{\rho^2 + z^2}}$$

3. El euclídon

Método de las fuentes singulares

$$\Delta\Psi = -4\pi\sigma(\rho, z)$$

$$\Psi = \frac{1}{4\pi} \int_{V'} \frac{\sigma(\rho', z') dV'}{|\vec{r} - \vec{r}'|}$$

$$\sigma(\rho', z') = \frac{\delta(\rho' - \rho_0)}{\rho'} \sigma(\rho_0, z')$$

$$\rho_0 = 0, \quad \sigma(\rho_0, z') = \sigma_0(z') \Theta(z')$$

$$\Theta(z') = \begin{cases} 0, & m < z' < -m \\ 1, & -m < z' < m \end{cases}$$

$$\Psi(\rho, z) = \frac{1}{2} \int_{-m}^m \frac{\sigma_0(z') dz'}{\sqrt{\rho^2 + (z - z')^2}}$$

a. Si $\sigma_0(z') = 1$ entonces

$$\Psi(\rho, z) = \frac{1}{2} \ln \frac{z - m + \sqrt{\rho^2 + (z - m)^2}}{z + m + \sqrt{\rho^2 + (z + m)^2}}$$

- b. Si $\sigma_0(z') = \delta_0 = \text{const}$ entonces $f = \left(1 - \frac{2m}{r}\right)^{\delta_0}$
- c. Si $\sigma_0(z') = \delta(z')/2$ entonces $\Psi(\rho, z) = -\frac{m}{\sqrt{\rho^2 + z^2}}$
- d. Si $\sigma_0(z') = \begin{cases} 1/2, & z' > 0 \\ -1/2, & z' < 0 \end{cases}$

entonces obtenemos el EUCLIDON

$$\Psi = \frac{1}{2} \ln \left[z + W + \sqrt{\rho^2 + (z + W)^2} \right]$$

4. Generación de nuevas soluciones con ayuda de euclídones

Mediante la combinación lineal de euclídones se pueden generar nuevas soluciones

$$\Psi = \sum C_k \Psi_k, \quad C_k = \pm 1, \quad k = 1, \dots, 2N$$

De acuerdo a si el desplazamiento es + o - y al valor de C_k , se obtienen diferentes resultados.

Soluciones bi-euclídónicas

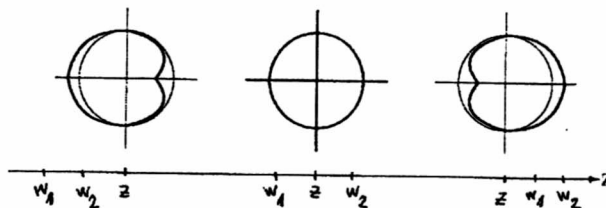
Combinación de 2 euclídones desplazados a la izquierda, con $C_1=1$ y $C_2=-1$

$$\Psi = \frac{1}{2} \ln \left[\frac{z + W_1 + \sqrt{\rho^2 + (z + W_1)^2}}{z + W_2 + \sqrt{\rho^2 + (z + W_2)^2}} \right]$$

Así, podemos escribir las otras combinaciones.

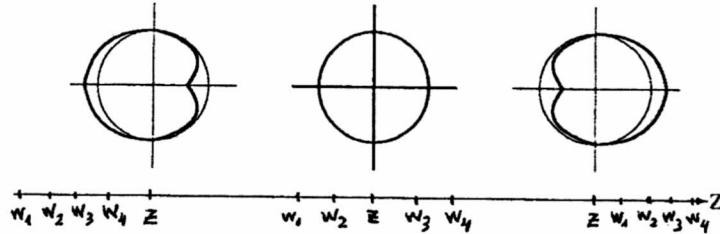
No todas las combinaciones son soluciones físicamente aceptables. Sólo las que a distancias grandes tienden a un espacio plano

En la combinación mostrada en el centro, si $C_1=-1$, $C_2=+1$, y además $W_1=W_2$, entonces se obtiene la solución de Schwarzschild.

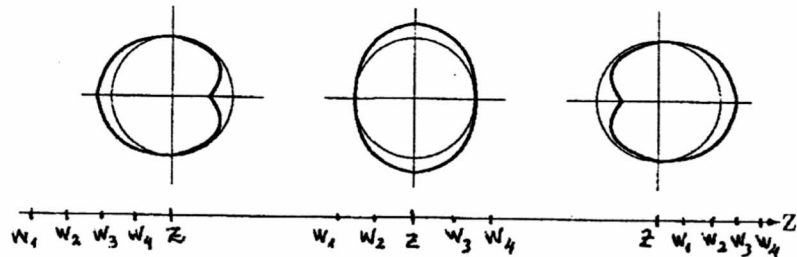


Soluciones tetra-euclídónicas

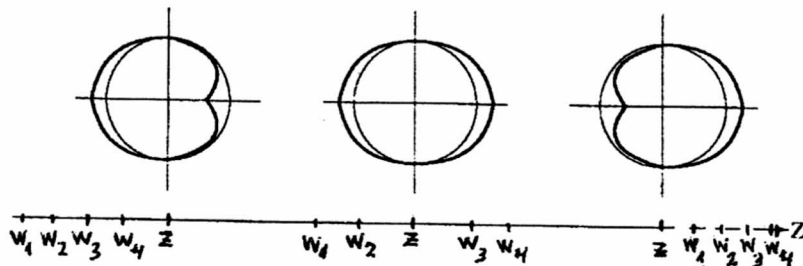
a)
$$\Psi = \frac{1}{2} \ln \frac{\left[z + W_3 + \sqrt{\rho^2 + (z + W_3)^2} \right] \left[z + W_4 + \sqrt{\rho^2 + (z + W_4)^2} \right]}{\left[z + W_1 + \sqrt{\rho^2 + (z + W_1)^2} \right] \left[z + W_2 + \sqrt{\rho^2 + (z + W_2)^2} \right]}$$



b)
$$\Psi = \frac{1}{2} \ln \frac{\left[z + W_2 + \sqrt{\rho^2 + (z + W_2)^2} \right] \left[z + W_4 + \sqrt{\rho^2 + (z + W_4)^2} \right]}{\left[z + W_1 + \sqrt{\rho^2 + (z + W_1)^2} \right] \left[z + W_3 + \sqrt{\rho^2 + (z + W_3)^2} \right]}$$



c)
$$\Psi = \frac{1}{2} \ln \frac{\left[z + W_2 + \sqrt{\rho^2 + (z + W_2)^2} \right] \left[z + W_3 + \sqrt{\rho^2 + (z + W_3)^2} \right]}{\left[z + W_1 + \sqrt{\rho^2 + (z + W_1)^2} \right] \left[z + W_4 + \sqrt{\rho^2 + (z + W_4)^2} \right]}$$



Conclusion

De esta manera, combinando euclidones se obtienen nuevas soluciones, las cuales deben ser analizadas e interpretadas físicamente.

A note from the Editors:

This talk was based on the authors PhD Thesis. So, for additional references we suggest to contact the author to the email: jpauyac@hotmail.com

OA-UNI: an Astronomical Observatory at UNI

A. Pereyra^a, W. Cori^b, E. Meza^b, J. Ricra^b, and G. Granda^b

^a Instituto de Astrofísica de Canarias (IAC)

C/ Vía Láctea, s/n, La Laguna

Tenerife - Spain

^b Grupo Astronomía

Facultad de Ciencias

Universidad Nacional de Ingeniería

Av. Túpac Amaru, s/n, Rimac

Lima - Perú

Abstract

We summarized the status of an ongoing project which aims to implement an astronomical observatory for the Faculty of Science of the National University of Engineering (FC-UNI). This observatory (OA-UNI) will be installed at the peruvian central Andes using the facilities of the Instituto Geofísico del Peru at Huancayo. We shown examples of science projects that could be initiated along with the future perspectives for its upgrade. We also comment the impact of the OA-UNI to open new possibilities to develop pre-professionals training in astronomy for last-year students at FC-UNI.

1 Introduction

The project of an astronomical observatory for the National University of Engineering (OA-UNI project) started in early 2009. The enthusiasm and strong collaboration of the members of the Astronomy Group (GA) at FC-UNI was fundamental to this purpose. Recently, the acquisition of a high quality optical tube of a 0.5m Ritchey-Chrétien telescope was completed at FC-UNI with local funds. The implementation of OA-UNI includes the construction of a dome building along with the acquisition of a robotic astronomical mount, CCD detectors and a spectrograph. OA-UNI will be located at the Peruvian Andes mountains and will let to develop the astronomy in undergraduate and graduate levels at FC-UNI. Astronomical projects could be feasible with a special focus in site testing, stellar photometry including mid and long term variability, spectroscopy and polarimetry. Astronomical outreach also will benefit with OA-UNI.

2 Antecedents

We began in 2001 a program to measure the quality of the sky (seeing) for astronomical observations at the site of the Huancayo's Observatory (OH, 3300 m.a.s.l, central Peruvian Andes, 300km from Lima). This site has a full structure (accommodations, electric power, etc.) to let several astronomical observational runs per year. Historically, the OH was used to collect measurements of the earth magnetic field. Our program of astronomical observations used small telescopes with the appropriate adaptations. Since then this site was routinely used by the members of GA to develop astronomical undergraduate programs. The program of site testing will be extended to other sites in the future and will be improved by the implementation of OA-UNI. The previous work of members of GA in this and other programs yielded several publications ([1, 2, 3, 4]).

The philosophy of the OA-UNI project is driven by the following aims: (1) to implement an astronomical observatory which contributes to the development of the astronomy in Peru; (2) to strengthen the UNI in Peru, as an institution with research in astronomy; and, (3) to promote the inclusion and/or collaboration between professionals, students and researchers of several areas in science/engineering to apply modern technologies to the development of the astronomical observatory.

We consider important to appoint that the OA-UNI project is a joint effort between two institutions dedicated to the development of astronomy in Peru: FC-UNI and the Astronomy Area at IGP. We consider that this collaboration must be preserved in time.

3 Science Projects

The professional astronomy in Peru is giving its first steps and it is necessary to prepare to new potential astronomers with a solid background. Observational projects with small telescopes are the first interaction of the students with the astronomical techniques. In particular, the experience in the site testing program will have as final aim to select the best place to install the first professional astronomical observatory in Peru.

The science cases where we believe that OA-UNI will be competitive include:

- (a) To detect extrasolar planets using photometric methods.
- (b) To study light curves of variables stars of short and middle period.
- (c) To implement the astronomical polarimetry technique to investigate the interstellar medium.
- (d) To continue the site testing program in potential astronomical sites.
- (e) To develop the Differential Imaging Motion Monitor (DIMM) technique to monitor the influence of atmospheric turbulence in astronomical imaging quality.
- (f) To promote the teaching/outreach of astronomy in schools, institutes and other universities.

4 The next steps

As we mentioned above, in this phase of the project, the OA-UNI will be located in the installations of IGP at Huancayo. Later studies of site testing, inside peruvian territory, will let to find the best site for astronomical observations. The portability of OA-UNI will be extremely useful for this program. In the middle term, OA-UNI will be relocated to this new site. Actually, we are focused to the acquisition of the remaining equipments for the OA-UNI, including the robotic mount and dome. When this phase was completed, we will be ready to begin the first observations of science cases.

The possibility of develop astronomy programs at FC-UNI will have a positive impact for the physics students. In this sense, we believe that an important number of Physics thesis in astronomy could be carried out helped by data gathered with OA-UNI. The acquired observational experience at OA-UNI also will be an additional plus when postgraduate studies in astronomy will be the natural next step in the academic life of a graduate student.

5 Acknowledgments

AP thanks to Dr. Humberto Asmat, Dr. Abel Gutarra and Dr. S. Petrick (from FC-UNI) along with Dr. José Ishitsuka (IGP) and Dr. A. Bernui (UNIVAP/Brazil) for their support to this project. We also thank the financial support of the UNI's Rectorate.

References

- [1] Pereyra, A. & Baella, N., 2003, *Medidas de Seeing en el Observatorio de Huancayo*, REVCUNI, 7, 1, 103
- [2] Baella, N. & Pereyra, A., 2003, *Calibraciones Fotométricas en el Observatorio de Huancayo*, TECNIA, 13, 1, 15 (UNI, Lima-Perú)
- [3] Dalmau, A. & Pereyra, A., 2004, *Medidas fotométricas en el Observatorio de Huancayo*, TECNIA, 14, 1, 49 (UNI, Lima-Perú)
- [4] Meza E. & Pereyra, A., 2005, *Linealidad de un CCD para uso astronómico*, Memorias de la Segunda Reunión Peruana de Óptica, PUC, Lima Perú, p. 69

Generalized AKNS System, Non-vanishing Boundary Conditions and N-Dark-Dark Solitons

A. de O. Assunção^a, H. Blas^a and M. J. B. F. da Silva^b

^a Instituto de Física
Universidade Federal de Mato Grosso
Av. Fernando Correa, s/n, Coxipó
78060-900, Cuiabá - MT - Brazil

^b Departamento de Matemática
Universidade Federal de Mato Grosso
Av. Fernando Correa, s/n, Coxipó
78060-900, Cuiabá - MT - Brazil

Abstract

We consider certain boundary conditions supporting soliton solutions in the generalized non-linear Schrödinger equation (AKNS). Using the dressing transformation (DT) method and the related tau functions we study the AKNS_r system for the vanishing, (constant) non-vanishing and the mixed boundary conditions, and their associated bright, dark and bright-dark N-soliton solutions, respectively. Moreover, we introduce a modified DT related to the dressing group in order to consider the free field boundary condition and derive generalized N-dark-dark solitons. We have shown that two-dark-dark-soliton bound states exist in the AKNS₂ system, and three- and higher-dark-dark-soliton bound states can not exist. As a reduced submodel of the AKNS_r system we study the properties of the focusing, defocusing and mixed focusing-defocusing versions of the so-called coupled non-linear Schrödinger equation (*r*-CNLS), which has recently been considered in many physical applications. The properties and calculations of some matrix elements using level one vertex operators are outlined.

1 Introduction

Many soliton equations in $1 + 1$ dimensions have integrable multi-component generalizations or more generally integrable matrix generalizations. It is well-known that certain coupled multi-field generalizations of the non-linear Schrödinger equation (CNLS) are integrable and possess soliton type solutions with rich physical properties (see e.g. [1, 2, 3, 4]). The model defined by two coupled NLS systems was earlier studied by Manakov [5]. Another remarkable example of a multi-field generalization of an integrable model is the so-called generalized sine-Gordon model which is integrable in some regions of its parameter space and it has many physical applications [6, 7]. The type of coupled NLS equations find applications in diverse areas of physics such as non-linear optics, optical communication, biophysics, multicomponent Bose-Einstein condensate, etc (see e.g. [1, 2, 3, 8, 9]). The multi-soliton solutions of these systems have recently been considered using a variety of methods depending on the initial-boundary values imposed on the solution. For example, the direct methods, mostly the Hirota method, have been applied in [10, 3] and in ref. [11] a wide class of NLS models have been studied in the framework of the Darboux-dressing transformation. Recently, some properties have been investigated, such as the appearance of stationary bound states formed by two dark-dark solitons in the mixed-nonlinearity case (focusing and defocusing) of the 2-CNLS systems [4]. The inverse scattering transform method (IST) for the defocusing CNLS model with nonvanishing boundary conditions (NVBC) has been an open problem for over 30 years. The two-component case was solved in [12] and the multi-component model has very recently presented in [13]. The inverse scattering [12, 13] and Hirota's method [10] results on N-dark-dark solitons in the defocusing 2-CNLS model have presented only the degenerate case, i.e. the multi-solitons of the both components are proportional to each other and therefore are reducible to the dark solitons of the scalar NLS model.

The r -CNLS model is related to the AKNS $_r$ system, which is a model with $2r$ dependent variables. As we will show below this system reduces to the r -CNLS model under a particular reality condition. Moreover, general multi-dark-dark soliton solutions of generalized AKNS type systems available in the literature, to our knowledge, are scarce. Recently, there appeared some reports on the general non-degenerate N-dark-dark solitons in the r -CNLS model [4, 14]. In [4] the model with mixed nonlinearity is studied in the context of the KP-hierarchy reduction approach. The mixed focusing and defocusing nonlinearity 2-CNLS model presents a two dark-dark soliton stationary bound state [4]. Here we show that similar phenomenon is present in the AKNS $_2$ system. However, three or more dark-dark solitons can not form bound states. In [14] general multi-dark-dark solitons have been obtained in the framework of the algebro-geometric approach of the CNLS $_r$ model.

In this paper we will consider the general constant (vanishing, nonvanishing and mixed vanishing-nonvanishing) boundary value problem for the AKNS $_r$ ($r = 1, 2$) model and show that its particular complexified and reduced version incorporates the focusing and defocusing scalar NLS system for $r = 1$, and the focusing, defocusing and mixed nonlinearity versions of the 2-CNLS system, i.e. the Manakov model, for $r = 2$, respectively. Recently, an extended version of this paper has been published in [15]. We will consider the dressing transformation (DT) method to solve integrable nonlinear equations, which is based on the Lax pair formulation of the system. In this approach the so-called integrable highest weight

representation of the underlying affine Lie algebra is essential to find soliton type solutions (see [16] and references therein). According to the approach of [16] a common feature of integrable hierarchies presenting soliton solutions is the existence of some special “vacuum solutions” such that the Lax operators evaluated on them lie in some Abelian (up to the central term) subalgebra of the associated Kac-Moody algebra. The boundary conditions imposed to the system of equations must be related to the relevant vacuum connections lying in certain Abelian sub-algebra. The soliton type solutions are constructed out of those “vacuum solutions” through the so called soliton specialization in the context of the DT. These developments lead to a quite general definition of tau functions associated to the hierarchies, in terms of the so called “integrable highest weight representations” of the relevant Kac-Moody algebra. However, the free field boundary condition does not allow the vacuum connections to lie in an Abelian sub-algebra; so, we will introduce a modified DT in order to deal with this case. We consider a modified DT relying on the dressing group composition law of two successive DT’s [30]. The corresponding tau functions will provide generalized dark-dark soliton solutions possessing additional parameters as compared to the ones obtained with constant boundary conditions.

Here we adopt a hybrid of the DT and Hirota methods [7, 19] to obtain soliton solutions of the AKNS system. As pointed out in [18], the Hirota method presents some drawbacks in order to derive a general type of soliton solutions, noticeably it is not appropriate to handle vector NLS type models in a general form since the process relies on a insightful guess of the functional forms of each component, whereas in the group theoretical approach adopted here the dependence of each component on the generalized tau functions become dictated by the DT and the solutions in the orbit of certain vacuum solutions are constructed in a uniform way. The generalized tau-functions for the nonlinear systems are defined as an alternative set of variables corresponding to certain matrix elements evaluated in the integrable highest-weight representations of the underlying affine Kac-Moody algebra. In this way we overcome two difficulties of the methods. The Hirota method needs the tau-functions and an expansion for them, but it does not provide a recipe how to construct them. That is accurately solved through the DT method, which in its turn needs the evaluation of certain matrix elements in the vertex operator representation. We may avoid the cumbersome matrix element calculations through the Hirota expansion method. Since this method is recursive it allows a simple implementation on a computer program for algebraic manipulation like MATHEMATICA.

The paper is organized as follows. In section 2 we present the generalized non-linear Schrödinger equation (AKNS_2) associated to the homogeneous gradation of the Kac-Moody algebra sl_3 . The 2-component coupled non-linear Schrödinger equation (CNLS_2) is defined as a particular reduction by imposing certain conditions on the AKNS_2 fields. In section 3 we review the theory of the DT, describe the various constant boundary conditions and define the tau-functions. In 3.1 we present the modified DT associated to the dressing group suitable for free field boundary conditions. In section 4 we apply the DT method to the vanishing boundary conditions and derive the bright solitons. In section 5 we consider the non-vanishing boundary conditions and derive the AKNS_r ($r = 1, 2$) dark solitons. Moreover, the general N-dark-dark solitons with free field NVBC of the AKNS_2 are derived through the modified DT method. In 5.3 we discuss the dark-dark soliton bound states. In 6 the

dark-bright solitons are derived associated to the mixed boundary conditions. Finally, we have included the $sl(2)$ and $sl(3)$ affine Kac-Moody algebra properties, in appendices A and B, respectively. In C the matrix elements in $\hat{sl}(3)$ have been computed using vertex operator representations.

2 The model

The AKNS₂ model can be constructed in the framework of the Zakharov-Shabat formalism using the hermitian symmetric spaces [20]. In [21] an affine Lie algebraic formulation based on the loop algebra $g \otimes \mathbb{C}[\lambda, \lambda^{-1}]$ of $g \in sl(3)$ has been presented. Here instead we construct the model associated to the full affine Kac-Moody algebra $\mathcal{G} = \hat{sl}(3)$ with homogeneous gradation and a semi-simple element $E^{(l)}$. So, the connections are given by

$$A = E^{(1)} + \sum_{i=1}^2 \Psi_i^+ E_{\beta_i}^{(0)} + \sum_{i=1}^2 \Psi_i^- E_{-\beta_i}^{(0)} + \phi_1 C, \quad (2.1)$$

$$\begin{aligned} B = & E^{(2)} + \sum_{i=1}^2 \Psi_i^+ E_{\beta_i}^{(1)} + \sum_{i=1}^2 \Psi_i^- E_{-\beta_i}^{(1)} + \sum_{i=1}^r \partial_x \Psi_i^+ E_{\beta_i}^{(0)} - \sum_{i=1}^r \partial_x \Psi_i^- E_{-\beta_i}^{(0)} - \\ & [\Psi_1^+ \Psi_1^- - \Omega_1](H_1^{(0)} + H_2^{(0)}) - [\Psi_2^+ \Psi_2^- - \Omega_2]H_2^{(0)} - \\ & \Psi_1^+ \Psi_2^- E_{\beta_1 - \beta_2}^{(0)} - \Psi_2^+ \Psi_1^- E_{\beta_2 - \beta_1}^{(0)} + \phi_2 C, \end{aligned} \quad (2.2)$$

where Ψ_i^+ , Ψ_i^- , ϕ_1 and ϕ_2 are the fields of the model. Notice that the auxiliary fields ϕ_1 and ϕ_2 lie in the direction of the affine Lie algebra central term C . The existence of this term plays an important role in the theory of the so-called integrable highest weight representations of the Kac-Moody algebra and they will be important below in order to find the soliton type solutions of the model. The connections (2.1)-(2.2) conveniently incorporate the terms with the constant parameters $\Omega_{1,2}$ in order to take into account the various boundary conditions, so, these differ slightly from the ones in [22, 21].

In the basis considered in this paper one has

$$E^{(l)} = \frac{1}{3} \sum_{a=1}^2 a H_a^{(l)} \quad (2.3)$$

The β_i in (2.1)-(2.2) are positive roots defined in (B.24). Notice that the connections lie in the subspaces

$$A \in \hat{g}_0 + \hat{g}_1, \quad B \in \hat{g}_0 + \hat{g}_1 + \hat{g}_2. \quad (2.4)$$

The zero curvature condition $[\partial_t - B, \partial_x - A] = 0$ provides the following system of equations

$$\partial_t \Psi_i^+ = +\partial_x^2 \Psi_i^+ - 2 \left[\sum_{j=1}^2 \Psi_j^+ \Psi_j^- - \frac{1}{2}(\beta_i, \vec{\Omega}) \right] \Psi_i^+, \quad i = 1, 2 \quad (2.5)$$

$$\partial_t \Psi_i^- = -\partial_x^2 \Psi_i^- + 2 \left[\sum_{j=1}^2 \Psi_j^+ \Psi_j^- - \frac{1}{2}(\beta_i, \vec{\Omega}) \right] \Psi_i^-, \quad (2.6)$$

$$\begin{aligned}\partial_t \bar{\phi}_1 - \partial_x \bar{\phi}_2 &= 0; \\ \vec{\Omega} &\equiv \sum_{i=1}^2 \Omega_i \beta_i, \quad (\beta_1 \cdot \vec{\Omega}) = 2\Omega_1 + \Omega_2, \quad (\beta_2 \cdot \vec{\Omega}) = 2\Omega_2 + \Omega_1\end{aligned}\quad (2.7)$$

Notice that the auxiliary fields $\phi_{1,2}$ completely decouple from the AKNS₂ fields Ψ_j^\pm . The integrability of the system of equations (2.5)-(2.6), for $\Omega_{1,2} = 0$, and its multi-Hamiltonian structure have been established [22]. The system of equations (2.5)-(2.6) supplied with a convenient complexification can be related to some versions of the so-called coupled non-linear Schrödinger equation (CNLS) [5, 1, 2, 3]. For example by making

$$t \rightarrow -it, \quad [\Psi_i^+]^* = -\mu \sigma_i \Psi_i^- \equiv -\mu \sigma_i \psi_i, \quad (2.8)$$

where \star means complex conjugation, $\mu \in \mathbb{R}_+$, $\sigma_i = \pm 1$, we may reduce the system (2.5)-(2.6) to the well known 2-coupled non-linear Schrödinger system (2-CNLS)

$$i \partial_t \psi_k + \partial_x^2 \psi_k + 2\mu \left(\sum_{j=1}^2 \sigma_j |\psi_j|^2 - \frac{1}{2} |(\beta_k \cdot \vec{\Omega})| \right) \psi_k = 0, \quad k = 1, 2. \quad (2.9)$$

The parameter $\mu > 0$ represents the strength of nonlinearity and the coefficients σ_j define the sign of the nonlinearity. The system (2.9) can be classified into three classes depending on the signs of the nonlinearity coefficients σ_i . For $\sigma_1 = \sigma_2 = 1$ this system is the focusing Manakov model which supports bright-bright solitons [5]. For $\sigma_1 = \sigma_2 = -1$, it is the defocusing Manakov model which supports bright-dark and dark-dark solitons [10, 12, 23]. In the cases $\sigma_1 = -\sigma_2 = \pm 1$ one has the mixed focusing-defocusing nonlinearities. In this case, these equations support bright-bright solitons [24, 3], bright-dark solitons [25]. This mixed nonlinearity has recently been considered in [4] through the KP-hierarchy reduction method and dark-dark solitons have been obtained. The system (2.9) has also been considered in the study of oscillations and interactions of dark and dark-bright solitons in Bose-Einstein condensates [26, 27]. The multi-dark-dark solitons in the mixed non-linearity case are useful for many physical applications such as nonlinear optics, water waves and Bose-Einstein condensates, where the generally coupled NLS equations often appear.

The focusing CNLS system possesses a remarkable type of soliton solution undergoing a shape changing (inelastic) collision property due to intensity redistribution among its modes. In this context, it has been found a novel class of solutions called partially coherent solitons (PCS) which are of substantially variable shape, such that under collisions the profiles remain stationary [1, 28, 29]. Interestingly, the PCSs, namely, 2-PCS, 3-PCS, ..., r-PCS, are special cases of the well known 2-soliton, 3-soliton, ..., r-soliton solutions of the 2-CNLS, 3-CNLS, ..., r-CNLS equations, respectively [2, 3]. So, the understanding of the variable shape collisions and many other properties of these partially coherent solitons can be studied by providing the higher-order soliton solutions of the r-CNLS ($r \geq 2$) system considered as submodel of the relevant AKNS_r. We believe that the group theoretical point of view of finding the analytical results for the general case of N -soliton interactions would facilitate the study of their properties; for example, the asymptotic behavior of trains of N solitonlike pulses with approximately equal amplitudes and velocities, as studied in [17]. Notice that the set of solutions of the AKNS model (2.5)-(2.6) is much larger than the solutions of the CNLS

system (2.9), since only the solutions of the former which satisfy the constraints (2.8) will be solutions of the CNLS model (2.9). This fact will be seen below in many instances; so, we believe that the AKNS soliton properties with relevant boundary conditions deserve a further study.

We will show that the three classes mentioned above, i.e. the *focusing*, *defocusing* and the *mixed focusing-defocusing* CNLS model can be related respectively to the vanishing, nonvanishing and mixed vanishing-nonvanishing boundary conditions of the AKNS model in the framework of the DT approach adapted conveniently to each case. The (constant) non-vanishing boundary conditions require the extension of the DT method to incorporate non-zero constant vacuum solutions. Therefore, the vertex operators corresponding to the vanishing boundary case undergo a generalization in such a way that the nilpotency property, which is necessary in obtaining soliton solutions, should be maintained. Using the modified vertex operators we construct multi-soliton solutions in the cases of (constant) non-vanishing and *mixed* vanishing-nonvanishing boundary conditions. The free field NVBC requires a modification of the usual DT of [16] by considering two successive DT's in the context of the dressing group [30]. However, the same vertex operator generating the dark-dark solitons in the (constant) NVBC will be used in the free field NVBC case with a modified τ functions.

3 Dressing transformations for AKNS₂

In this section we summarize the so-called DT procedure to find soliton solutions, which works by choosing a vacuum solution and then mapping it into a non-trivial solution, following the approach of [16]. For simplicity, we concentrate on a version of the DT suitable for vanishing, (constant) non-vanishing and mixed boundary conditions of the AKNS₂ model (2.5)-(2.7). The free field boundary condition requires a modified DT which is developed in subsection 3.1. So, let us consider

$$\lim_{|x| \rightarrow \infty} \Psi_j^\pm \rightarrow \rho_j^\pm; \quad \phi_{1,2} \rightarrow 0; \quad \rho_j^\pm = \text{const.} \quad (3.1)$$

We may identify the NVBC (3.1) to certain classes of trivial vacuum solutions of the system (2.5)-(2.6):

1) the trivial zero vacuum solution

$$\Psi_{j,vac}^\pm = \rho_j^\pm = 0, \quad j = 1, 2 \quad (3.2)$$

2) the trivial constant vacuum solution

$$\Psi_{j,vac}^\pm = \rho_j^\pm \neq 0, \quad j = 1, 2. \quad (3.3)$$

Notice that (3.3) is a trivial constant vacuum solution of (2.5)-(2.6) provided that the expression $[\sum_{j=1}^2 \rho_j^+ \rho_j^- - \frac{1}{2}(\beta_i \cdot \vec{\Omega})]$ vanishes for any $i = 1, 2$; which is achieved if $\Omega_1 = \Omega_2 \equiv \Omega$, implying $\sum_{j=1}^2 \rho_j^+ \rho_j^- = \frac{3}{2}\Omega$.

3) the mixed constant-zero (zero-constant) vacuum solutions

$$i) \Psi_{1,vac}^\pm = \rho_1^\pm \neq 0, \quad \Psi_{2,vac}^\pm = \rho_2^\pm = 0, \quad (3.4)$$

$$ii) \Psi_{1,vac}^\pm = \rho_1^\pm = 0, \quad \Psi_{2,vac}^\pm = \rho_2^\pm \neq 0. \quad (3.5)$$

The first mixed trivial solution (3.4) requires $2\rho_1^+\rho_1^- = 2\Omega_1 + \Omega_2$, whereas for the second trivial solution (3.5) it must be $2\rho_2^+\rho_2^- = 2\Omega_2 + \Omega_1$.

The connections (2.1)-(2.2) for the above boundary conditions (3.1) take the form

$$A^{vac} \equiv E^{(1)} + \rho_1^+ E_{\beta_1}^{(0)} + \rho_1^- E_{-\beta_1}^{(0)} + \rho_2^+ E_{\beta_2}^{(0)} + \rho_2^- E_{-\beta_2}^{(0)}, \quad (3.6)$$

$$B^{vac} \equiv E^{(2)} + \rho_1^+ E_{\beta_1}^{(1)} + \rho_1^- E_{-\beta_1}^{(1)} + \rho_2^+ E_{\beta_2}^{(1)} + \rho_2^- E_{-\beta_2}^{(1)} - \rho_1^+ \rho_2^- E_{\beta_1-\beta_2}^{(0)} - \rho_2^+ \rho_1^- E_{\beta_2-\beta_1}^{(0)} - (\rho_1^+ \rho_1^- - \Omega_1)(H_1^{(0)} + H_2^{(0)}) - (\rho_2^+ \rho_2^- - \Omega_2)H_2^{(0)}. \quad (3.7)$$

Notice that $[A^{vac}, B^{vac}] = 0$. These connections are related to the group element $\Psi^{(0)}$ through

$$A^{(vac)} = \partial_x \Psi^{(0)} [\Psi^{(0)}]^{-1}; \quad B^{(vac)} = \partial_t \Psi^{(0)} [\Psi^{(0)}]^{-1}, \quad (3.8)$$

where

$$\Psi^{(0)} = e^{xA^{vac} + tB^{vac}} \quad (3.9)$$

The DT is implemented through two gauge transformations generated by Θ_{\pm} such that the nontrivial gauge connections in the vacuum orbit become [16]

$$A = \Theta_{\pm}^h A^{(vac)} [\Theta_{\pm}^h]^{-1} + \partial_x \Theta_{\pm}^h [\Theta_{\pm}^h]^{-1} \quad (3.10)$$

$$B = \Theta_{\pm}^h B^{(vac)} [\Theta_{\pm}^h]^{-1} + \partial_t \Theta_{\pm}^h [\Theta_{\pm}^h]^{-1} \quad (3.11)$$

where

$$\Theta_{-}^h = \exp \left(\sum_{n>0} \sigma_{-n} \right), \quad \Theta_{+}^h \equiv M^{-1} N; \quad M = \exp(\sigma_o), \quad N = \exp \left(\sum_{n>0} \sigma_n \right), \quad (3.12)$$

where $[D, \sigma_n] = n \sigma_n$. Therefore, from the relationships

$$A = \partial_x (\Psi^h) [\Psi^h]^{-1}; \quad B = \partial_t (\Psi^h) [\Psi^h]^{-1}, \quad \Psi^h \equiv \Theta_{\pm}^h \Psi^{(0)}, \quad (3.13)$$

one has

$$[\Theta_{-}^h]^{-1} \Theta_{+}^h = \Psi^{(0)} h [\Psi^{(0)}]^{-1}, \quad (3.14)$$

where h is some constant group element.

One can relate the fields Ψ_i^{\pm} , ϕ_1 and ϕ_2 to some of the components in σ_n . One has

$$A = A^{vac} + [\sigma_{-1}, E^{(1)}] + \text{terms of negative grade.} \quad (3.15)$$

$$B = B^{vac} + [\sigma_{-1}, E^{(2)}] + [\sigma_{-2}, E^{(2)}] + \frac{1}{2} [\sigma_{-1}, [\sigma_{-1}, E^{(2)}]] + \left[\sigma_{-1}, \sum_{i=1}^2 \rho_i^+ E_{\beta_i}^{(1)} \right] + \left[\sigma_{-1}, \sum_{i=1}^2 \rho_i^- E_{-\beta_i}^{(1)} \right] + \text{terms of negative grade} \quad (3.16)$$

Taking into account the grading structure of the connection A in (2.4) we may write the σ'_n s in terms of the fields of the model. In order to match the zero grade terms of the both sides of the equation (3.15) one must have

$$\sigma_{-1} = - \sum_{i=1}^2 (\Psi_i^+ - \rho_i^+) E_{\beta_i}^{(-1)} + \sum_{i=1}^2 (\Psi_i^- - \rho_i^-) E_{-\beta_i}^{(-1)} + \sum_{a=1}^2 \sigma_{-1}^a H_a^{(-1)}. \quad (3.17)$$

In the equation above, the explicit form of σ_{-1}^a in terms of the fields Ψ_i^\pm can be obtained by setting the sum of the (-1) grade terms to zero. Nevertheless, the form of the σ_{-1}^a will not be necessary for our purposes.

Following the above procedure to match the gradations on both sides of eqs. (3.15)-(3.16) one notices that the σ_{-n} 's with $n \geq 1$ are used to cancel out the undesired components on the r.h.s. of the equations.

From the equations (3.12)-(3.13) and (3.14) one has

$$\langle \lambda | M^{-1} | \lambda' \rangle = \langle \lambda | [\Psi^{(0)} h \Psi^{(0)-1}] | \lambda' \rangle, \quad (3.18)$$

where $\langle \lambda |$ and $| \lambda' \rangle$ are certain states annihilated by $\mathcal{G}_<$ and $\mathcal{G}_>$, respectively. Defining

$$\sigma_o = \sum_{\alpha>0} \sigma_o^\alpha E_\alpha^{(0)} + \sum_{\alpha>0} \sigma_o^{-\alpha} E_{-\alpha}^{(0)} + \sum_{a=1}^2 \sigma_o^a H_a^{(0)} + \eta C \quad (3.19)$$

and choosing a specific matrix element one gets a space time dependence for the field η

$$e^{-\eta} = \langle \lambda_o | [\Psi^{(0)} h \Psi^{(0)-1}] | \lambda_o \rangle \quad (3.20)$$

$$\equiv \tau_0, \quad (3.21)$$

where we have defined the tau function τ_0 . Next, we will write the fields Ψ_i^\pm in terms of certain matrix elements. These functions will be represented as matrix elements in an appropriate representation of the affine Lie algebra $\widehat{sl}(3)$. We proceed by writing the eq. (3.14) in the form

$$\exp \left(- \sum_{n>0} \sigma_{-n} \right) | \lambda_o \rangle = [\Psi^{(0)} h \Psi^{(0)-1}] | \lambda_o \rangle \tau_0^{-1}, \quad (3.22)$$

where the eqs. (3.12), (3.19) and (3.20)-(3.21) have been used.

Then the terms with grade (-1) in the both sides of (3.22) can be written as

$$-\sigma_{-1} | \lambda_o \rangle = \frac{[\Psi^{(0)} h \Psi^{(0)-1}]_{(-1)} | \lambda_o \rangle}{\tau_0(x, t)} \quad (3.23)$$

or equivalently

$$\sum_{i=1}^2 (\Psi_i^+ - \rho_i^+) E_{\beta_i}^{(-1)} - \sum_{i=1}^2 (\Psi_i^- - \rho_i^-) E_{-\beta_i}^{(-1)} - \sum_{a=1}^2 \sigma_{-1}^a H_a^{(-1)} | \lambda_o \rangle = \frac{[\Psi^{(0)} h \Psi^{(0)-1}]_{(-1)} | \lambda_o \rangle}{\tau_0(x, t)}. \quad (3.24)$$

Acting on the left in eq. (3.24) by $E_{\pm\beta_i}^{(1)}$ and taking the relevant matrix element with the dual highest weight state $\langle \lambda_o |$ we may have

$$\Psi_i^+ = \rho_i^+ + \frac{\tau_i^+}{\tau_0} \quad \text{and} \quad \Psi_i^- = \rho_i^- - \frac{\tau_i^-}{\tau_0}; \quad i = 1, 2. \quad (3.25)$$

where the *tau* functions τ_i^\pm , τ_0 are defined by

$$\tau_i^+(x, t) \equiv \langle \lambda_o | E_{-\beta_i}^{(1)} [\Psi^{(0)} h \Psi^{(0)-1}]_{(-1)} | \lambda_o \rangle, \quad (3.26)$$

$$\tau_i^-(x, t) \equiv \langle \lambda_o | E_{\beta_i}^{(1)} [\Psi^{(0)} h \Psi^{(0)-1}]_{(-1)} | \lambda_o \rangle, \quad (3.27)$$

$$\tau_0(x, t) \equiv \langle \lambda_o | [\Psi^{(0)} h \Psi^{(0)-1}]_{(o)} | \lambda_o \rangle. \quad (3.28)$$

Notice that in order to get the above relationships we have used the commutation rules for the corresponding affine Kac-Moody algebra elements, as well as their properties acting on the highest weight state (see Appendix B).

According to the solitonic specialization in the context of the DT method the soliton solutions are determined by choosing suitable constant group elements h in the eqs. (3.26)-(3.28). In order to obtain n -soliton solutions the general prescription is to parameterize the orbit of the vacuum as a product of exponentials of eigenvectors of the operators ε_l ($\varepsilon_1 = A^{vac}$, $\varepsilon_2 = B^{vac}$) defined in (3.6)-(3.7); i.e $h = \prod_{i=1}^n e^{F_i}$, where $[\varepsilon_l, F_i] = \lambda_i^l F_i$, such that $(F_i)^m \neq 0$ only for $m < m_i$, m_i being some positive integer. The relationships between DT, solitonic specialization and the Hirota method have been presented in [16] for any hierarchy of integrable models possessing a zero curvature representation in terms of an affine KacMoody algebra. The DT method provides a relationship between the fields of the model and the relevant tau functions, and it explains the truncation of the Hirota expansion. The Hirota method is a recursive method which can be implemented through a computer program for algebraic manipulation like MATHEMATICA. On the other hand, the DT method requires the computation of matrix elements as in eqs. (3.26)-(3.28) in the vertex operator representations of the affine KacMoody algebra. Actually, these matrix element calculations are very tedious in the case of higher soliton solutions.

3.1 Free field boundary conditions and dressing group

As a generalization of the constant NVBC (3.3) consider the free field NVBC

$$\lim_{x \rightarrow \pm\infty} \Psi_j^\pm \rightarrow \hat{\rho}_i^\pm(x, t) \equiv \rho_i^\pm e^{a_i^\pm x + b_i^\pm t} \quad (3.29)$$

where ρ_i^\pm , a_i^\pm , b_i^\pm are some constants such that $a_i^+ = -a_i^-$; $b_i^+ = -b_i^-$. Notice that (3.29) is a free field solution of (2.5)-(2.6) such that $b_i^\pm = \pm(a_i^\pm)^2 \mp 2\Lambda_i$, for $\Lambda_i \equiv [\sum_{j=1}^2 \hat{\rho}_j^+ \hat{\rho}_j^- - \frac{1}{2}(\beta_i, \vec{\Omega})] = \text{const.}$ However, a direct application of the DT approach of [16] is not possible since the relevant connections \hat{A}^{vac} , \hat{B}^{vac} do not belong to an abelian subalgebra (up to the central term). So, in order to consider the NVBC (3.29) we resort to the dressing group [30] composition law of two successive DTs. In fact, the relevant connections can be related

to the A^{vac}, B^{vac} in (3.6)-(3.7) through certain gauge transformations generated by Θ_{\pm}^g such that

$$\hat{V}_i^{vac} = \Theta_{\pm}^g(x, t) V_i^{vac} [\Theta_{\pm}^g(x, t)]^{-1} + \partial_{x_i} \Theta_{\pm}^g(x, t) [\Theta_{\pm}^g(x, t)]^{-1}; \quad \hat{V}_1 = \hat{A}^{vac}, \quad \hat{V}_2 = \hat{B}^{vac}, \quad (3.30)$$

where $i = 1, 2$, $x_1 = x$, $x_2 = t$, $V_1 = A^{vac}$, $V_2 = B^{vac}$, and $\Theta_{\pm}^g \in \widehat{SL}(3)$. One must have

$$\Psi^{(0)} \rightarrow \hat{\Psi}^{(0)} = \Theta_{+}^g(x, t) \Psi^{(0)} = \Theta_{-}^g(x, t) \Psi^{(0)} g, \quad \hat{V}_i^{(vac)} = \partial_{x_i} \hat{\Psi}^{(0)} [\hat{\Psi}^{(0)}]^{-1}, \quad (3.31)$$

g is an arbitrary constant group element, and

$$\Theta_{+}^g = e^{\chi_0} e^{\sum_{n>0} \chi_n}, \quad \Theta_{-}^g = e^{-\chi_0} e^{\sum_{n>0} \chi_{-n}}, \quad [D, \chi_n] = n \chi_n. \quad (3.32)$$

From (3.30) one has $e^{-\chi_0} E^{(l)} e^{\chi_0} = E^{(l)}$, so

$$\chi_0 = \chi_0^{+} E_{\beta_1 - \beta_2}^{(0)} + \chi_0^{-} E_{-\beta_1 + \beta_2}^{(0)} + \sum_{a=1}^2 \chi_0^a H_a^{(0)} + \chi C \quad (3.33)$$

$$\chi_{-1} = - \sum_{i=1}^2 [f_i^{+}(x, t) - \rho_i^{+}] E_{\beta_i}^{(-1)} + \sum_{i=1}^2 [f_i^{-}(x, t) - \rho_i^{-}] E_{-\beta_i}^{(-1)} + \sum_{a=1}^2 \chi_{-1}^a H_a^{(-1)}. \quad (3.34)$$

Therefore, a modified DT procedure can be implemented with [30]

$$[\Theta_{-}^{h'g}]^{-1} \Theta_{+}^{h'g} \equiv \hat{\Psi}^{(0)} h' [\hat{\Psi}^{(0)}]^{-1} = \Theta_{-}^g(x, t) \Psi^{(0)} h [\Psi^{(0)}]^{-1} [\Theta_{+}^g(x, t)]^{-1}; \quad h \equiv gh', \quad (3.35)$$

through

$$A' = \Theta_{\pm}^{h'g} A^{(vac)} [\Theta_{\pm}^{h'g}]^{-1} + \partial_x \Theta_{\pm}^{h'g} [\Theta_{\pm}^{h'g}]^{-1}; \quad B' = \Theta_{\pm}^{h'g} B^{(vac)} [\Theta_{\pm}^{h'g}]^{-1} + \partial_t \Theta_{\pm}^{h'g} [\Theta_{\pm}^{h'g}]^{-1}. \quad (3.36)$$

So, the equation (3.35) can be used instead of (3.14) in order to derive general dark-dark solitons with free field NVBC.

4 Bright solitons and vanishing boundary conditions

For simplicity we first apply the DT method to the VBC (3.2) and show the existence of bright-bright soliton solutions [31]. The connections $A^{vac} \equiv E^{(1)}$, $B^{vac} \equiv E^{(2)}$ are related to the group element Ψ_0 as

$$A^{(vac)} = \partial_x \Psi_0 [\Psi_0]^{-1}; \quad B^{(vac)} = \partial_t \Psi_0 [\Psi_0]^{-1}; \quad \Psi_0 \equiv e^{xE^{(1)} + tE^{(2)}}. \quad (4.1)$$

In order to obtain soliton solutions the simultaneous adjoint eigenstates of the elements $E^{(1)}$, $E^{(2)}$ play a central role. In the case at hand one has the eigenstates

$$F_j = \sum_{n=-\infty}^{+\infty} \nu_j^n E_{-\beta_j}^{(-n)}; \quad G_k = \sum_{n=-\infty}^{+\infty} \rho_k^n E_{\beta_k}^{(-n)}; \quad j, k = 1, 2 \quad (4.2)$$

such that

$$[xE^{(1)} + tE^{(2)}, F_j] = -\varphi_j(x, t) F_j; \quad \varphi_j(x, t) = \nu_j(x + \nu_j t) \quad (4.3)$$

$$[xE^{(1)} + tE^{(2)}, G_j] = \eta_j(x, t) G_j, \quad \eta_j(x, t) = \rho_j(x + \rho_j t). \quad (4.4)$$

4.1 j^{th} -component one bright-soliton solution

Consider the product

$$h = e^{a_{j_1} F_{j_1}} e^{b_{j_2} G_{j_2}}, \quad (4.5)$$

where j_1 and j_2 are some indexes chosen from $\{1, 2\}$. Using the nilpotency properties of F_{j_1} and G_{j_2} (see Appendix C) one gets

$$[\Psi_0 h \Psi_0^{-1}] = (1 + e^{-\varphi_{j_1}} a_{j_1} F_{j_1}) (1 + e^{\eta_{j_2}} b_{j_2} G_{j_2}) \quad (4.6)$$

$$= 1 + e^{-\varphi_{j_1}} a_{j_1} F_{j_1} + e^{\eta_{j_2}} b_{j_2} G_{j_2} + a_{j_1} b_{j_2} e^{-\varphi_{j_1}} e^{\eta_{j_2}} F_{j_1} G_{j_2}, \quad (4.7)$$

with φ_{j_1} and η_{j_2} given in (4.3)-(4.4). The corresponding tau functions become

$$\tau_0 = 1 + a_{j_1} b_{j_2} C_{j_1, j_2} e^{-\varphi_{j_1}} e^{\eta_{j_2}}, \quad C_{j_1, j_2} = \frac{\nu_{j_1} \rho_{j_2}}{(\nu_{j_1} - \rho_{j_2})^2} \delta_{j_1, j_2}, \quad (4.8)$$

$$\tau_i^+ = \delta_{i, j_2} b_{j_2} \rho_{j_2} e^{\eta_{j_2}}, \quad \tau_i^- = \delta_{i, j_1} a_{j_1} \nu_{j_1} e^{-\varphi_{j_1}}, \quad (4.9)$$

where the matrix element C_{j_1, j_2} has been presented in (C.2). In order to construct *one-soliton* solutions we must have $j_1 = j_2 \equiv j$ in (4.8). Therefore one gets ¹

$$\Psi_i^+ = \frac{b_i \rho_i e^{\eta_i}}{1 + a_i b_i C_{i, i} e^{-\varphi_i} e^{\eta_i}}, \quad \Psi_i^- = -\frac{a_i \nu_i e^{-\varphi_i}}{1 + a_i b_i C_{i, i} e^{-\varphi_i} e^{\eta_i}}, \quad i = j; \quad (4.10)$$

$$\Psi_i^\pm = 0, \quad i \neq j \quad (4.11)$$

Impose the relationships $\rho_j^* = -\nu_j$, $b_j^* = -\mu \sigma a_j$, $\nu_j, a_j \in \mathbb{C}$; so from (4.8) one has $C_{j, j} = -(\frac{\nu_j}{2\sigma a_j})^2$. The equations (4.10)-(4.11) with the complexification (2.8) provide a solution of the CNLS system (2.9)

$$i(x, t) = \begin{cases} -\frac{(a_i \nu_i) \nu_{iR}}{\sqrt{\mu \sigma |a_j \nu_j|^2}} e^{i\varphi_{iI}} \operatorname{sech}\left(\varphi_{iR} + \frac{X_0}{2}\right), & i = j \\ 0 & i \neq j \end{cases} \quad (4.12)$$

where $e^{X_0} = \frac{\mu \sigma |a_j \nu_j|^2}{(\nu_j + \nu_j^*)^2}$, $\varphi_j = \nu_j (x - i\nu_j t) \equiv \varphi_{jR} + i\varphi_{jI}$.

The solution (4.12) for the j 'th component is known as a 'bright soliton' in the context of the scalar non-linear Schrödinger equation (NLS).

4.2 1-bright-bright soliton solution

The main observation in the last construction of the j^{th} -component one-soliton is that it has been excited by the group element h in (4.5) such that $j_1 = j_2 = j$. So, in order to excite the two components of Ψ_i^\pm ($i = 1, 2$) and reproduce a bright soliton for each component let us consider the group element

$$h = e^{a_1 F_1} e^{b_1 G_1} e^{a_2 F_2} e^{b_2 G_2}, \quad (4.13)$$

¹If $j_1 \neq j_2$ in (4.8)-(4.9) one still has certain trivial solutions, since in this case $C_{j_1, j_2} = 0$ implying $\tau_0 = 1$

where the exponential factors contain F 's and G 's of type (C.1). The tau functions become

$$\tau_j^+ = b_j \rho_j e^{\eta_j}, \quad \eta_j = \rho_j (x + \rho_j t) + \rho_{0j}, \quad j = 1, 2. \quad (4.14)$$

$$\tau_j^- = a_j \nu_j e^{-\varphi_j}, \quad \varphi_j = \nu_j (x + \nu_j t) + \nu_{0j}, \quad j = 1, 2. \quad (4.15)$$

$$\tau_0 = 1 + a_1 b_1 C_{11} e^{-\varphi_1} e^{\eta_1} + a_2 b_2 C_{22} e^{-\varphi_2} e^{\eta_2}, \quad (4.16)$$

where $C_{jj} = \frac{\rho_j \nu_j}{(\rho_j - \nu_j)^2}$. Let us set $b_j^* = -\mu \sigma_j a_j$, $\rho_j^* = -\nu_j \equiv -\nu_1$, then $\eta_j^* = -\varphi_j \equiv -\varphi_1 = -(\varphi_{1R} + i\varphi_{1I})$, $(\nu_j, a_j \in \mathbb{C})$ in the relations (4.14)-(4.16). Therefore, the vector soliton solution of the 2-CNLS eq. (2.9) arises

$$\begin{pmatrix} 1(x, t) \\ 2(x, t) \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \nu_{1R} \operatorname{sech} \left(\varphi_{1R} + \frac{X_0}{2} \right) e^{i\varphi_{1I}}, \quad (4.17)$$

where $A_i = -\frac{a_i \nu_i}{\mu^{1/2} \sum_{j=1}^2 (\sigma_j |a_j \nu_j|^2)^{1/2}}$; $e^{X_0} = \mu \frac{\sum_j \sigma_j |a_j \nu_j|^2}{(\nu_1 + \nu_1^*)^2}$.

Notice that this 1-bright-bright soliton possesses 3 arbitrary complex parameters, namely, a_1, a_2 and $\nu_1 = \nu_{1R} + i\nu_{1I}$. This solution in the case of mixed nonlinearity $\sigma_1 = -\sigma_2 = 1$, may have a singular behavior if the sum $\sum_{j=1}^2 (\sigma_j |a_j \nu_j|^2)^{1/2}$ in the denominator of the expression of A_i vanishes, such that the soliton amplitude in (4.17) diverges. The N -bright-bright soliton requires the generalization of the group element in (4.13) as $h = e^{a_1 F_1} e^{b_1 G_1} \dots e^{a_N F_N} e^{b_N G_N}$.

5 AKNS_r ($r = 1, 2$), NVBC and dark solitons

In this section we tackle the problem of finding dark soliton type solutions of the system (2.5)-(2.6). The associated coupled NLS model (2.9) with nonvanishing boundary conditions has been considered in the framework of direct methods, such as the Hirota tau function approach (see e.g. [10, 32, 23, 33]), and recently in the inverse scattering transform approach [12, 13]. In the last approach the relevant Lax operators have remarkable differences and rather involved spectral properties as compared to their counterparts with vanishing boundary conditions (see [34] and references therein), e.g. in the NVBC case the spectral parameter requires the construction of certain Riemann sheets [34, 35, 36, 37]. So, it would be interesting to give the full Lie algebraic construction of the tau functions and soliton solutions for the system (2.5)-(2.6) with NVBC.

5.1 AKNS₁: N-dark solitons

For simplicity, firstly we describe the entire process for the system (2.5)-(2.6) with just two fields Ψ^\pm . So, in order to study a NVBC for the system $sl(2)$ -AKNS consider the Lax pair

$$A = E^{(1)} + \Psi^+ E_+^{(0)} + \Psi^- E_-^{(0)} + \phi_1 C, \quad (5.1)$$

$$B = E^{(2)} + \Psi^+ E_+^{(1)} + \Psi^- E_-^{(1)} + \partial_x \Psi^+ E_+^{(0)} - \partial_x \Psi^- E_-^{(0)} - 2(\Psi^+ \Psi^- - \rho^+ \rho^-) H^{(0)} + \phi_2 C, \quad (5.2)$$

where Ψ^+ , Ψ^- are the fields of the model ($\rho^\pm = \text{constant}$). In this case one considers the generator $E^{(l)} \equiv H^{(l)}$. The ϕ_1 and ϕ_2 are introduced as auxiliary fields. Therefore the equations of motion suitable to treat NVBC become

$$\partial_t \Psi^+ = \partial_x^2 \Psi^+ - 2(\Psi^+ \Psi^- - \rho^+ \rho^-) \Psi^+, \quad (5.3)$$

$$\partial_t \Psi^- = -\partial_x^2 \Psi^- + 2(\Psi^+ \Psi^- - \rho^+ \rho^-) \Psi^-, \quad (5.4)$$

$$\partial_t \phi_1 - \partial_x \phi_2 = 0. \quad (5.5)$$

We will show that the $\hat{sl}(2)$ -AKNS model, through a particular reduction, contains as sub-model the scalar defocusing NLS system

$$i \partial_t \psi + \partial_x^2 \psi - 2(|\psi|^2 - \rho^2) \psi = 0. \quad (5.6)$$

This equation is suitable for treating nonvanishing boundary conditions (NVBC) [37, 34, 35]

$$(x, t) = \begin{cases} \rho, & x \rightarrow -\infty \\ \rho \epsilon^2, & x \rightarrow +\infty \end{cases}; \quad \epsilon = e^{i\theta/2}, \quad \rho \in \mathbb{R}. \quad (5.7)$$

In fact, this boundary condition is manifestly compatible with the eq. of motion (5.6). The defocusing NLS equation (5.6) with the boundary condition (5.7) is exactly integrable by the inverse scattering technique [38]. This model has soliton solutions in the form of localized “dark” pulses created on a constant or stable *continuous* wave background solution.

In the system of eqs. (5.3)-(5.4) consider the transformation

$$t \rightarrow -it, \quad x \rightarrow -ix, \quad (5.8)$$

$$\Psi^\pm \rightarrow i \Psi^\pm \epsilon^{\mp 2}, \quad (5.9)$$

where the factor $\epsilon^{\mp 2}$ is introduced for later convenience. Furthermore, the identification

$$\equiv \Psi^+ = (\Psi^-)^*, \quad (5.10)$$

where the star stands for complex conjugation, such that $\Psi_0^+ \Psi_0^- \rightarrow -\rho^2$ allows one to reproduce the defocusing NLS equation (5.6).

Let us take as vacuum solution of (5.3)-(5.4) the constant background configuration

$$\Psi^\pm = \rho^\pm \equiv \rho \epsilon^{\mp 2}, \quad \phi_{1,2} = 0, \quad \rho, \epsilon = \text{const.} \quad (5.11)$$

Therefore the gauge connections (5.1)-(5.2) for the vacuum solution (5.11) become

$$A^{vac} \equiv \varepsilon_1 = E^{(1)} + \rho^+ E_+^{(0)} + \rho^- E_-^{(0)} \quad (5.12)$$

$$B^{vac} = \varepsilon_2 = E^{(2)} + \rho^+ E_+^{(1)} + \rho^- E_-^{(1)} \quad (5.13)$$

We follow eq. (3.8) in order to write the connections in the form

$$A^{(vac)} = \partial_x \Psi_{nvbc}^{(0)} [\Psi_{nvbc}^{(0)}]^{-1}, \quad B^{(vac)} = \partial_t \Psi_{nvbc}^{(0)} [\Psi_{nvbc}^{(0)}]^{-1}, \quad (5.14)$$

with the group element

$$\begin{aligned}\Psi_{nvbc}^{(0)} &\equiv e^{x\varepsilon_1+t\varepsilon_2} \\ &= \exp \left[x(E^{(1)} + \rho^+ E_+^{(0)} + \rho^- E_-^{(0)}) + t(E^{(2)} + \rho^+ E_+^{(1)} + \rho^- E_-^{(1)}) \right]\end{aligned}\quad (5.15)$$

Let us emphasize that this group element differs fundamentally from the one associated to the case with VBC. First, the difference originates from the constant boundary condition terms added to the relevant vacuum gauge connections. However, they must be related since $\lim_{\rho \rightarrow 0} \Psi_{nvbc}^{(0)} \rightarrow \Psi_0$, where Ψ_0 is the group element in the VBC case (4.1). Second, the vacuum connections $\varepsilon_{1,2}$ in the DT procedure will require another eigenvectors under the adjoint actions in analogy to eqs. (4.3)-(4.4), as we will see below.

Another proposal for this group element $\Psi_{nvbc}^{(0)}$ has been introduced in [39]. There it has been considered a group element inspired in the inverse scattering method, possessing double-valued functions of the spectral parameter λ . This fact motivates the introduction of an affine parameter to avoid constructing Riemann sheets. This construction involves some complications when used in the context of the DT method. The main difficulty arises in the computation of the matrix elements associated to the highest weight states in order to find the relevant tau functions. However, the group element given in (5.15) will turn out to be more suitable in the DT procedure. In fact, a similar group element has been proposed in [40] to tackle a NVBC soliton solutions in the so-called negative even grade mKdV hierarchy.

The DT procedure in this case follows all the way verbatim as in section 3 adapted to the $sl(2)$ case. It follows by relating the relevant connections (5.1)-(5.2) with the connections in eqs. (5.12)-(5.13) corresponding to the vacuum solution (5.11).

Next, we will write the fields Ψ^\pm in terms of the relevant tau functions. According to the development in section 3 these functions will be written as certain matrix elements in an integrable highest weight representation of the affine Lie algebra $\widehat{sl}(2)$. So, it is a straightforward task to get the following relationships

$$\Psi^+ = \rho^+ + \frac{\hat{\tau}^+}{\hat{\tau}_0} \quad \text{and} \quad \Psi^- = \rho^- - \frac{\hat{\tau}^-}{\hat{\tau}_0}, \quad (5.16)$$

where the *tau* functions τ^\pm , τ_0 are given by

$$\tau^+ \equiv \langle \lambda_o | E_-^{(1)} \left[\Psi_{nvbc}^{(0)} h \Psi_{nvbc}^{(0)-1} \right]_{(-1)} | \lambda_o \rangle, \quad (5.17)$$

$$\tau^- \equiv \langle \lambda_o | E_+^{(1)} \left[\Psi_{nvbc}^{(0)} h \Psi_{nvbc}^{(0)-1} \right]_{(-1)} | \lambda_o \rangle, \quad (5.18)$$

$$\tau_0 \equiv \langle \lambda_o | \left[\Psi_{nvbc}^{(0)} h \Psi_{nvbc}^{(0)-1} \right]_{(o)} | \lambda_o \rangle, \quad (5.19)$$

with the group element $\Psi_{nvbc}^{(0)}$ given by (5.15).

5.1.1 AKNS₁ and 1-dark soliton of defocusing scalar NLS

Let us consider the group element

$$h^q = e^{\alpha^q V^q(\gamma_q, \rho)}, \quad (q = 1, 2) \quad (5.20)$$

where a^q is some constant.

The vertex operators V^1 and V^2 in (A.10), respectively, satisfy

$$[\varepsilon_1, V^q] = 2\gamma V^q, \quad [\varepsilon_2, V^q] = (-1)^{q-1} 2\gamma(\gamma^2 - \rho^2)^{1/2} V^q, \quad q = 1, 2. \quad (5.21)$$

Notice that these eigenstates V^q exhibit a more complex structure in comparison to their counterparts F_j and G_j corresponding to the VBC as in (4.3)-(4.4). Therefore one has

$$[x\varepsilon_1 + t\varepsilon_2, V^q(\gamma_q, \rho)] = [2\gamma_q(x + (-1)^{q-1}v_q t)] V^q(\gamma_q, \rho), \quad (5.22)$$

where $v_q = \sqrt{\gamma_q^2 - \rho^2}$. Denoting $\varphi_q = 2\gamma_q(x + (-1)^{q-1}v_q t)$ one can write

$$[\Psi_{nbc}^{(0)} h \Psi_{nbc}^{(0)-1}] = \exp(e^{\varphi_q} a_q V^q) \quad (5.23)$$

$$= 1 + e^{\varphi_q} a_q V^q, \quad (5.24)$$

where we have used $(V^q)^n = 0$, for $n \geq 2$, that is, V^q is nilpotent (A.15)-(A.16).

So, in order to find the explicit tau functions in (5.17)-(5.19) it remains to compute the relevant matrix elements. Using the properties presented in the Appendix A and the eqs. (3.26)-(3.28) one gets the tau functions $\tau_{(q)}^0 = 1 + c^{(q)} e^{\varphi_q}$; $\tau_{(q)}^\pm = s^\pm(\rho) \rho^\pm e^{\varphi_q}$, where the matrix elements $c^{(q)} = a^{(q)} \langle \lambda_o | V^1 | \lambda_o \rangle$, $s^\pm(\rho) = a^{(q)} (\rho^\pm)^{-1} \langle \lambda_o | E_\mp^{(1)} V^q | \lambda_o \rangle$ have been considered. They are given in the Appendix A, eqs. (A.11)-(A.12), respectively. Therefore, the relations (5.16) provide

$$\Psi^{\pm(q)}(x, t) = \rho^\pm [1 \pm \frac{s^\pm(q)}{2c^{(q)}}] \pm \frac{\rho^\pm s^\pm(q)}{2c^{(q)}} \tanh\{\gamma_q[x + (-1)^{q-1}v_q t] + \frac{\log c^{(q)}}{2}\}. \quad (5.25)$$

The solution in the case $q = 1$ allows the imposition of the conditions (5.10) in order to satisfy the defocusing NLS equation (5.6). So, let us make the transformation (5.8) and the complexification $\gamma_1 \rightarrow i\gamma$, $\rho \rightarrow i\rho$, $v_\gamma \rightarrow \bar{v}_\gamma = \sqrt{\rho^2 - \gamma^2}$; ($\rho > \gamma$). Choose $a_1 = -i\gamma/\bar{v}$. Moreover, considering the transformation (5.9) one gets the functions satisfying the complexification condition (5.10). Therefore, the function $\psi(x, t) \equiv -i\epsilon^2 \Psi^+$, where $\Psi^+ = i\epsilon^{-2} \{\rho + (\frac{i\gamma\rho}{\bar{v}_\gamma - i\gamma}) \frac{e^{\varphi(x, t)}}{1 + e^{\varphi(x, t)}}\}$, $\varphi(x, t) = 2\gamma(x - \bar{v}_\gamma t) + \gamma_0$, obtained in this way is a solution of the defocusing NLS equation (5.6). In fact, the function ψ can be written as

$$(x, t) = e^{i(\alpha + \pi/2)} \left\{ \frac{\gamma}{2} \tanh[\gamma(x - \bar{v}_\gamma t) + \gamma_0/2] + \frac{\gamma}{2} - i\rho e^{-i\alpha} \right\}, \quad \alpha = \tan^{-1}(\frac{\gamma}{\bar{v}_\gamma}), \quad (5.26)$$

which is the dark soliton solution of the defocusing NLS equation (5.6) (see e.g. [32]). Notice that this solution satisfy the NVBC $\lim_{x \rightarrow \pm\infty} |(x, t)|^2 \rightarrow \rho^2$.

Similarly, one can consider the case $q = 2$. In this case one gets $\Psi^\pm = i\epsilon^{\mp 2} \rho [\mp \frac{\gamma}{2\rho} e^{\pm i\delta} \tanh(\gamma x + \gamma \bar{v}_\gamma t) + 1 \mp \frac{\gamma}{2\rho} e^{\pm i\delta}]$, $\delta = \tan^{-1}(\bar{v}_\gamma/\gamma)$. However, they do not satisfy the condition (5.10), hence these functions would not be solutions of (5.6).

From the previous constructions it is clear that the two types (each one associated to the index $q = 1, 2$) of N-dark solitons of the AKNS₁ model will be associated to the group element

$$h^q = e^{a_1^1 V^q(\gamma_1^1, \rho)} e^{a_2^2 V^q(\gamma_2^2, \rho)} \dots e^{a_N^N V^q(\gamma_N^N, \rho)}, \quad q = 1, 2 \quad (5.27)$$

5.2 AKNS₂: N-dark-dark solitons

Consider the non-vanishing boundary condition (NVBC) for the system (2.5)-(2.6) in the form (3.1) and its associated constant vacuum solution (3.3). The vacuum connections (3.6)-(3.7) associated to this trivial solution commute $[A^{vac}, B^{vac}] = 0$ if one takes $\Omega_1 = \Omega_2 = \frac{2}{3}(\rho_1^+ \rho_1^- + \rho_2^+ \rho_2^-)$. Consider the notation $A^{vac} = \varepsilon_1$, $B^{vac} = \varepsilon_2$.

1-dark-dark soliton. Let us choose the group element $h^q = e^{a^q W^q(k)}$, a^q and k^q being some parameters, in the eqs. (3.26)-(3.28), such that

$$[\varepsilon_1, W^q] = k^q W^q; \quad q = 1, 2; \quad (5.28)$$

$$[\varepsilon_2, W^q] = w^q W^q \quad (5.29)$$

where

$$w^q = e_q k^q \sqrt{(k^q)^2 - 4\rho_1^+ \rho_1^- - 4\rho_2^+ \rho_2^-}, \quad e_q \equiv (-1)^{q-1} \quad (5.30)$$

So, one has

$$[x\varepsilon_1 + t\varepsilon_2, W^q] = (k^q x + w^q t)W^q \quad (5.31)$$

The vertex operator W^q (see (C.7)) is associated to the 1-dark-dark soliton and it can be shown to be nilpotent. So, using (3.26)-(3.28) the following tau functions correspond to the element h^q given above

$$\tau_0^{(q)} = 1 + c^{(q)} e^{k^q x + w^q t}, \quad (5.32)$$

$$\tau_i^{\pm(q)} = \rho_i^{\pm} s^{\pm(q)} e^{k^q x + w^q t}, \quad i = 1, 2; \quad (5.33)$$

where $c^{(q)} = a^q < \lambda_o |W^q| \lambda_o >$, $s^{\pm(q)} = (\rho_i^{\pm})^{-1} a^q < \lambda_o |E_{\mp\beta_i}^{(1)} W^q| \lambda_o >$. The above tau functions provide a 1-dark-dark solution for the fields Ψ_i^{\pm} of the system (2.5)-(2.6), such that the following relationships hold between the parameters

$$c^{(q)} = \frac{s^{+(q)} s^{-(q)}}{s^{+(q)} - s^{-(q)}}; \quad (k^q)^2 = -\frac{(s^{+(q)} - s^{-(q)})^2 (\sum_k \rho_k^+ \rho_k^-)}{s^{+(q)} s^{-(q)}}; \quad w^q = \frac{s^{+(q)} + s^{-(q)}}{s^{+(q)} - s^{-(q)}} (k^q)^2. \quad (5.34)$$

Then, the relations (3.25) provide

$$\Psi_j^{\pm(q)} = \frac{\rho_j^{\pm}}{2} \left[\left(1 + (y^q)^{\pm 1} \right) + \left(-1 + (y^q)^{\pm 1} \right) \tanh\{(k^q x + w^q t + \log c^{(q)})/2\} \right], \quad (5.35)$$

where $j = 1, 2$; $y^q \equiv \frac{s^{+(q)}}{s^{-(q)}}$. The condition $c^{(q)} > 0$ requires that $y^q > 1$ and $y^q < 0$.

Taking into account (C.7) one can compute the relevant matrix elements, which define $s^{\pm(q)}$, in order to get

$$y^q = -\frac{k^q + e_q \sqrt{(k^q)^2 - 4 \sum_j \rho_j^+ \rho_j^-}}{k^q - e_q \sqrt{(k^q)^2 - 4 \sum_j \rho_j^+ \rho_j^-}}. \quad (5.36)$$

Notice that for any value of the index q , if $\sum_j \rho_j^+ \rho_j^- > 0$ the last relationship implies $y^q < 0$ and if $\sum_j \rho_j^+ \rho_j^- < 0$ then $y^q > 0$.

Some remarks are in order here.

First, the solitons associated to the pair of components (Ψ_1^+, Ψ_2^+) and (Ψ_1^-, Ψ_2^-) , respectively, are proportional, so they are degenerate and reducible to the single dark soliton of the $sl(2)$ AKNS₁ model (5.25). Whereas, the solitons associated to the pair of components (Ψ_i^+, Ψ_j^-) , respectively, are not proportional, so they are non-degenerate presenting in general different degrees of ‘darkness’ in each component. In this way the solution in (5.35) presents a non-degenerate 1-dark-dark soliton in the components, say (Ψ_1^+, Ψ_2^-) of the AKNS₂ model. A nondegenerate 1-dark-dark soliton in the defocusing 2-CNLS has been given in [23, 10] through the Hirota’s method. The inverse scattering [12, 13] and Hirota’s method [10] results on N-dark-dark solitons in the defocusing 2-CNLS model have presented only the degenerate case.

Second, as in the AKNS₁ solution (5.25), one can show that the 1-dark-dark solitons (5.35) allow the imposition of the conditions (2.8) for the pair of fields (Ψ_i^+, Ψ_i^-) , $i = 1, 2$, in order to satisfy the 2-CNLS equation (2.9). In fact, consider the parameters $y^q = \frac{z^q+1}{z^q-1}$ and a complexification procedure as in (2.8) provided with a convenient set of complex parameters in (5.35), $k^q \rightarrow ik^q$, $\rho_i^\pm \rightarrow i\rho_i^\pm$ such that $y^q = e^{2i\phi_q}$, $z^q = \frac{e_q}{ik^q} \sqrt{4 \sum_j \rho_j^+ \rho_j^- - (k^q)^2}$, ϕ_q being real, so the equation (2.8) can be satisfied if

$$(\rho_j^+)^* = \mu \sigma_j \rho_j^-. \quad (5.37)$$

Notice that, as it was shown in the AKNS₁ case and its 1-dark soliton (5.26) particular reduction, in order to get 2-CNLS 1-dark solitons one must consider $q = 1$ in the both components (Ψ_j^+, Ψ_j^-) . Therefore, the relationship between the parameters ρ_i^\pm determine clearly if these solutions will correspond to the defocusing ($\sigma_j = -1$) or to the mixed nonlinearity ($\sigma_1 = -\sigma_2 = \pm 1$) 2-CNLS system, respectively.

Third, let us discuss the extension of our results to the case of free field NVBC (3.29). In order to compute the relevant tau functions analog to the ones in (3.26)-(3.28) one must consider the expression (3.35) instead of (3.14). So, the tau functions become

$$\hat{\tau}_i^{+g}(x, t) \equiv \langle \lambda_o | E_{-\beta_i}^{(1)} \Theta_-^g(x, t) \Psi^{(0)} h[\Psi^{(0)}]^{-1} | \lambda_o \rangle e^{-\chi}, \quad (5.38)$$

$$= (\hat{\rho}_i^+(x, t) - \rho_i^+) \hat{\tau}_0^g(x, t) + \langle \lambda_o | [e^{-\chi_o} E_{-\beta_i}^{(1)} e^{\chi_o}] \Psi^{(0)} h[\Psi^{(0)}]^{-1} | \lambda_o \rangle \quad (5.39)$$

$$\hat{\tau}_i^{-g}(x, t) \equiv \langle \lambda_o | E_{\beta_i}^{(1)} \Theta_-^g(x, t) \Psi^{(0)} h[\Psi^{(0)}]^{-1} | \lambda_o \rangle e^{-\chi}, \quad (5.40)$$

$$= (-\hat{\rho}_i^-(x, t) + \rho_i^-) \hat{\tau}_0^g(x, t) + \langle \lambda_o | [e^{-\chi_o} E_{\beta_i}^{(1)} e^{\chi_o}] \Psi^{(0)} h[\Psi^{(0)}]^{-1} | \lambda_o \rangle \quad (5.41)$$

$$\hat{\tau}_0^g(x, t) \equiv \langle \lambda_o | \Psi^{(0)} h[\Psi^{(0)}]^{-1} | \lambda_o \rangle. \quad (5.42)$$

where the group element $\Theta_-^g(x, t)$ is defined in (3.31)-(3.32) and $\chi(x, t)$ is an ordinary function. We have used Θ_+^g from eq. (3.32) and the decomposition (3.33) for χ_o , as well as the properties (B.15)-(B.16) of the highest weight representation. The tau functions (5.38)-(5.42) exhibit the modifications to be done on the previous eqs. (3.26)-(3.28) in order to satisfy the NVBC (3.29). These amount to introduce the group element Θ_-^g and the factor $e^{-\chi}$. Then from (5.38)-(5.42) and the analog to (3.25) $\Psi_i^+ = \rho_i^+ + \frac{\hat{\tau}_i^{+g}}{\hat{\tau}_0^g}$ and $\Psi_i^- = \rho_i^- - \frac{\hat{\tau}_i^{-g}}{\hat{\tau}_0^g}$ one can get

$$\Psi_i^+ = \hat{\rho}_i^+(x, t) + \frac{\hat{\tau}_i^+}{\hat{\tau}_0}; \quad \Psi_i^- = \hat{\rho}_i^-(x, t) - \frac{\hat{\tau}_i^-}{\hat{\tau}_0} \quad (5.43)$$

$$\hat{\tau}_{\pm} \equiv \langle \lambda_o | [e^{-\chi_o} E_{-\beta_i}^{(1)} e^{\chi_o}] \Psi^{(0)} h[\Psi^{(0)}]^{-1} | \lambda_o \rangle; \quad \hat{\tau}_0^g \equiv \hat{\tau}_0 = \langle \lambda_o | \Psi^{(0)} h[\Psi^{(0)}]^{-1} | \lambda_o \rangle \quad (5.44)$$

In fact, the relations (5.43) can be formally obtained from the ones in (3.25) by making the changes $\rho_i^{\pm} \rightarrow \hat{\rho}_i^{\pm}(x, t)$, $\tau_i^{\pm} \rightarrow \hat{\tau}_i^{\pm}$, $\tau_0 \rightarrow \hat{\tau}_0$. The soliton type solutions can be obtained following similar steps as the previous ones; i.e. for 1-soliton choose $h^q = e^{a^q W^q(k)}$ and the adjoint eigenvectors of $\varepsilon_{1,2}$ become precisely the vertex operator W^q as in (5.28)-(5.29), then the expressions $(k^q x + w^q t)$ of the tau functions (5.32)-(5.33) do not change.

Next, let us describe the general properties of the solutions (5.43)-(5.44). i) Notice that the form of $\hat{\tau}_0$ remains the same as the one in (5.32). ii) The modified DT and the related tau functions $\hat{\tau}_{\pm}$ (5.44) will introduce some new parameters into the tau functions τ_i^{\pm} in (5.33). In (5.33) the factor ρ_i^{\pm} must be replaced by $\hat{\rho}_i^{\pm}(x, t)$ and $s^{\pm(q)}$ in general will depend on the index 'i' and the space-time, i.e. $s^{\pm(q)} \rightarrow s^{\pm(q,i)}(x, t)$. The explicit form one gets from $s^{\pm(q,i)}(x, t) = (\hat{\rho}_i^{\pm}(x, t))^{-1} a^q < \lambda_o | [e^{-\chi_o} E_{\mp\beta_i}^{(1)} e^{\chi_o}] W^q | \lambda_o >$. In the case of 1-dark-dark-soliton it is required an element χ_o such that $s^{\pm(q,i)}(x, t) \equiv s^{\pm(q)} \chi_i^{\pm(q)}$, $s^{\pm(q)}$, $\chi_i^{\pm(q)}$ being constant parameters. So, under the above modified DT one has the general form of the 1-dark-dark soliton

$$\Psi_j^{\pm(q)} = \frac{\rho_j^{\pm} e^{a_j^{\pm} x + b_j^{\pm} t}}{2} \left[\left(1 + (y_j^q)^{\pm 1} \right) + \left(-1 + (y_j^q)^{\pm 1} \right) \tanh\{(k^q x + w^q t + \log c^{(q)})/2\} \right], \quad (5.45)$$

where $j = 1, 2$; $y_j^q \equiv \frac{s^{+(q,j)}}{s^{-(q,j)}}$, $a_j^+ = -a_j^-$, $b_j^+ = -b_j^-$ and

$$b_i^{\pm} = \pm(a_i^{\pm})^2 \mp 2 \left[\sum_{j=1}^2 \rho_j^+ \rho_j^- - \frac{1}{2}(\beta_i \cdot \vec{\Omega}) \right]; \quad \beta_1 \cdot \vec{\Omega} \neq \beta_2 \cdot \vec{\Omega}. \quad (5.46)$$

The solutions (5.45) are the general 1-dark-dark solitons of the AKNS₂ model and the two versions $q = 1, 2$, to our knowledge, have never been reported yet. A complexified version of this soliton for $q = 1$, which is a solution of the 2-CNLS model, has recently been reported in [4, 14].

2-dark-dark solitons. In order to get the two-dark-dark soliton ($N = 2$) solution one must choose the group element $h_2^q = e^{a_1 W^q(k_1)} e^{a_2 W^q(k_2)}$. Then, following similar steps one has

$$\tau_0 = 1 + c_1 e^{k_1 x + w_1 t} + c_2 e^{k_2 x + w_2 t} + c_1 c_2 d_0 e^{k_1 x + w_1 t} e^{k_2 x + w_2 t}, \quad (5.47)$$

$$\tau_i^{\pm} = \rho_i^{\pm} \left[s_1^{\pm} e^{k_1 x + w_1 t} + s_2^{\pm} e^{k_2 x + w_2 t} + s_1^{\pm} s_2^{\pm} d^{\pm} e^{k_1 x + w_1 t} e^{k_2 x + w_2 t} \right], \quad i = 1, 2 \quad (5.48)$$

where

$$d_0 = \left(\frac{\sqrt{s_1^+ s_2^-} - \sqrt{s_1^- s_2^+}}{\sqrt{s_1^- s_2^-} - \sqrt{s_1^+ s_2^+}} \right)^2; \quad d^{\pm} = -\frac{(s_1^- s_2^- - s_1^+ s_2^+)}{(s_1^- - s_1^+)(s_2^- - s_2^+)} d_0; \quad (5.49)$$

$$c_n = \frac{s_n^+ s_n^-}{s_n^+ - s_n^-}; \quad (k_n)^2 = -\frac{(s_n^+ - s_n^-)^2 (\sum_{k=1}^2 \rho_k^+ \rho_k^-)}{s_n^+ s_n^-}; \quad w_n = \frac{s_n^+ + s_n^-}{s_n^+ - s_n^-} k_n^2, \quad (n = 1, 2) \quad (5.50)$$

The above process can be extended for N -dark-dark soliton solutions, in that case the group element in (3.26)-(3.28) must take the form $h_2^q = e^{a_1 W_1^q(k_1)} e^{a_2 W_2^q(k_2)} \dots e^{a_N W_N^q(k_N)}$. The

relevant tau functions follow

$$\tau_0 = \left| \delta_{mn} + \frac{s_m^+ s_n^-}{\sqrt{s_m^+ s_n^+} - \sqrt{s_m^- s_n^-}} e^{k_n x + w_n t} \right|_{N \times N}; \quad m, n = 1, 2, 3 \dots N \quad (5.51)$$

$$\tau_i^\pm = \rho_i^\pm \left\{ \left| \delta_{mn} + s_n^\pm \sqrt{1 + c_{mn}} e^{k_n x + w_n t} \right|_{N \times N} - 1 \right\}, \quad i = 1, 2. \quad (5.52)$$

where $|\cdot|_{N \times N}$ stands for determinant, $c_{mn} = \frac{s_m^- s_n^- - s_m^+ s_n^+}{(s_m^- - s_m^+)(s_n^- - s_n^+)} \left(\frac{\sqrt{s_m^+ s_n^-} - \sqrt{s_m^- s_n^+}}{\sqrt{s_m^- s_n^-} - \sqrt{s_m^+ s_n^+}} \right)^2$, and w_n, k_n are the same as in (5.50). Notice that $c_{mn} = c_{nm}$, $c_{nn} = 0$. We have omitted the index (q) in all the parameters above. Let us mention that we have verified these solutions up to $N = 3$ using the MATHEMATICA program. The above solutions associated to the two (5.47)-(5.48) or higher order (5.51)-(5.52) dark-dark-solitons are in general non-degenerate. Notice that in the CNLS case the two- and higher-dark-dark solitons derived in [10] are actually degenerate and reducible to scalar NLS dark solitons. So, regarding this property our solutions resemble to the ones recently obtained in [4, 14] for the r-CNLS system.

Let us remark that the free field NVBC (3.29) requires the introduction of more parameters into the N-dark-dark tau functions (5.51)-(5.52) through the modified DT. So, following similar steps to get (5.45) one has: i) the parameters s_n^\pm entering the tau function τ_0 in (5.51) remain the same. ii) in (5.52) the factor ρ_j^\pm must be changed to $\rho_j^\pm e^{a_j^\pm x + b_j^\pm t}$. iii) the parameters s_n^\pm entering the tau functions τ_i^\pm in (5.52) must be changed as $s_n^\pm \rightarrow s_n^\pm \chi_i^\pm$.

So far, to our knowledge, general multi-dark-dark solitons in the AKNS model incorporating the cases $q = 1, 2$ have never been reported yet. In [4] the general non-degenerate N-dark-dark solitons in the r-CNLS model with mixed nonlinearity has been reported in the context of the KP-hierarchy reduction approach, and in [14] the r-CNLS model has been addressed in the algebro-geometric approach.

5.3 Dark-dark soliton bound states

So far, reports on multi-dark-dark soliton bound states in integrable systems, to our knowledge, are very limited. Recently, it has been shown that in the mixed-nonlinearity case of the 2-CNLS system, two dark-dark solitons can form a stationary bound state [4]. Then, in order to have multi-dark-dark soliton bound states in the N-soliton solution (5.51)-(5.52) the constituent solitons should have the same velocity, i.e. denoting $y_n \equiv \frac{s_n^+}{s_n^-}$, ($y_n \notin [0, 1]$) then $\frac{w_n}{k_n} = \frac{y_n + 1}{y_n - 1} k_n \equiv v$ (assume $v > 0$) for certain soliton parameters labeled by $n = 1, 2, \dots$. We show that the signs of the sum $\sum_j \rho_j^+ \rho_j^-$ determine the existence of these bound states, for the positive sign bound states can be formed, whereas for the negative sign they do not exist.

First, consider $\sum_i \rho_i^+ \rho_i^- > 0$, then from (5.50) it follows that $k_n = \frac{|y_n| - 1}{\sqrt{|y_n|}} (\sum_i \rho_i^+ \rho_i^-)^{1/2}$ for $y_n < 0$. Therefore, in order for multi-dark-dark soliton bound states to exist the equation

$$\frac{(|y_n| - 1)^2}{(|y_n| + 1)\sqrt{|y_n|}} = \frac{v}{\sqrt{\sum_i \rho_i^+ \rho_i^-}} \equiv c > 0 \quad (5.53)$$

must give at least two distinct positive solutions for $|y_n|$. In fact, one has the following two solutions

$$|y_{1,2}| = 1 + \frac{c^2}{4} + c \frac{\sqrt{16+c^2}}{4} \pm \frac{\sqrt{12c^2+c^4+\Delta}}{2\sqrt{2}}, \quad (5.54)$$

where $\Delta \equiv \frac{64c+20c^3+c^5}{\sqrt{16+c^2}}$. These exhaust all the possible solutions with the condition (5.53). The right hand side of (5.54) with either + or - signs provides $|y_{1,2}| > 0$ for any $c > 0$. The other possible solutions $y_{3,4}$ do not satisfy the above requirements, since they possess a term of the form $\sqrt{12c^2+c^4-\Delta}$, which is imaginary for any positive value c . Therefore, two-dark-dark-soliton bound states exist in the $sl(3)$ -AKNS system, and three- and higher-dark-dark-soliton bound states can not exist. These results hold for any value of the index q . Notice that in order to reduce to the 2-CNLS system the parameter relationship (5.37) must be satisfied. So, from (5.37) one has $\sum_i \rho_i^+ \rho_i^- = \frac{1}{\mu} [\frac{|\rho_1^+|^2}{\sigma_1} + \frac{|\rho_2^+|^2}{\sigma_2}] < 0$ (remember that under $\rho_i^\pm \rightarrow i\rho_i^\pm$ the sum $\sum_i \rho_i^+ \rho_i^-$ reverses sign); so, it is possible if $\sigma_1 = -\sigma_2$ (e.g. $\sigma_1 = -\sigma_2 = -1$ for $|\rho_1^+| > |\rho_2^+|$), this case corresponds to the 2-CNLS with mixed focusing and defocusing nonlinearities. Thus, the two-dark-dark soliton bound state solution we have obtained here corresponds to the Manakov model with mixed nonlinearity. The case $\sigma_i = -1$ ($i = 1, 2$) defines the defocusing Manakov model which does not support multi-dark-dark-soliton bound states [4].

Second, the condition $\sum_i \rho_i^+ \rho_i^- < 0$ will provide only a single positive solution for $y^q > 0$ in the equation $\frac{w_n}{k_n} = v > 0$. So, one can not obtain two or more solitons with the same velocity and therefore bound states in this case are not possible.

6 Mixed boundary conditions and dark-bright solitons

One may ask about the mixed boundary conditions for the system (2.5)-(2.6), i.e. NVBC for one of the field components of the system, say Ψ_1^\pm , and VBC for the other field component, Ψ_2^\pm . So, we will deal with the mixed boundary condition (3.4). Moreover, taking into account sequence of conditions as in (2.8)

$$t \rightarrow -it, \quad [\Psi_i^+]^* = -\mu \sigma_i \Psi_i^- \equiv -\mu \sigma_i \psi_i, \quad (6.1)$$

the system (2.5)-(2.6), for $(\beta_i, \vec{\rho})$ being real, can be written as

$$i \partial_t \psi_k + \partial_x^2 \psi_k + 2\mu \left[\sum_{j=1}^2 \sigma_j |\psi_j|^2 - \frac{1}{2} (\beta_k \cdot \vec{\rho}) \right] \psi_k = 0, \quad k = 1, 2. \quad (6.2)$$

Precisely, the system (6.2) in the defocusing case ($\sigma_1 = \sigma_2 = -1$) and possessing the first trivial solution (3.4), i.e. $\rho_1 \neq 0$; $\rho_2^\pm = 0$, has been considered in [27] in order to investigate dark-bright solitons which describe an inhomogeneous two-species Bose-Einstein condensate. The system (2.5)-(2.6) with the mixed trivial solution (3.4) can be written as

$$\partial_t \Psi_1^\pm = \pm \partial_x^2 \Psi_1^\pm \mp 2 \left[\sum_{j=1}^2 \Psi_j^+ \Psi_j^- - \rho_1^+ \rho_1^- \right] \Psi_1^\pm, \quad (6.3)$$

$$\partial_t \Psi_2^\pm = \pm \partial_x^2 \Psi_2^\pm \mp 2 \left[\sum_{j=1}^2 \Psi_j^+ \Psi_j^- - \frac{1}{2} \rho_1^+ \rho_1^- - \frac{3}{2} \Omega_2 \right] \Psi_2^\pm. \quad (6.4)$$

From the previous sections we can make the following observation: the vacuum connections relevant to each type of solutions must exhibit the fact that dark solitons are closely related to NVBC, whereas bright solitons to the relevant VBC one. Since Ω_2 and Ω_1 are some parameters satisfying $2\Omega_1 + \Omega_2 = 2\rho_1^+ \rho_1^-$, for simplicity we will assume $\Omega_2 = 0$, $\Omega_1 = \rho_1^+ \rho_1^-$ in (3.6)-(3.7). Therefore, the connections (3.6)-(3.7) for the mixed (constant-zero) boundary condition (3.4) take the form

$$\hat{A}^{vac} \equiv \tilde{\varepsilon}_{\beta_1}^1 = E^{(1)} + \rho_1^+ E_{\beta_1}^{(0)} + \rho_1^- E_{-\beta_1}^{(0)}, \quad (6.5)$$

$$\hat{B}^{vac} \equiv \tilde{\varepsilon}_{\beta_1}^2 = E^{(2)} + \rho_1^+ E_{\beta_1}^{(1)} + \rho_1^- E_{-\beta_1}^{(1)} \quad (6.6)$$

The relevant group element $\Psi_{mbc}^{(0)}$ is given by

$$\Psi_{mbc}^{(0)} \equiv e^{x\tilde{\varepsilon}_{\beta_1}^1 + t\tilde{\varepsilon}_{\beta_1}^2}, \quad (6.7)$$

In order to construct the soliton solutions we must look for the common eigenstates of the adjoint action of the vacuum connections (6.5)-(6.6). So, one has

$$[\tilde{\varepsilon}_{\beta_1}^1, \Gamma_{\pm\beta_2}] = \pm k^\pm \Gamma_{\pm\beta_2}, \quad (6.8)$$

$$[\tilde{\varepsilon}_{\beta_1}^2, \Gamma_{\pm\beta_2}] = \pm w^\pm \Gamma_{\pm\beta_2}, \quad w^\pm = (k^\pm)^2 - \rho_1^+ \rho_1^- \quad (6.9)$$

where the vertex operators $\Gamma_{\pm\beta_2}(k^\pm)$ are defined in (C.3). We expect that these vertex operators will be associated to the bright-dark soliton solutions of the model.

Let us write the following expressions

$$[x\tilde{\varepsilon}_{\beta_1}^1 + t\tilde{\varepsilon}_{\beta_1}^2, \Gamma_{\beta_2}] = \varphi^+(x, t) \Gamma_{\beta_2}; \quad \varphi^+(x, t) = k^+ x + w^+ t; \quad (6.10)$$

$$[x\tilde{\varepsilon}_{\beta_1}^1 + t\tilde{\varepsilon}_{\beta_1}^2, \Gamma_{-\beta_2}] = -\varphi^-(x, t) \Gamma_{-\beta_2}; \quad \varphi^-(x, t) = k^- x + w^- t. \quad (6.11)$$

The 1-dark-bright soliton is constructed taking h in (3.26)-(3.28) as

$$h = e^{\gamma^+ \Gamma_{+\beta_2}} e^{\gamma^- \Gamma_{-\beta_2}}, \quad \gamma^\pm = \text{constants}, \quad (6.12)$$

where the vertex operators in (C.3) have been considered. Notice that due to the nilpotency property of the vertex operators, as presented in the appendix C, the exponential series must truncate. So, replacing the group element (6.12) in (3.26)-(3.28) one has the following tau functions ²

$$\bar{\tau}_1^\pm = e^{\varphi^+ - \varphi^-} \gamma^+ \gamma^- \langle \lambda_o | E_{\mp\beta_1}^{(1)} \Gamma_{+\beta_2} \Gamma_{-\beta_2} | \lambda_o \rangle \quad (6.13)$$

$$\bar{\tau}_2^\pm = e^{\pm\varphi^\pm} \gamma^\pm \langle \lambda_o | E_{\mp\beta_2}^{(1)} \Gamma_{\pm\beta_2} | \lambda_o \rangle \quad (6.14)$$

$$\bar{\tau}_0 = 1 + e^{\varphi^+ - \varphi^-} \gamma^+ \gamma^- \langle \lambda_o | \Gamma_{+\beta_2} \Gamma_{-\beta_2} | \lambda_o \rangle \quad (6.15)$$

²There exist other eigenstates $[\tilde{\varepsilon}_{\beta_1}^1, V_{\beta_1}^q] = \lambda V_{\beta_1}^q$; $[\tilde{\varepsilon}_{\beta_1}^2, V_{\beta_1}^q] = (-1)^{q-1} \lambda (\lambda^2 - \rho_0^2)^{1/2} V_{\beta_1}^q$, where $\rho_0^2 = 4\rho_1^+ \rho_1^-$ and $V_{\beta_1}^q$ ($q = 1, 2$) is given in (C.5). However, these eigenstates are related to purely dark-solitons for the first component Ψ_1^\pm , e.g. if one takes the group element $h = e^{V_{\beta_1}^q}$, it will not excite the second component Ψ_2^\pm since the matrix elements of type $\langle \lambda_o | E_{\mp\beta_2}^{(1)} V_{\beta_1}^q | \lambda_o \rangle$ vanish.

These tau functions resemble the ones in (5.32)-(5.33) for the 1-dark soliton (5.35) and the eqs. (4.8)-(4.9) for the 1-bright soliton (4.10) components, respectively. The matrix elements in (6.13)-(6.15) can be computed, and will depend only on the parameters k^\pm, ρ_1^\pm . So, one has six independent parameters $\gamma^\pm, k^\pm, \rho_1^\pm$ associated to the these tau functions. Let us write the tau functions in the form $\bar{\tau}_0 = 1 + c_0 e^{k^+x+w^+t} e^{-k^-x-w^-t}$, $\bar{\tau}_1^\pm = a^\pm e^{k^+x+w^+t} e^{-k^-x-w^-t}$, $\bar{\tau}_2^\pm = b^\pm e^{\pm k^\pm x \pm w^\pm t}$, where $w^\pm = (k^\pm)^2 - \rho_1^\pm \rho_1^-$; $c_0 = \frac{a^+ a^-}{a^+ \rho_1^- - a^- \rho_1^+}$; $\frac{a^\pm}{a^\mp} = \frac{k^+ \rho_1^+}{k^- \rho_1^-}$; $b^+ b^- = (\frac{a^-}{\rho_1^-})(\frac{k^+}{k^-} - 1)(k^+ k^- + \rho_1^+ \rho_1^-)$. So, using these tau functions in (3.25) one has

$$\Psi_1^\pm = \rho_1^\pm \pm \left(\frac{a^\pm}{2c_0}\right) \left[1 + \tanh \frac{(k^+ - k^-)x + (w^+ - w^-)t}{2}\right] \quad (6.16)$$

$$\Psi_2^\pm = \frac{b^\pm}{2\sqrt{c_0}} e^{\pm \frac{1}{2}[(k^+ + k^-)x + (w^+ + w^-)t]} \operatorname{sech} \frac{(k^+ - k^-)x + (w^+ - w^-)t}{2}. \quad (6.17)$$

This is the 1-dark-bright soliton of the $\hat{sl}(3)$ AKNS model (6.3)-(6.4) (for $\Omega_2 = 0$). Notice that this solution has six independent complex parameters, say $a^\pm, b^\pm, k^\pm, \rho_1^\pm$.

The construction of the 2-dark-bright solitons follows similar steps. The group element $h = e^{\gamma_1^+ \Gamma + \beta_2(k_1^+)} e^{\gamma_1^- \Gamma - \beta_2(k_1^-)} e^{\gamma_2^+ \Gamma + \beta_2(k_2^+)} e^{\gamma_2^- \Gamma - \beta_2(k_2^-)}$ does the job. We record the relevant tau functions

$$\bar{\tau}_0 = 1 + \sum_{m,n=1}^2 c_{mn} e^{k_m^+ x + w_m^+ t} e^{-k_n^- x - w_n^- t} + c_0 e^{k_1^+ x + w_1^+ t} e^{-k_1^- x - w_1^- t} e^{k_2^+ x + w_2^+ t} e^{-k_2^- x - w_2^- t} \quad (6.18)$$

$$\bar{\tau}_1^\pm = \sum_{m,n=1}^2 a_{mn}^\pm e^{k_m^+ x + w_m^+ t} e^{-k_n^- x - w_n^- t} + d_1^\pm e^{k_1^+ x + w_1^+ t} e^{-k_1^- x - w_1^- t} e^{k_2^+ x + w_2^+ t} e^{-k_2^- x - w_2^- t} \quad (6.19)$$

$$\bar{\tau}_2^\pm = \sum_{m=1}^2 b_m^\pm e^{\pm k_m^\pm x \pm w_m^\pm t} + e^{\pm k_1^\pm x \pm w_1^\pm t} e^{\pm k_2^\pm x \pm w_2^\pm t} \left[\sum_{m=1}^2 q_m^\pm e^{\mp k_m^\mp x \mp w_m^\mp t} \right], \quad (6.20)$$

where $w_n^\pm = (k_n^\pm)^2 - \rho_1^\pm \rho_1^-$; $c_{mn} = \frac{a_{mn}^+ a_{mn}^-}{a_{mn}^+ \rho_1^- - a_{mn}^- \rho_1^+}$; $\frac{a_{mn}^\pm}{a_{mn}^\mp} = \frac{k_m^+ \rho_1^+}{k_n^- \rho_1^-}$; $c_0 = \frac{d_1^+ d_1^-}{d_1^+ \rho_1^- - d_1^- \rho_1^+}$; $\frac{d_1^\pm}{d_1^\mp} = \frac{k_1^+ k_2^+ \rho_1^+}{k_1^- k_2^- \rho_1^-}$.
 $\frac{a_{12}^-(k_1^+ - k_2^-)(k_1^+ k_2^- + \rho_1^+ \rho_1^-)}{a_{11}^-(k_1^+ - k_1^-)(k_1^+ k_1^- + \rho_1^+ \rho_1^-)} + \frac{a_{22}^-(k_2^+ - k_2^-)(k_2^+ k_2^- + \rho_1^+ \rho_1^-)}{a_{21}^-(k_1^+ - k_2^-)(k_1^+ k_2^- + \rho_1^+ \rho_1^-)} = 0$; $b_i^+ b_i^- = \frac{(k_i^+ - k_i^-)(k_i^+ k_i^- + \rho_1^+ \rho_1^-)}{k_i^- \rho_1^-} a_{ii}^-$;
 $d_1^- = -\frac{a_{12}^+ a_{21}^-(k_1^+ - k_2^-)^2 (k_1^+ - k_2^-)^2 (k_1^+ k_2^+ - k_1^- k_2^-)(k_1^+ k_2^+ + \rho_1^+ \rho_1^-)(k_1^- k_2^- + \rho_1^- \rho_1^-)}{(k_1^+ - k_1^-)^2 (k_1^+ - k_2^-)(k_1^+ - k_2^-)^2 \rho_1^- (k_1^+ k_1^- + \rho_1^+ \rho_1^-)(k_2^+ k_2^- + \rho_1^+ \rho_1^-)}$, $\frac{b_2^-}{b_1^-} = \frac{a_{12}^+ k_1^- (k_1^+ - k_2^-)(k_1^+ k_2^- + \rho_1^+ \rho_1^-)}{a_{11}^+ k_2^- (k_1^+ - k_1^-)(k_1^+ k_1^- + \rho_1^+ \rho_1^-)}$;
 $q_i^\pm = \mp \frac{a_{ii}^\pm (k_i^\pm)^2}{b_i^\pm k_1^\pm k_2^\pm (\rho_1^\pm)^2} \frac{1}{(k_i^\pm - k_i^\mp)(k_i^\mp - k_{3-i}^\pm)} (k_1^\pm k_2^\pm + \rho_1^\pm \rho_1^-) a_{m_i^\pm n_i^\pm}^\pm$; $m_i^+ = n_i^- = 3 - i$; $m_i^- = n_i^+ = i$.

These tau functions resemble to the ones in [25] provided for the 2-CNLS model. The generalization to N-dark-bright solitons require the group element h to be

$$h = e^{\gamma_1^+ \Gamma + \beta_2(k_1^+)} e^{\gamma_1^- \Gamma - \beta_2(k_1^-)} e^{\gamma_2^+ \Gamma + \beta_2(k_2^+)} e^{\gamma_2^- \Gamma - \beta_2(k_2^-)} \dots e^{\gamma_N^+ \Gamma + \beta_2(k_N^+)} e^{\gamma_N^- \Gamma - \beta_2(k_N^-)}. \quad (6.21)$$

7 Discussion

We have considered soliton type solutions of the AKNS model supported by the various boundary conditions (3.2)-(3.5): vanishing, (constant) non-vanishing and mixed vanishing-nonvanishing boundary conditions related to bright, dark and bright-dark soliton solutions,

respectively, by applying the DT approach as presented in [16]. The set of solutions of the AKNS_r system (2.5)-(2.6) is much larger than the solutions of the r-CNLS system (2.9). A subset of solutions of the AKNS_r system, (2.5)-(2.6) for $r = 2$ and (5.3)-(5.4) for $r = 1$, respectively, solve the scalar NLS (5.6) and 2-CNLS system (2.9), under relevant complexifications.

Moreover, the free field boundary condition (3.29) for dark solitons is considered in the context of a modified DT approach associated to the dressing group [30], and the general N-dark-dark soliton solutions of the AKNS₂ system have been derived. These soliton components are not proportional to each other and thus they do not reduce to the AKNS₁ solitons, in this sense they are not degenerate. We showed that these solitons under convenient complexifications reduce to the general N-dark-dark solitons derived previously in the literature for the CNLS model [4, 14]. In addition, we have shown that two-dark-dark-soliton bound states exist in the $sl(3)$ -AKNS system, and three- and higher-dark-dark-soliton bound states can not exist. These results hold for any value of the index q . In the case of reduced 2-CNLS when focusing and defocusing nonlinearities are mixed, this result corresponds to 2-dark-dark soliton stationary bound state [4].

In the mixed constant boundary conditions we derived the dark-bright solitons of the $\hat{sl}(3)$ AKNS model. These solitons under the complexification (6.1) reduce to the solitons of the 2-CNLS model (6.2) which will be useful in order to investigate dark-bright solitons appearing in an inhomogeneous two-species Bose-Einstein condensate [27].

Another point we should highlight relies upon the possible relevance of the CNLS tau functions to its higher-order generalizations. We expect that the tau functions of the higher-order CNLS generalization are related somehow to the basic tau functions of the usual CNLS equations. This fact is observed for example in the case of the coupled scalar NLS+derivative-NLS system in which the coupled system possesses a composed tau function depending on the basic scalar NLS tau functions [41].

8 Acknowledgments

HB has been partially supported by CNPq. AOA thanks the support of the brazilian CAPES.

In memory of our friend S. S. Costa .

A $\hat{sl}(2)$ matrix elements

The commutation relations for the $\hat{sl}(2)$ affine KacMoody algebra elements are

$$[H^{(m)}, H^{(n)}] = 2m\delta_{m+n,0}C, \quad (A.1)$$

$$[H^{(m)}, E_{\pm}^{(n)}] = \pm 2E_{\pm}^{(m+n)}, \quad (A.2)$$

$$[E_{+}^{(m)}, E_{-}^{(n)}] = H^{(m+n)} + 2m\delta_{m+n,0}C; \quad (A.3)$$

$$[D, T_a^{(m)}] = mT_a^{(m)}; \quad T_a^{(m)} = \{H^{(m)}, E_{\pm}^{(m)}\} \quad (A.4)$$

The central extension ensures highest weight representations (h.w.r.) of the affine algebra (see e.g. [16]). So, in the h.w.r. $\{|\lambda_0\rangle, |\lambda_1\rangle\}$ one has the following relationships

$$E_+^{(0)}|\lambda_a\rangle = 0 \quad (\text{A.5})$$

$$E_{\pm}^{(m)}|\lambda_a\rangle = 0, \quad m > 0 \quad (\text{A.6})$$

$$H^{(m)}|\lambda_a\rangle = 0, \quad m > 0; \quad (\text{A.7})$$

$$H^{(0)}|\lambda_a\rangle = \delta_{a,1}|\lambda_a\rangle, \quad (\text{A.8})$$

$$C|\lambda_a\rangle = |\lambda_a\rangle \quad (\text{A.9})$$

where $a = 0, 1$. The adjoint relations $(E_{\pm}^{(m)})^{\dagger} = E_{\mp}^{(-m)}$, $(H^{(m)})^{\dagger} = H^{(-m)}$ allow one to know their actions on the $\langle\lambda_a|$. Next, consider the vertex operators

$$V^q(\gamma, \rho) = \sum_{n=-\infty}^{\infty} \{(\gamma^2 - \rho^2)^{-n/2} e_q \left[H^{(n)} - \frac{\rho^+}{\gamma - e_q(\gamma^2 - \rho^2)^{1/2}} E_+^{(n)} + \frac{\rho^-}{\gamma + e_q(\gamma^2 - \rho^2)^{1/2}} E_-^{(n)} \right] + e_q \left(\frac{\gamma_-^2 - \rho_-^2}{\gamma^2} \right)^{1/2} \delta_{n,0} C\}; \quad q = 1, 2; \quad (\text{A.10})$$

where $e_q \equiv (-1)^{q-1}$.

The following matrix elements can be computed using the properties (A.5)-(A.9)

$$\langle\lambda_0| V^q |\lambda_0\rangle, = e_q \frac{(\gamma^2 - \rho^2)^{1/2}}{\gamma}, \quad (\text{A.11})$$

$$\langle\lambda_0| E_{\mp}^{(1)} V^q |\lambda_0\rangle = \mp \frac{2\rho^{\pm}}{\gamma \mp e_q(\gamma^2 - \rho^2)^{1/2}} (\gamma^2 - \rho^2)^{1/2}. \quad (\text{A.12})$$

The matrix element $\langle\lambda_0| V^q(\gamma_1, \rho) V^q(\gamma_2, \rho) |\lambda_0\rangle$ can be computed by developing the products and keeping only non-trivial terms, then one makes use of the commutation rules to change the order, and eventually to get some central terms C . The double sum can be simplified to a single sum and each term can be substituted by power series like $\sum_{n=0}^{\infty} x^n = \frac{1}{(1-x)}$, $\sum_{n=1}^{\infty} x^n = \frac{x}{(1-x)}$, and $\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}$. So, one can get

$$\langle\lambda_0| V^q(\gamma_1, \rho) V^q(\gamma_2, \rho) |\lambda_0\rangle = [2 + 2K(\gamma_1, \gamma_2)] \frac{K_0(\gamma_1, \gamma_2)}{[1 - K_0(\gamma_1, \gamma_2)]^2} + \left[\frac{K(\gamma_1, \gamma_2)\rho^2}{\gamma_1\gamma_2} + 1 \right]; \quad q = 1, 2 \quad (\text{A.13})$$

$$\text{where } K_0(\gamma_1, \gamma_2) \equiv \frac{\sqrt{\gamma_2^2 - \rho^2}}{\sqrt{\gamma_1^2 - \rho^2}}; \quad K(\gamma_1, \gamma_2) \equiv \frac{\sqrt{\gamma_1^2 - \rho^2} \sqrt{\gamma_2^2 - \rho^2} - \gamma_1\gamma_2}{\rho^2} \quad (\text{A.14})$$

In order to prove the nilpotency property of the vertex operator $V^q(\gamma_1, \rho)$, when evaluated within the state $|\lambda_0\rangle$, it is convenient to write (A.13) in the following Laurent series expansion

$$\langle\lambda_0| V^q(\gamma_1, \rho) V^q(\gamma_2, \rho) |\lambda_0\rangle = -6\rho^2 \sqrt{\gamma_1^2 - \rho^2} \sqrt{\gamma_2^2 - \rho^2} \left(\frac{\sqrt{\gamma_1^2 - \rho^2} + \sqrt{\gamma_2^2 - \rho^2}}{\gamma_2^2 - \rho^2} \right)^2 \left(\frac{\gamma_1 - \gamma_2}{\gamma_1 + \gamma_2} \right)^2 \times$$

$$\left[\frac{1}{4!} - \frac{10\gamma_2}{(\gamma_2^2 - \rho^2)} \frac{(\gamma_1 - \gamma_2)}{5!} + \frac{15(6\gamma_2^2 + \rho^2)}{(\gamma_2^2 - \rho^2)^2} \frac{(\gamma_1 - \gamma_2)^2}{6!} - \frac{420\gamma_2(2\gamma_2^2 + \rho^2)}{(\gamma_2^2 - \rho^2)^3} \frac{(\gamma_1 - \gamma_2)^3}{7!} + \dots \right]. \quad (\text{A.15})$$

From this, it is clear that

$$\lim_{\gamma_1 \rightarrow \gamma_2} \langle \lambda_o | V^q(\gamma_1, \rho) V^q(\gamma_2, \rho) | \lambda_o \rangle \rightarrow 0, \quad q = 1, 2. \quad (\text{A.16})$$

B The affine Kac-Moody algebra $\widehat{sl}_3(C)$

In the following we provide some results about the affine Kac-Moody algebra $\mathcal{G} = \widehat{sl}_3(C)$ relevant to our discussions above. We follow closely [42, 19]. The elements of the $sl_3(C)$ Lie algebra are all 3×3 complex matrices with zero trace. Consider the corresponding root system $\Delta = \{\pm\alpha_1, \pm\alpha_2, \pm\alpha_3\}$, such that the three positive roots are α_i , $i = 1, 2, 3$, with α_a , $a = 1, 2$, being the simple roots and $\alpha_3 = \alpha_1 + \alpha_2$. We choose a standard basis for the Cartan subalgebra \mathcal{H} such that

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (\text{B.1})$$

and the generators of the root subspaces corresponding to the positive roots are chosen as

$$E_{+\alpha_1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad E_{+\alpha_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad E_{+\alpha_3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (\text{B.2})$$

For negative roots one has $E_{-\alpha} = (E_{+\alpha})^T$. The invariant bilinear form on $sl_3(C)$, $(x|y) = \text{tr}(xy)$; $x, y \in sl_3(C)$ induces a nondegenerate bilinear form on \mathcal{H}^* which we also denote by $(\cdot|\cdot)$. This definition allows one to write

$$(\alpha_1|\alpha_1) = 2, \quad (\alpha_2|\alpha_2) = 2, \quad (\alpha_1|\alpha_2) = -1. \quad (\text{B.3})$$

On the other hand, the generators $T^{(m)} \equiv \{H_1^{(m)}, H_2^{(m)}, E_\alpha^{(m)}\}$, where $m \in \mathbb{Z}$ and $\alpha \in \Delta$, together with the central C and the 'derivation' operator D ($[D, T^{(m)}] = mT^{(m)}$) form a basis for $\widehat{sl}_3(C)$. These generators satisfy the commutation relations

$$[H_a^{(m)}, H_b^{(n)}] = m(\alpha_a|\alpha_b) C \delta_{m+n,0}, \quad (\text{B.4})$$

$$[H_a^{(m)}, E_{\pm\alpha_i}^{(n)}] = \pm(\alpha_a|\alpha_i) E_{\pm\alpha_i}^{(m+n)}, \quad (\text{B.5})$$

$$[E_{\alpha_a}^{(m)}, E_{-\alpha_a}^{(n)}] = H_a^{(m+n)} + m C \delta_{m+n,0}, \quad (\text{B.6})$$

$$[E_{\alpha_3}^{(m)}, E_{-\alpha_3}^{(n)}] = H_1^{(m+n)} + H_2^{(m+n)} + m C \delta_{m+n,0}, \quad (\text{B.7})$$

$$[E_{\alpha_1}^{(m)}, E_{\alpha_2}^{(n)}] = E_{\alpha_3}^{(m+n)}, \quad (\text{B.8})$$

$$[E_{\alpha_3}^{(m)}, E_{-\alpha_1}^{(n)}] = -E_{\alpha_2}^{(m+n)}, \quad (\text{B.9})$$

$$[E_{\alpha_3}^{(m)}, E_{-\alpha_2}^{(n)}] = E_{\alpha_1}^{(m+n)}, \quad (\text{B.10})$$

$$[D, C] = 0, \quad (\text{B.11})$$

where $a, b = 1, 2$, $i = 1, 2, 3$ and $m, n \in \mathbb{Z}$. The remaining non-vanishing commutation relations are obtained by using the relation $[E_{\alpha}^{(m)}, E_{\beta}^{(n)}]^{\dagger} = -[E_{-\alpha}^{(-m)}, E_{-\beta}^{(-n)}]$.

In this paper we use the homogeneous \mathbb{Z} -gradation of $\hat{sl}_3(C)$ which is defined by the grading operator D , such that

$$\hat{sl}_3(C) = \bigoplus_{m \in \mathbb{Z}} \mathcal{G}_m, \quad [\mathcal{G}_m, \mathcal{G}_n] \subset \mathcal{G}_{m+n}, \quad (\text{B.12})$$

Where $\mathcal{G}_m = \{x \in \hat{sl}_3(C) \mid [D, x] = m x; m \in \mathbb{Z}\}$.

The subspace \mathcal{G}_0 is a subalgebra of $\hat{sl}_3(C)$ given by

$$\hat{\mathcal{G}}_0 = C H_1 \oplus C H_2 \oplus C E_{\alpha}^{(0)} \oplus C C \oplus C D \quad (\text{B.13})$$

and for the subspaces \mathcal{G}_m ($m \neq 0$) we have

$$\mathcal{G}_m = C H_1^{(m)} \oplus C H_2^{(m)} \oplus C E_{\alpha_1}^{(m)} \oplus C E_{\alpha_2}^{(m)} \oplus C E_{\alpha_3}^{(m)} \oplus C E_{-\alpha_1}^{(m)} \oplus C E_{-\alpha_2}^{(m)} \oplus C E_{-\alpha_3}^{(m)}. \quad (\text{B.14})$$

We use in the paper the fundamental highest weight representation $|\lambda_0\rangle$, satisfying

$$H_a^{(0)} |\lambda_0\rangle = 0, \quad E_{\alpha}^{(0)} |\lambda_0\rangle = 0, \quad C |\lambda_0\rangle = |\lambda_0\rangle \quad (\text{B.15})$$

for $a, b = 1, 2$, and $\alpha \in \Delta$. Such state is annihilated by all positive grade subspaces

$$\mathcal{G}_m |\lambda_0\rangle = 0, \quad m > 0, \quad (\text{B.16})$$

and all the representation space is spanned by the states obtained by acting on $|\lambda_0\rangle$ with negative grade generators. This representation space can be supplied with a scalar product such that one has

$$(H_a^m)^{\dagger} = H_a^{-m}, \quad (E_{\alpha}^m)^{\dagger} = E_{-\alpha}^{-m}, \quad (\text{B.17})$$

$$C^{\dagger} = C, \quad D^{\dagger} = D. \quad (\text{B.18})$$

It follows from (B.13) and (B.14) that

$$(\mathcal{G}_m)^{\dagger} = \mathcal{G}_{-m}, \quad (\text{B.19})$$

and therefore

$$\langle \lambda_0 | \mathcal{G}_{-m} = 0, \quad m > 0. \quad (\text{B.20})$$

In addition to the subalgebra \mathcal{g}_0 it is also convenient to consider two additional subalgebras

$$\mathcal{G}_{<0} = \bigoplus_{m>0} \mathcal{G}_m, \quad \mathcal{G}_{>0} = \bigoplus_{m>0} \mathcal{G}_{-m}. \quad (\text{B.21})$$

These subalgebras and the corresponding Lie groups play important role in the DT method.

The next relationships are useful in the AKNS_r ($r = 2$) model construction. The special element $E^{(l)}$ in the basis presented above can be written as

$$E^{(l)} = \frac{1}{3}(H_1^{(l)} + 2H_2^{(l)}), \quad [D, E^{(l)}] = lE^{(l)}. \quad (\text{B.22})$$

Then the matrix $E^{(0)}$ becomes

$$E^{(0)} = \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \quad (\text{B.23})$$

The roots entering in the AKNS₂ construction are

$$\beta_1 \equiv \alpha_3 = \alpha_1 + \alpha_2; \quad \beta_2 = \alpha_2. \quad (\text{B.24})$$

Moreover, the following commutation relations hold

$$[E^{(l)}, H_a^{(m)}] = l \delta_{2a} C \delta_{l+m,0}, \quad a = 1, 2, \quad (\text{B.25})$$

$$[E^{(l)}, E_{\pm\beta_j}^{(m)}] = \pm E_{\pm\beta_j}^{(l+m)}, \quad j = 1, 2. \quad (\text{B.26})$$

$$[E^{(l)}, E_{\pm\beta_1 \mp \beta_2}^{(m)}] = 0, \quad (\text{B.27})$$

$$[H_1^{(m)}, E_{\pm\beta_1}^{(n)}] = \pm E_{\pm\beta_1}^{(m+n)}, \quad [H_1^{(m)}, E_{\pm\beta_2}^{(n)}] = \mp E_{\pm\beta_2}^{(m+n)} \quad (\text{B.28})$$

$$[H_2^{(m)}, E_{\pm\beta_1}^{(n)}] = \pm E_{\pm\beta_1}^{(m+n)}, \quad [H_2^{(m)}, E_{\pm\beta_2}^{(n)}] = \pm 2E_{\pm\beta_2}^{(m+n)} \quad (\text{B.29})$$

$$[H_1^{(m)}, E_{\pm\beta_1 \mp \beta_2}^{(n)}] = \pm 2E_{\pm\beta_1 \mp \beta_2}^{(m+n)} \quad (\text{B.30})$$

$$[H_2^{(m)}, E_{\pm\beta_1 \mp \beta_2}^{(n)}] = \mp E_{\pm\beta_1 \mp \beta_2}^{(m+n)} \quad (\text{B.31})$$

C $\hat{sl}(3)$ matrix elements

Consider the vertex operators associated to bright soliton solutions

$$F_j = \sum_{n=-\infty}^{+\infty} \nu_j^n E_{-\beta_j}^{(-n)}, \quad G_j = \sum_{n=-\infty}^{+\infty} \rho_j^n E_{\beta_j}^{(-n)}; \quad j = 1, 2; \quad \nu_j, \rho_j \in \mathbb{C}. \quad (\text{C.1})$$

It can be shown that they are nilpotent, i.e. $F_j^2 = 0, G_j^2 = 0$. The matrix element $\langle \lambda_o | F_j G_k | \lambda_o \rangle$ can be computed by developing the products and keeping only non-trivial terms, then one makes use of the commutation rules to get the central term C . The double sum can be simplified to a single sum, which provide the power series $\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}$. So, one has

$$\langle \lambda_o | F_j G_k | \lambda_o \rangle = \frac{\nu_j \rho_k}{(\nu_j - \rho_k)^2} \delta_{j,k} \quad (\text{C.2})$$

Let us consider the deformation of the vertex operators F_2, G_2 as

$${}_{\pm\beta_2}(k^\pm) = \sum_{n=-\infty}^{+\infty} \left(\frac{w^\pm}{k^\pm}\right)^{-n} [k^\pm E_{\beta_2}^{(n)} - \rho_1^\mp E_{\pm\beta_2 \mp \beta_1}^{(n)}], \quad w^\pm = (k^\pm)^2 - \rho_1^+ \rho_1^-. \quad (\text{C.3})$$

It is a direct computation to show the nilpotency of these operators, i.e. $\Gamma_{\pm\beta_2}^2 = 0$. Similar computations to the one in (C.2) provide the following matrix element

$$\langle \lambda_o | \Gamma_{\beta_2}(k^+) \Gamma_{-\beta_2}(k^-) | \lambda_o \rangle = \frac{w^+ w^- k^+ k^-}{(k^+ k^- + \rho_1^+ \rho_1^-)(k^+ - k^-)^2} \quad (\text{C.4})$$

Consider the vertex operator analog to the one in (A.10)

$$V_{\beta_1}^q(\lambda, \rho_0) = \sum_{n=-\infty}^{\infty} \left\{ (\lambda^2 - \rho_0^2)^{-n/2} [e_q]^n \left[\frac{1}{2} (H_1^{(n)} + H_2^{(n)}) - \frac{\rho_1^+}{\lambda - e_q (\lambda^2 - \rho_0^2)^{1/2}} E_{\beta_1}^{(n)} + \frac{\rho_1^-}{\lambda + e_q (\lambda^2 - \rho_0^2)^{1/2}} E_{-\beta_1}^{(n)} \right] + e_q \frac{(\lambda^2 - \rho_0^2)^{1/2}}{2\lambda} \delta_{n,0} C \right\}; \quad q = 1, 2 \quad (\text{C.5})$$

where $e_q \equiv (-1)^{q-1}$ and $\rho_0^2 = 4\rho_1^+ \rho_1^-$. The next matrix element computation follows similar steps to the one performed to arrive at (A.13), except that one must take into account the $\hat{sl}(3)$ commutation rules. So, one has

$$\langle \lambda_o | V_{\beta_1}^q(\lambda_1, \rho_0) V_{\beta_1}^q(\lambda_2, \rho_0) | \lambda_o \rangle = \frac{1}{4} \{ [2 + 2K(\lambda_1, \lambda_2)] \frac{K_0(\lambda_1, \lambda_2)}{[1 - K_0(\lambda_1, \lambda_2)]^2} + \left[\frac{K(\lambda_1, \lambda_2) \rho_0^2}{\lambda_1 \lambda_2} + 1 \right] \}; \quad q = 1, 2, \quad (\text{C.6})$$

where K_0 and K are given in (A.14). Since this two-point function, except for an overall constant factor, is similar to the one in (A.13) one can use the relationships (A.15)-(A.16) to show that the operator $V_{\beta_1}^q(\lambda_1, \rho_0)$ is nilpotent.

The vertex operator generating the dark-dark soliton solution becomes

$$W^q(k, \rho_{1,2}^\pm) = \sum_{n=-\infty}^{\infty} \left\{ (k^2 - 4 \sum_{i=1}^2 \rho_i^+ \rho_i^-)^{-n/2} [e_q]^n \left[s_1 H_1^{(n)} + s_2 H_2^{(n)} + \sum_{i=1}^2 e_{i(q)}^+ E_{\beta_i}^{(n)} + \sum_{i=1}^2 e_{i(q)}^- E_{-\beta_i}^{(n)} + e_{12}^+ E_{\beta_1 - \beta_2}^{(n)} + e_{12}^- E_{\beta_2 - \beta_1}^{(n)} \right] + e_q \frac{(k^2 - 4 \sum_{i=1}^2 \rho_i^+ \rho_i^-)^{1/2}}{2k} \delta_{n,0} C \right\} \quad (\text{C.7})$$

$$e_{i(q)}^\pm = \frac{\mp \rho_i^\pm}{k \mp e_q (k^2 - 4 \sum_j \rho_j^+ \rho_j^-)^{1/2}}, \quad s_2 = \frac{1}{2}; \quad s_1 = \frac{1}{2} \frac{\rho_1^+ \rho_1^-}{\sum_i \rho_i^+ \rho_i^-}; \quad e_{12}^\pm = \frac{1}{2} \frac{\rho_1^\pm \rho_2^\mp}{\sum_i \rho_i^+ \rho_i^-};$$

$$e_q \equiv (-1)^{q-1}$$

Notice that the vertex operator (C.7) reduces to the one in (C.5) in the limit $\rho_2^\pm \rightarrow 0$. The nilpotent property of this vertex operator can be verified as follows

$$\langle \lambda_o | W^1(k_1, \rho_{1,2}^\pm) W^1(k_2, \rho_{1,2}^\pm) | \lambda_o \rangle = \left(\frac{x_1 x_2}{4} \frac{x_2 - x_1}{x_1^2 + S} \quad x_1 + \frac{1}{4} \frac{S - 4x_1^2}{x_1^2 + S} (x_2 - x_1) + \dots \right) \quad (\text{C.8})$$

where $x_1 = \sqrt{k_1^2 - 4S}$, $x_2 = \sqrt{k_2^2 - 4S}$, $S = \sum_j \rho_j^+ \rho_j^-$. In the limit $x_2 \rightarrow x_1$ (or $k_2 \rightarrow k_1$) the r. h. s. of eq. (C.8) vanishes.

References

- [1] T. Kanna and M. Lakshmanan, *Phys. Rev. Lett.* **86** (2001) 5043.
- [2] T. Kanna and M. Lakshmanan, *Phys. Rev.* **E67** (2003) 046617.
- [3] T. Kanna, M. Lakshmanan, P. Tchofo Dinda and N. Akhmediev, *Phys. Rev.* **E73** (2006) 026604.
- [4] Y. Ohta, D-S Wang and J. Yang, *Studies in Applied Mathematics* **0** (2011) 1; see also [arXiv:nlin.PS/10112522].
- [5] S.V. Manakov, *Sov. Phys. JETP Lett.* **38** (1974) 248.
- [6] H. Blas, *JHEP* **311** (2003) 054.
- [7] H. Blas, *JHEP* **03** (2007) 055;
H. Blas and H.L. Carrion, *JHEP* **0701** (2007) 027.
- [8] A.V. Mikhailov, E.A. Kuznetsov, A.C. Newell and V.E. Zakharov (Eds.), Proceedings of the Conference on *The Nonlinear Schrödinger Equation*, Chernogolovka, 25 July-3 August 1994. Published in *Physica* **D87** (1995) 1.
- [9] T. Kanna, M. Vijayajayanthi and M. Lakshmanan, *Phys. Rev.* **A76** (2007) 013808.
- [10] R. Radhakrishnan and M. Lakshmanan, *J. Phys.* **A28** (1995) 2683.
- [11] A. Degasperis and S. Lombardo, *J. Phys.* **A40** (2007) 961.
- [12] B. Prinari, M. J. Ablowitz and G. Biondini, *J. Math. Phys.* **47** (2006) 063508.
- [13] B. Prinari, G. Biondini, and A. D. Trubatch, *Studies in Applied Mathematics* **126** (2011) 245.
- [14] C. Kalla, *J. Phys. A: Math. Theor.* **44** (2011) 335210.
- [15] A de O Assunção, H Blas and M J B F da Silva, *J. Phys.* **A45** (2012) 085205.
- [16] L.A. Ferreira, J. L. Miramontes and J. Sánchez Guillen, *J. Math. Phys.* **38** (1997) 882 (hep-th/9606066).
- [17] V.S. Gerdjikov, D.J. Kaup, I.M. Uzunov and E.G. Evstatiev, *Phys. Rev.* **E77** (1996) 3943.
- [18] Q-Han Park and H. J. Shin, *Phys. Rev.* **E61** (2000) 3093.
- [19] A.G. Bueno, L.A. Ferreira, A.V. Razumov, *Nucl. Phys.* **B626** (2002) 463.
- [20] A.P. Fordy and P.P. Kulish, *Commun. Math. Phys.* **89** (1983) 427;
A.P. Fordy, in *Soliton Theory: a Survey of Results*, (ed. A.P. Fordy) University Press, Manchester (1990), p. 315.

- [21] H. Aratyn, J.F. Gomes and A.H. Zimerman, *J. Math. Phys.* **36** (1995) 3419.
- [22] H.S. Blas Achic, L.A. Ferreira, J.F. Gomes and A.H. Zimerman, *Phys. Lett.* **237A** (1998) 225.
- [23] A. P. Sheppard and Y. S. Kivshar, *Phys. Rev.* **E55** (1997) 4773.
- [24] D.S. Wang, D. Zhang and J. Yang, *J. Math. Phys.* **51** (2010) 023510.
- [25] M. Vijayajayanthi, T. Kanna, and M. Lakshmanan, *Phys. Rev.* **A77** (2008) 013820.
- [26] C. Becker, S. Stellmer, P. Soltan-Panahi, S. Dorschner, M. Baumert, E. M. Richter, J. Kronjager, K. Bongs, and K. Sengstock, *Nature Physics* **4** (2008) 496.
- [27] Th. Busch and J.R. Anglin, *Phys. Rev. Lett.* **87** (2001) 010401.
- [28] N. Akhmediev, W. Królikowski and A.W. Snyder, *Phys. Rev. Lett.* **81** (1998) 4632.
- [29] A. Ankiewicz, W. Królikowski and N.N. Akhmediev, *Phys. Rev.* **E59** (1999) 6079.
- [30] O. Babelon, D Bernard and M. Talon, *Introduction to Classical Integrable Systems*, Cambridge Univ. Press, Cambridge, 2003.
- [31] H. Blas, Vector NLS hierarchy solitons revisited: dressing transformation and tau function approach, [arxiv:solv-int/9912015].
- [32] Y. S. Kivshar and B. Luter-Davies, *Phys. Reports* **298** (1998) 81
- [33] K. Nakkeeran, *Phys. Rev.* **E64** (2001) 046611.
- [34] V. S. Gerdjikov, Selected Aspects of Soliton Theory Constant boundary conditions, [arXiv:nlin.SI/0604005]
- [35] L.D. Faddeev and L.A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons*, Springer-Verlag, London, (1987).
- [36] V.V. Konotop and V. E. Vekslerchik, *Phys. Rev.* **E49** (1994) 2397.
- [37] E. V. Doktorov, *J. Math. Phys.* **38** (1997) 4138.
- [38] V.E. Zakharov, A.B. Shabat, *Zh. Eksp. Teor. Fiz.* **64** (1973) 1627 [*Sov. Phys. JETP* **37** (1973) 823].
- [39] L. F. dos Santos, H. Blas and M. J. B. da Silva, *PoSISFTG* (2009) 043 [arXiv:0906.0131 [nlin.SI]].
- [40] J. F. Gomes, G. S. Franca, G R de Melo and A H Zimerman, *J. Phys. A: Math. Theor.* **42** (2009) 445204.
- [41] Shan-liang Liu and Wen-zheng Wang, *Phys. Rev.* **E48** (1993) 3054.
- [42] P. Goddard and D. Olive, *Int. J. Mod. Phys.* **A1** (1986) 303.

AKNS solitons with constant boundary conditions via dressing transformation and tau function approach

H. Blas and M. Zambrano

Instituto de Física
Universidade Federal de Matogrosso - UFMT
Av. Fernando Corrêa, N°2367 - Bairro Boa Esperança, Cuiabá.
MT-Brasil

Abstract

We consider constant boundary conditions (cbc) for the Ablowitz-Kaup-Newell-Segur(AKNS) system in the framework of the $sl(2)$ affine Kac-Moody algebra formulation. Through the dressing transformation and the tau function approach we obtain its soliton solutions and classify them according to their regular or singular behaviors. Moreover, by the analytic continuation method it is obtained the dark soliton solution of the well-known nonlinear Schrodinger equation (NLS) with cbc, which turns out to be a submodel of the AKNS system.

1 Introduction

An interesting area of research in the context of integrable systems is the study of algebraic methods to solve the nonlinear partial differential equations which arise in those systems. Among the various methods available one has the dressing transformation method (DT) which turned out to be a convenient tool to tackle the problem with nonvanishing boundary conditions, as recently presented in [1]. The dressing transformation method works by mapping certain trivial solutions into new solutions, in particular the soliton type solutions.

In this work one follows the DT method as presented in [1, 2, 3] to find certain solutions of the Ablowitz-Kaup-Newell-Segur (AKNS) system associated to constant boundary conditions (cbc). The model is defined by [1, 3]

$$q_t - \frac{1}{2}q_{xx} + (qr - q_0r_0)q = 0, \quad r_t + \frac{1}{2}r_{xx} - (qr - q_0r_0)r = 0, \quad (1)$$

where $q = q(x, t)$, $r = r(x, t)$ are real functions and $q_0, r_0 \in \mathbb{R}$ are some parameters. When $q_0 = r_0 = 0$ one has the usual AKNS system conveniently written to consider vanishing boundary condition (vbc), it has been studied in ref. [4] in the context of the DT method and the soliton type solutions have been found. For an earlier reference on this model see [5] in which the theory of highest weight representations of the affine Kac-Moody algebra $sl(2)$ is applied to find soliton type solutions with vanishing boundary condition.

The AKNS model (1), through convenient change of variables which involves analytic continuation of the parameters (see below), can be reduced to the model

$$i\varphi_t - \frac{1}{2}\varphi_{xx} \pm (|\varphi|^2 - A^2)\varphi = 0, \quad A \in \mathbb{R}, \quad (2)$$

which is precisely the nonlinear Schrodinger model (NLS) conveniently written to study cbc. In this model one has $\varphi = \varphi(x, t) \in \mathbb{C}$. The signs \pm define the defocusing and focusing NLS model for (+) and (−) signs, respectively¹. The dark soliton solution of the first model will be obtained in this paper. The focusing model has recently been considered in [6].

An extension of the above model for 4 AKNS fields has been considered in [1, 7] in which the cbc, vbc and the mixed cbc-vbc have been discussed.

2 AKNS system with cbc

Let us consider the Lie algebra valued connections $U, V \in \widehat{sl}_2(\mathbb{C})$

$$\begin{aligned} U &= H^{(1)} + qE_+^{(0)} + rE_-^{(0)} + \eta_V \widehat{c} \\ V &= H^{(2)} + qE_+^{(1)} + rE_-^{(1)} + \eta_V \widehat{c} - \frac{1}{2}(qr - q_0r_0)H^{(0)} + \frac{1}{2}q_x E_+^{(0)} - \frac{1}{2}r_x E_-^{(0)} \end{aligned} \quad (3)$$

¹Some care must be taken in the literature about the focusing/defocusing convention. The present paper convention differs from the one in [1], there the sign of the $\partial_x \varphi$ term has been fixed to be positive, so the \pm signs above become multiplied by a minus sign, therefore one recovers the standard convention (+) for focusing and (−) for defocusing.

where the fields q, r define the AKNS model and η_U, η_V are auxiliary fields. These fields will play an important role in our approach since they lie in the direction of the central term of the affine $sl(2)$ Kac-Moody algebra, the central term is essential in order to define an integrable highest weight representation of this algebra. See the appendix for the $sl(2)$ Kac-Moody notations and conventions. The connections (3) allow us to construct the system of eqs. (1) through the zero-curvature representation (27). Next, let us discuss the vacuum solutions of (1) and the related boundary conditions. First, the vbc: $q(x), r(x) \rightarrow 0$ as $|x| \rightarrow \pm\infty$ is compatible with the trivial solution $q(x) = r(x) \equiv 0$ of the system. Second, the case with cbc defined as: $\{q(x), r(x)\} \rightarrow \{\hat{q}_0, \hat{r}_0\}$ as $|x| \rightarrow \pm\infty$ is compatible with the trivial solution $q(x) \equiv \hat{q}_0, r(x) \equiv \hat{r}_0$ provided that $\hat{q}_0\hat{r}_0 - q_0r_0 \equiv 0$. In what follows, for the sake of simplicity, let us assume $\hat{q}_0 = q_0, \hat{r}_0 = r_0$. Let us find the solution of the system (26) for Ψ related to the cbc discussed above. One has

$$U_0 = H^{(1)} + q_0 E_+^{(0)} + r_0 E_-^{(0)}, \quad V_0 = H^{(2)} + q_0 E_+^{(1)} + r_0 E_-^{(1)}, \quad (4)$$

therefore

$$\Psi_0 = \exp(xU_0 + tV_0) \quad (5)$$

2.1 Vertex operators

It will be useful to consider adjoint eigenstates of the the combination $xU_0 + tV_0$. The first eigenstate is

$$X = \sum_{-\infty}^{+\infty} \frac{1}{\theta^n} \left[H^{(n)} - \frac{q_0}{z-\theta} E_+^{(n)} + \frac{r_0}{z+\theta} E_-^{(n)} \right] + \frac{\theta}{z} \hat{c}, \quad (6)$$

where $\theta = \sqrt{z^2 - q_0 r_0}$. Then one has

$$[xU_0 + tV_0, X] = \xi X, \quad \xi = 2z(x + \theta t). \quad (7)$$

The group element

$$g = \exp(aX) \implies \Psi_0 g \Psi_0^{-1} = 1 + a \exp(\xi) X \quad (8)$$

defines a vertex operator which will be associated to soliton solutions. Notice that the eigenstate X is a nilpotent operator, i.e. $X^2 = 0$.

Next, consider another eigenstate, as it is showed in ([7]),

$$Y = \sum_{-\infty}^{+\infty} \frac{(-1)^n}{\theta^n} \left[H^{(n)} - \frac{q_0}{z+\theta} E_+^{(n)} + \frac{r_0}{z-\theta} E_-^{(n)} \right] - \frac{\theta}{z} \hat{c}. \quad (9)$$

In fact, one can see that

$$[xU_0 + tV_0, Y] = \eta Y, \quad \eta = 2z(x - \theta t). \quad (10)$$

In addition, it can be shown that it is also nilpotent $Y^2 = 0$.

3 Dressing transformation and tau functions

The gauge transformation (29), using (3)-(4), provides

$$D^{(-1)} = a_1 H^{(-1)} - \frac{1}{2}(q - q_0)E_+^{(-1)} + \frac{1}{2}(r - r_0)E_-^{(-1)}, \quad a_1 \in \mathbb{R} \quad (11)$$

where $D^{(-1)}$ is the expression in (30) for $j = -1$. Next, from (28) for an arbitrary degree $n \in \mathbb{Z}$ one has

$$\langle \mu_0 | \Psi_0 g \Psi_0^{-1} \rangle_{(n)} | \mu_0 \rangle = \langle \mu_0 | \Theta_0^{-1} \Theta_0^{-1} \Theta_+ \rangle_{(n)} | \mu_0 \rangle \quad (12)$$

where $|\mu_0\rangle$ is a highest weight vector in a fundamental representation of $\widehat{sl}_2(\mathbb{C})$.

The tau functions τ_a , ($a = 0, \pm$), by making use of (12), can be defined as

$$\tau_{\pm} = \langle \mu_0 | E_{\mp}^{(1)} \Psi_0 g \Psi_0^{-1} \rangle_{(-1)} | \mu_0 \rangle, \quad (13)$$

$$\tau_0 = \langle \mu_0 | \Psi_0 g \Psi_0^{-1} \rangle_{(0)} | \mu_0 \rangle \quad (14)$$

and using (11) one has that the fields of the model become parametrized as

$$q = q_0 + 2\frac{\tau_+}{\tau_0}, \quad r = r_0 - 2\frac{\tau_-}{\tau_0}. \quad (15)$$

In fact, these eqs. provide a solution of the system (1) provided that the tau functions τ_a are constructed by using the relationships (13)-(14). In the following we present the two types of solutions associated to the eigenstates X, Y , respectively.

3.1 X AKNS Solitons

From (15) and taking into account (8), (13)-(14) for $c = a\frac{\theta}{z}$, $f_c = \frac{c \exp(\xi)}{1 + c \exp(\xi)}$ one has

$$q = q_0 \left(1 - \frac{2zf_c}{z - \theta} \right), \quad r = r_0 \left(1 - \frac{2zf_c}{z + \theta} \right). \quad (16)$$

Notice that for $c > 0$ one has a regular solution, whereas for $c < 0$ the solution is singular. In particular for the numeric values $a = 1, z = 2, q_0 = 1, r_0 = 1$ one has plotted the Fig.1 representing a regular 1-soliton,

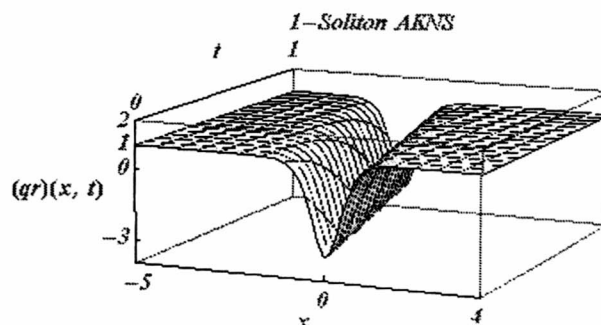


Fig. 1: 1-Soliton of AKNS for cbc: $q_0 = 1, r_0 = 1$ and $a = 1, z = 2$.

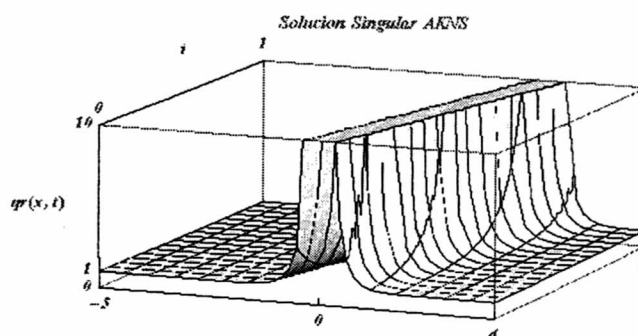


Fig. 2: Singular solution in AKNS for cbc: $q_0 = 1, r_0 = 1$.

3.2 Y AKNS solitons

From (15) and taking $g = \exp(aY)$ in the relationships (13)-(14) one gets

$$q = q_0 \left(1 + \frac{2zg_c}{z + \theta} \right), \quad r = r_0 \left(1 + \frac{2zg_c}{z - \theta} \right), \quad (17)$$

where $c = a \frac{\theta}{z}$ and $g_c = \frac{c \exp(\eta)}{1 - c \exp(\eta)}$. Similarly to the case above, one notices that for $c < 0$ one has a regular solution, whereas for $c > 0$ the solution is singular. For $a = 1, z = 2$ we have plotted the Fig. 2 which is a singular solution.

3.3 The NLS submodel and dark solitons

Let us make the following analytic continuation on the AKNS (1) system

$$(x, t) \rightarrow (ix, it), \quad (q_0, r_0) \rightarrow (iA, iA) \quad \text{and} \quad (q, r) \rightarrow (iq, ir). \quad (18)$$

So, one gets a defocalized NLS model in (2) provided the identification $\varphi = q = r^*$ is made. Moreover, in the solutions obtained for q, r in (16) let us make $a \rightarrow ia, z \rightarrow iz$ in order to get

$$\xi = -2z(x + \theta t), \quad \theta = \sqrt{A^2 - z^2} \quad (19)$$

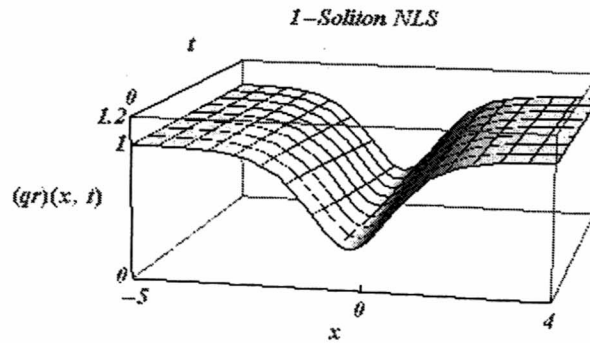


Fig 3: 1-soliton for defocusing-NLS model with cbc: $A = 1$.

and

$$\varphi = A \left(-1 + \frac{2izf_c}{iz - \theta} \right). \quad (20)$$

Then, defining $\psi^\pm = \frac{1}{2}(\xi + \ln(\pm c))$, one has

$$|\varphi|^2 = A^2 - z^2 \times \begin{cases} +\text{sech}^2(\psi^+), & c > 0, \\ -\text{csch}^2(\psi^-), & c < 0, \end{cases} \quad (21)$$

For $a = 1, z = \frac{5}{6}$ ($c > 0$) we have plotted the Fig. 3.

Now, one can use (17) to obtain NLS solutions. Thus using the same conditions and change of variables mentioned above, one has

$$\eta = -2z(x - \theta t), \quad \theta = \sqrt{A^2 - z^2} \quad (22)$$

and

$$\varphi = -A \left(1 + \frac{2izg_c}{iz + \theta} \right) \quad (23)$$

Then, defining $\psi^\pm = \frac{1}{2}(\eta + \ln(\pm c))$, one has

$$|\varphi|^2 = A^2 + z^2 \times \begin{cases} +\text{csch}^2(\psi^+), & c > 0, \\ -\text{sech}^2(\psi^-), & c < 0, \end{cases} \quad (24)$$

For $a = 1, z = \frac{5}{6}$ ($c > 0$) we have plotted the Fig. 4 which shows the singular solution for NLS. Finally, one notices that the both type of solutions for the AKNS, the singular and regular solitons, under the complexification process emerge as solutions of the NLS model.

3.4 2-solitons

In order to construct the 2-dark-soliton of the defocusing-NLS model one uses the group element $g = \exp(aX_a) \exp(bX_b)$, where X_a, X_b are two copies of the eigenstate defined in (6). Following similar steps as in the previous constructions one can get the 2-soliton. We simply provide its profile in Fig. 4.

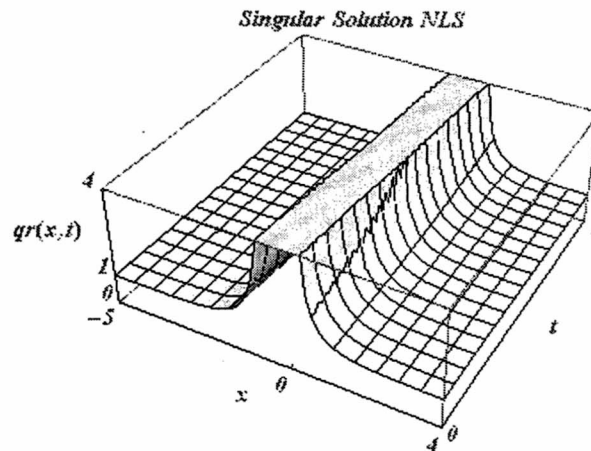


Fig 4: Singular Solution for defocusing-NLS model with cbc: $A = 1$.

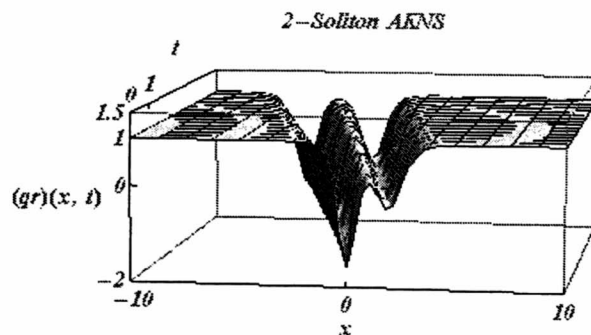


Fig 5: 2-soliton for $a = 3, b = -2$ and cbc: $q_0 = r_0 = 1$.

4 Discussion

We have obtained regular and singular soliton solutions of the AKNS model related to $sl(2)$ affine Kac-Moody algebra in the framework of the dressing method and its associated tau function parametrization of the fields. Through an adequate analytic continuation we have obtained a defocusing-NLS as a sub-model and constructed its dark soliton solutions.

Acknowledgements

MZ thanks FAPEMAT and CAPES for financial support and HB is grateful to FAPEMAT for partial support.

Dedicated to the memory of our mathematics teacher Eugen Blum.

A Algebraic formulation and dressing of the AKNS model

In this appendix we summarize some useful constructions, as well as the notations and conventions, regarding the $sl(2)$ affine Kac-Moody algebra formulation and the dressing transformation for the AKNS model. The $\widehat{sl}_2(\mathbb{C})$ algebra commutation relations for the elements $H^{(m)}$, $E_+^{(n)}$ and $E_-^{(n)}$ become

$$\left[H^{(m)}, H^{(n)}\right] = 2m\delta_{m+n,0}\widehat{c}, \quad \left[H^{(m)}, E_{\pm}^{(n)}\right] = \pm 2E_{\pm}^{(m+n)}, \quad \left[E_+^{(m)}, E_-^{(n)}\right] = H^{(m+n)} + m\delta_{m+n,0}\widehat{c}. \quad (25)$$

Let \mathcal{G} be a semisimple Lie algebra. The *loop algebra* $\widetilde{\mathcal{G}}$, with infinite dimension, is defined by

$$\widetilde{\mathcal{G}} \equiv \mathcal{G} \otimes \mathbb{C}[z, z^{-1}],$$

where $\mathbb{C}[z, z^{-1}]$ is the set of Laurent polynomials in the variable $z \in \mathbb{C}$. An affine Kac-Moody Lie algebra $\widehat{\mathcal{G}}$ is defined by

$$\widehat{\mathcal{G}} \equiv \widetilde{\mathcal{G}} \oplus \mathbb{C}\widehat{c} \oplus \mathbb{C}\widehat{d},$$

where \widehat{c} is the so-called *central extension* which commutes with every element of $\widetilde{\mathcal{G}}$. The operator \widehat{d} is a derivation operator and defines a homogeneous gradation of the algebra $\widetilde{\mathcal{G}}$ through the relationship $[\widehat{d}, \widehat{\mathcal{G}}^{(j)}] = j \widehat{\mathcal{G}}^{(j)}$.

The algebra $\widehat{\mathcal{G}}$ can be expanded as a direct sum of graded subspaces \mathbb{Z}

$$\widehat{\mathcal{G}} = \bigoplus_{j \in \mathbb{Z}} \widehat{\mathcal{G}}^{(j)}, \quad j - \text{grade of the subspace.}$$

Let us define a linear system of eqs

$$\partial_x \Psi = U \Psi, \quad \partial_t \Psi = V \Psi \quad (26)$$

where the Ψ is a group element and U, V are some Lie algebra valued connections. Many well known integrable systems admit the linear system of equations (26), whose compatibility condition implies the zero-curvature equation [8]

$$[\partial_x - U, \partial_t - V] = 0, \quad (27)$$

which is equivalent to the system of eqs. defining the integrable system.

A.1 Dressing transformation

The dressing method is related to the factorization problem

$$\Psi_0 g \Psi_0^{-1} = \Theta_-^{-1} \Theta_0^{-1} \Theta_+, \quad (28)$$

where Ψ_0 is a trivial solution of (27) and g is a constant group element. The result (28) guaranties that the *dressing operators*

$$\Theta_- := \exp\left(\sum D^{(-n)}\right), \quad \Theta_0 := \exp\left(D^{(0)}\right), \quad \Theta_+ := \exp\left(\sum D^{(n)}\right)$$

map a trivial solution (vacuum solution) of (26)-(27) into a new nontrivial solution through the gauge transformation

$$U = \Theta_{\pm} U_0 \Theta_{\pm}^{-1} + \partial_x \Theta_{\pm} \Theta_{\pm}^{-1}, \quad V = \Theta_{\pm} V_0 \Theta_{\pm}^{-1} + \partial_t \Theta_{\pm} \Theta_{\pm}^{-1}. \quad (29)$$

The operators $D^{(j)} \in \widehat{\mathcal{G}}^{(j)}$ for $j \in \mathbb{Z}$ become

$$D^{(j)} := a_j H^{(j)} + b_j E_+^{(j)} + c_j E_-^{(j)} \quad (30)$$

References

- [1] A de O Assunção, H. Blas and M J B F da Silva, New Derivation of Soliton Solutions to the AKNS₂ system via Dressing Transformation Methods, *J. Phys. ATM.* **45** (2011) 085205.
- [2] L.A. Ferreira, J. L. Miramontes and J. Sanchez Guillen, *J. Math. Phys.* **38** (1997) 882.
- [3] J. F. Gomes, G.S. Franca and A.H. Zimerman, Dressing Approach to the Nonvanishing Boundary value problem for the AKNS hierarchy, *J.Phys. Conf. Ser* **343** (2012) 012039
- [4] H. Blas, Alguns Aspectos Algébricos de Sistemas Generalizados da Equação de Schrödinger Não Linear (2 dimensões), Dissertação do Mestrado, 1996, ITF-UNESP. Veja também em “Vector NLS hierarchy solitons revisited: dressing transformation and tau function approach”, arXiv:solv-int/9912015.
- [5] V. Kac, Infinite Dimensional Lie Algebras, *Cambridge University Press.* (2003).
- [6] V. E. Zakharov and A.A. Gelash, Soliton on Unstable Condensate, *arXiv: 1109.0620v2*(2011).
- [7] M. Zambrano, Sólitos dos Modelos AKNS e ACNLS com Condições de Contorno Nulas e Constantes Através das Transformações Dressing e as Funções Tau, Dissertação de Mestrado, 2012, Insituto de Física, UFMT-Brasil.
- [8] L. Faddeev and L. Takhtajan, Hamiltonian Methods in the Theory of Solitons, *Springer Press.* (1987).

Formación de Comunidades Latinoamericanas de Investigación La experiencia de CLARA y las posibilidades de la RAAP

Encuentro Nacional de Física Holger Valqui Casas

UNI, AGOSTO 2011

Benjamín Marticorena

Contenido

1. Avances en la formación de comunidades LA de investigación
2. Perspectivas próximas
3. Asuntos claves

I. Avances en la formación de comunidades LA de investigación

1.1 Visita a ONCYT de los países miembros

Objetivo:

- Procurar un mayor consenso entre los actores de la investigación, la administración y la política, para la expansión de la infraestructura y los servicios de redes avanzadas para CTI.

Metodología:

- Visita a autoridades públicas de CTI
- Encuentros con las RNIE
- Talleres con grupos nacionales de investigación
- Visitas realizadas (doce países)

Conclusiones básicas:

- Los proyectos más adecuados para utilizar redes avanzadas son aquellos cuyo interés atraviesa las fronteras de los países y son de interés público. Esta última condición se relaciona con el hecho de que los proyectos que utilizan redes avanzadas deben contar, en muchos casos, con el apoyo financiero del Estado.

El mayor reto para las redes avanzadas es entender bien lo que los científicos, las comunidades de investigación y, en general, los actores académicos requieren para integrarse, comunicarse y colaborar más y mejor entre sí. La tarea más importante de la RAAP (Red Académica Avanzada Peruana) para el desarrollo de la e-Ciencia es explicar a los investigadores las innovaciones que introduce en sus trabajos e identificar

los campos en que la red avanzada se constituye en la más adecuada solución técnica para sus requerimientos, así como las ventajas del trabajo colaborativo entre países.

- Importantes funciones de la Red Avanzada son la creación de una base de datos sobre fuentes financieras y posibles socios de investigación en el Perú y en otros países y constituirse en foro temático virtual para atraer investigadores al uso de RA, por su mejor desempeño y el logro de más importantes resultados de investigación. La RAAP debe apoyar a sus instituciones socias a vincularse entre sí y con sus pares de otros países
- Una opinión difundida en AL es que deben usarse redes avanzadas para proyectos de clima, ambiente, ciencias sociales, genómica, diversidad biológica, altas energías, imágenes médicas, cartografía, datos espaciales, gobierno electrónico, bioinformática y, general, todos los que demandan una alta capacidad de almacenamiento, procesamiento y transmisión de información.
- Las Teleconferencias de Información sobre grandes convocatorias internacionales de cooperación científica, son formas efectivas de comprometer a más investigadores del Perú con los de otros países de AL y de otras regiones del mundo.
- Siendo Telesalud y Telemedicina, para servicios públicos de salud a distancia, entrenamiento médico profesional e investigación en salud, campos de gran importancia para RA, ellos deben constituir una preocupación central de la RAAP.
- El campo de la Cultura es uno de los más promisorios para las RA y tiene un gran potencial de innovación de estos sistemas.
- Repositorios con información de buena calidad certificada, pueden ser organizados en la RAAP para las comunidades nacionales de investigación y educación, para satisfacer sus principales necesidades de información.
- Un observatorio de TIC puede ser un mecanismo importante a ser montado por la RAAP en el Perú para informar a sus instituciones socias sobre los avances y pronósticos de avances de esas tecnologías y sus costos, a fin de que ellas tomen decisiones pertinentes sobre compras de equipos y software.
- Para permitir la expansión y autosostenimiento de la RAAP a mediano plazo es urgente incorporar mayor número de miembros y gestionar la red de manera que los investigadores estimen que es una herramienta eficiente y de fácil uso.
- Organizar talleres temáticos (desastres naturales, biodiversidad, telemedicina, etc.) para provocar lluvias de ideas y detectar las principales áreas de cooperación en investigación, con impacto regional, nacional e internacional

1.2 Identificación de servicios requeridos por comunidades

Un estudio fue realizado por CLARA (ALICE2 DO2.5-03) en el 2010, dedicado a identificar las necesidades generales y específicas de los grupos de investigación promovidos por CLARA para desarrollar sus trabajos y sus comunicaciones en los mejores términos y para consolidarse. Los servicios más requeridos son:

- 1) Servicios H.323
- 2) Browser para Redes Avanzadas (RAUDO)
- 3) Base de Datos de proyectos, fuentes financieras, etc.
- 4) Portal sobre fuentes financieras y servicios / sitio web para las comunidades
- 5) Wikis
- 6) Calendarios comunes
- 7) Gestión de documentos (DSPACE, Google docs)
- 8) Mapas de intereses (inteligo)
- 9) Indicadores de e-Ciencia
- 10) Observatorios
- 11) Servidor de flujo de video (HD)

1.3 Incentivos para crear y consolidar comunidades

- Pago de 12 meses a un coordinador de las actividades de la Comunidad para que realice su trabajo anual bajo la dirección del líder de ésta.
- Financiamiento para la participación del líder y del coordinador de la Comunidad en algún encuentro internacional de importancia en su especialidad en el que se utilicen RA.
- Financiamiento parcial para realizar un encuentro de los miembros de la comunidad con el objeto de dar forma final a proyectos que estén pensando presentar a alguna fuente financiera.
- Una vez que la comunidad ha definido sus aplicaciones necesarias, podrá desarrollarlas con apoyo técnico de CLARA-RAAP.
- Entrenamiento para técnicos e investigadores en e-Infraestructuras y sus usos.
- Entrenamiento para los investigadores de las comunidades para preparar proyectos exitosos a ser presentados a fuentes financieras tales como FP7, ALFAIII y CYTED.
- Participación remota en Conferencias Virtuales de Información para programas FP7.

- Información sobre fuentes de financiamiento para áreas de investigación seleccionadas.
- Espacio en el sitio web de la RAAP para su comunidad, con la información pertinente sobre sus trabajos.

Videoconferencias con el apoyo de la RAAP y CLARA

Apoyo para encuentros científicos orientados a la constitución de nuevas comunidades que utilicen redes avanzadas.

- Financiación de la participación de, a lo sumo, dos expositores en encuentros científicos (congresos, talleres, seminarios, etc.) de grupos de investigación que tratan, de constituirse como usuarios de redes avanzadas.
- La RAAP puede organizar, a solicitud de los investigadores interesados, conferencias de información virtual y talleres. Estas actividades pueden ser grabadas para usos futuros.

1.4 Trabajo con agencias internacionales

Son indispensables para establecer nuevas relaciones para impulsar la e-Ciencia entre los organismos socios de la RAAP, para fortalecer a sus institutos de investigación. Las agencias con las que pueden establecerse relaciones fructíferas son BID, OEA, NSF, CYTED, RICYT, OEI y UE.

<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/work-programme.pdf>

Lista de los 14 proyectos presentados en julio del 2010 a la conferencia virtual de información del FP7 sobre e=infraestructura:

- Repositorios de Datos Federados para el Ambiente, FEDRE NET
Rodrigo Torrens; Universidad de Los Andes, en Merida, Venezuela
- Opportunities for Federating Global Earth Science Digital Repositories: the experience of GENESI – DR. L Fusco, R. Gargana from Canada
- Customized Opportunistic Grid. Harold Castro, University of Los Andes, in Colombia
- Infrastructure for Microorganisms, Agriculture and Food Research. Doris Zúñiga, National Agrarian University La Molina, in Peru
- Tele Diagnosis of Tuberculosis and Multidrug Resistance Determination using internet and mobile telephony. Mirko Zimic, Peruvian University Cayetano Heredia, in Peru.
- Seismic Tomography for clusters and volcanoes. Sebastián Araujo, Salesian Politechnic University, in Ecuador,

1.5 Reuniones de información virtual con Unión Europea

- Learning and transferring best practices and tools from successful virtual research communities (Lea PT-VRV). Ysabel Briceño, Universidad of Los Andes, in Merida, Venezuela
- Model of the Ecuadorian Atmosphere. Enrique Palacios and Sheila Serrano, Salesian Politechnic University, in Ecuador.
- Model for evaluation of Climate Change in Tropical Ecosystems. Norma Sánchez Santillán and Víctor Hugo López. Autonomous Metropolitan University, in Mexico
- Latin American Virtual Platform for Nanotechnology Modeling. Allan Campos Gallo. High Technologies National Center, CENAT, in Costa Rica
- A Toolkit for VRE (TK-VRE). Carlos Jaime Barrios Hernández. Venezuela
- 2-Infrastructure for Chemical Weather SAEMC. Eugenio Almeida, Space Research National Institut, INPE, in Brasil
- Vegetal Biology Academic Network. Aida Osuna Fernández from the Metropolitan Autonomous University, UAM, in México
- Quality Water Laboratory. Aida Malpica from the Metropolitan Autonomous University, UAM, in México

Los temas tratados en las tres conferencias virtuales de informacion en julio del 2010 con la UE (por acuerdo con la International Scientific Cooperation EU Office (Dr. Cornelia Nauen) fueron:

- *19 July 2010 (9 points connected and about 350 by streaming)*
 - 1) *Food, Agriculture, Fisheries and Biotechnologies* (Presented by Matthiessen-Guyader / Dietlind Jering / Elisabetta Balzi), and
 - 2) *Social Sciences and Humanities* (Presented by María Pilar González Pantaleón).
- *22 July 2010 (27 AV connections & 347 by streaming)*
 - 1) *Environment* (including Climate Change, presented by Birgit de Boissezon / Nick Christoforides),
 - 2) *Transport* (including aeronautics, presented by Arnoldas Milukas / Lionel Banège / Karsten Krause / Pablo Pérez Illana), and
 - 3) *Nanosciences, Nanotechnologies, Materials and New Production Technologies* (NMP, presented by María Pilar Aguar Fernández).
- *27 July 2010 (23 points connected & 345 by streaming)*
 - 1) *Health* (Presented by Stephane Hogan / Gianpietro Van De Goor), and
 - 2) *People Programme* (Presented by George Bingen / Carmen Madrid / Vanessa Debiais-Sainton)

1.6 Base de datos, proyectos, instituciones, líderes e investigadores

- Se ha construido una Base de Datos para CLARA. Hasta el presente hay 140 proyectos LA que corren sobre redes avanzadas y se da cuenta de las instituciones e investigadores involucrados en ellos.
- Una segunda Base de Datos con alrededor de 1000 investigadores de AL.

II. Perspectivas Próximas

- Consolidar comunidades LA de investigación en los campos de mas relevante interés para la región.
- Fortalecer las relaciones entre las RNIE (RAAP en el Peru) y los ONCyT en cada país de AL.
- Fortalecer las relaciones de CLARA con las principales agencias internacionales de financiamiento para promover las comunidades LA de investigación y su asociación con comunidades de otras regiones del mundo.
- Proveer un paquete de servicios a las comunidades LA de investigación que utilizan RA.

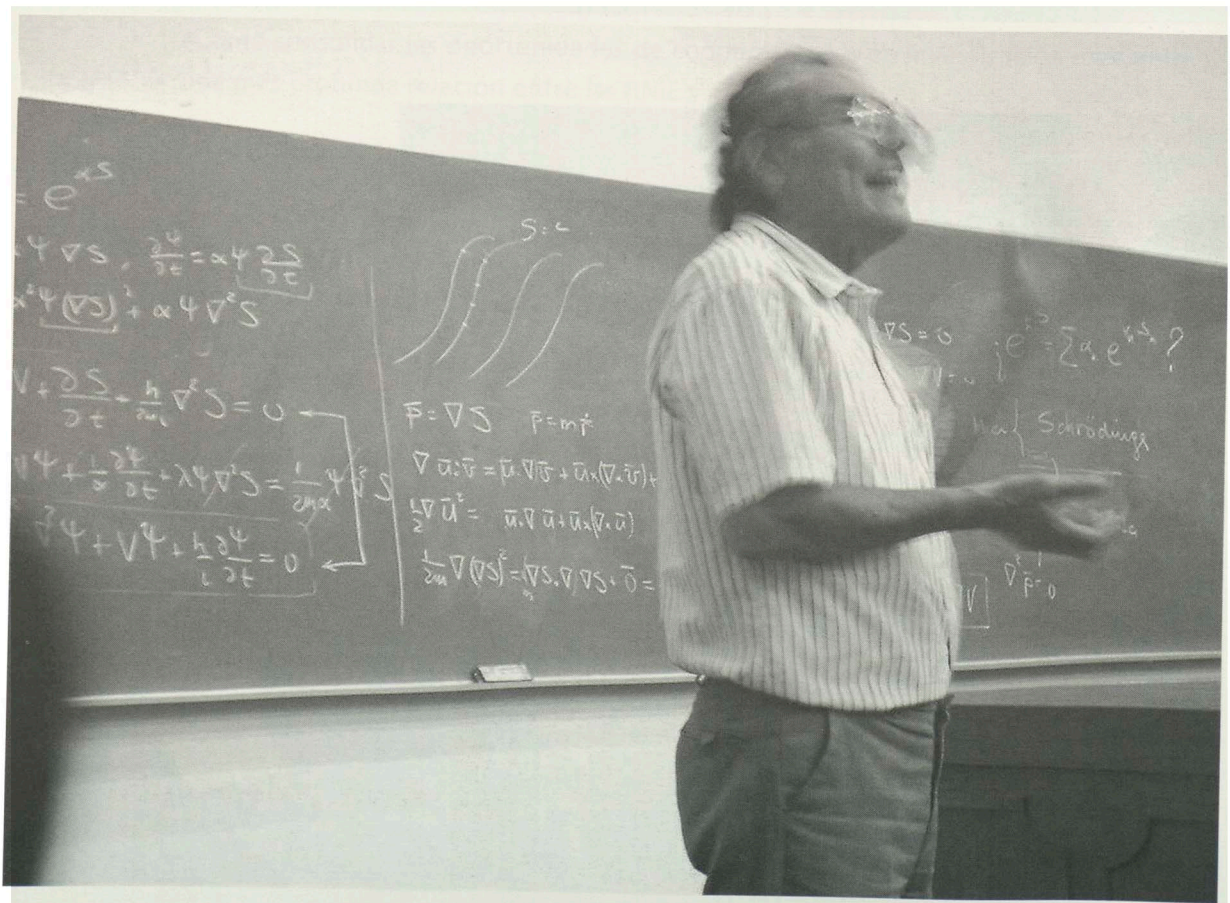
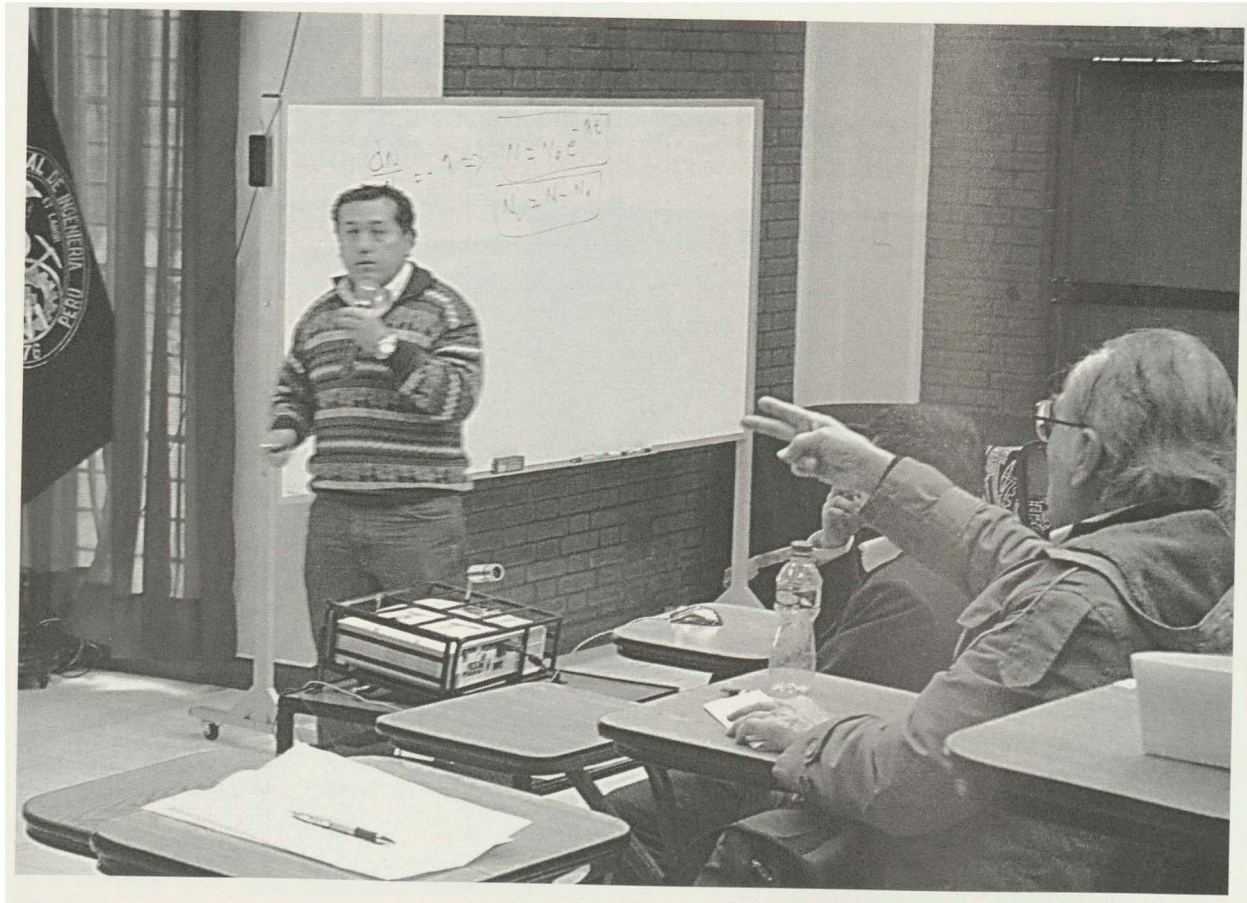
III. Asuntos Claves

- ¿Cómo consolidar las oportunidades de cooperación LA en investigación mediante una más profunda relación entre las RNIE y CLARA?
- ¿Cuáles son los compromisos específicos que las RNIE y CLARA deben asumir para lograr ese objetivo?
- ¿Debe CLARA preparar una propuesta de cooperación (ALICE3) con la UE, orientada principalmente a financiar investigación cooperativa y la infraestructura para los países menos beneficiados en las versiones 1 y 2 de ALICE?

Nota: Esta presentación está basada en información elaborada por el autor como Gerente de Relaciones Académicas de CLARA en el año 2010

Pictures of professor Valqui, his colleagues and students







Presentación

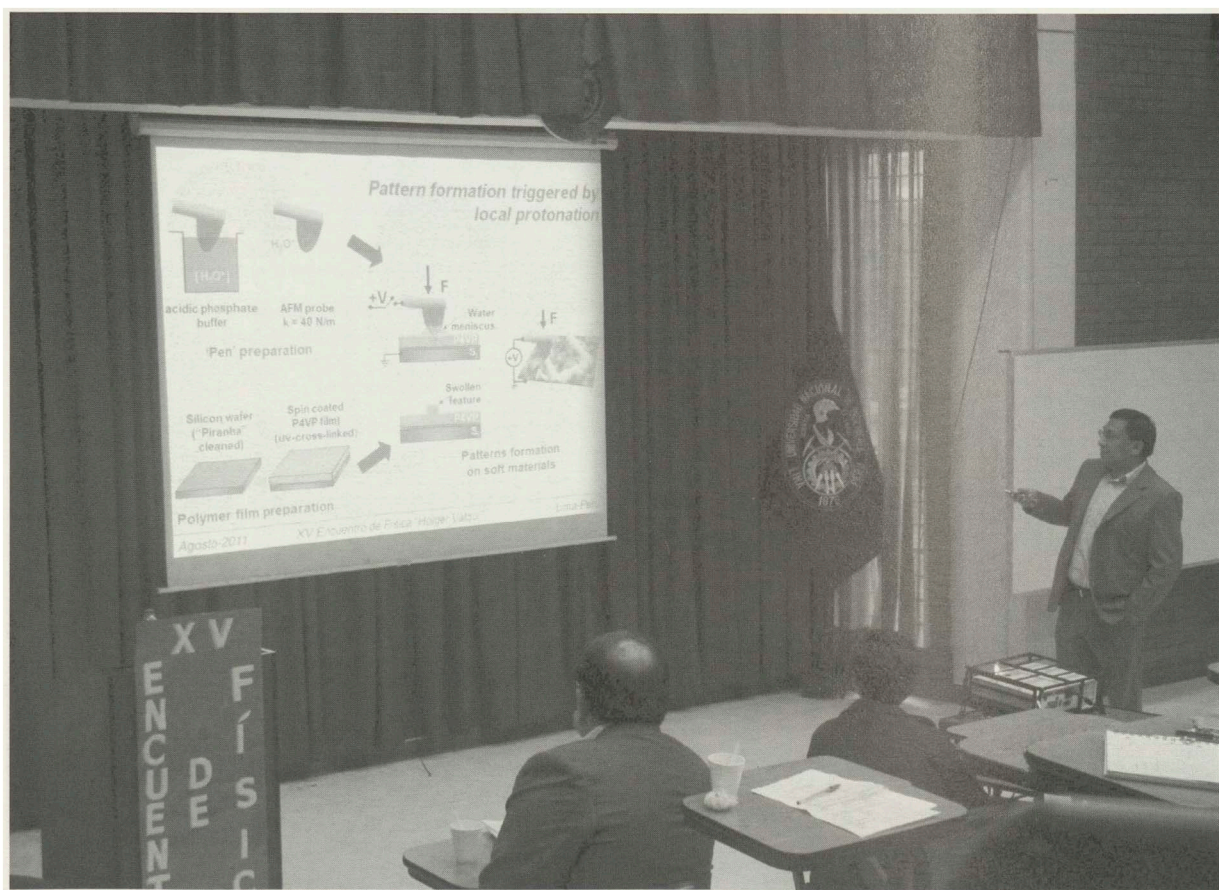
Este es el fruto de un trabajo dominical, que refuerza el aforismo que dice: Las personas muy ocupadas siempre encuentran tiempo para hacer lo que les gusta o lo que consideran importante.

Aunque siempre es dudoso que los demás lleguen a compartir mis gustos, confío en que, al amparo de las aventuras matemáticas, haya quienes disfruten de las situaciones presentadas. Propiamente sólo se requieren los conocimientos que suelen adquirirse en secundaria. Propiamente.

www.guzlop-editoras.com





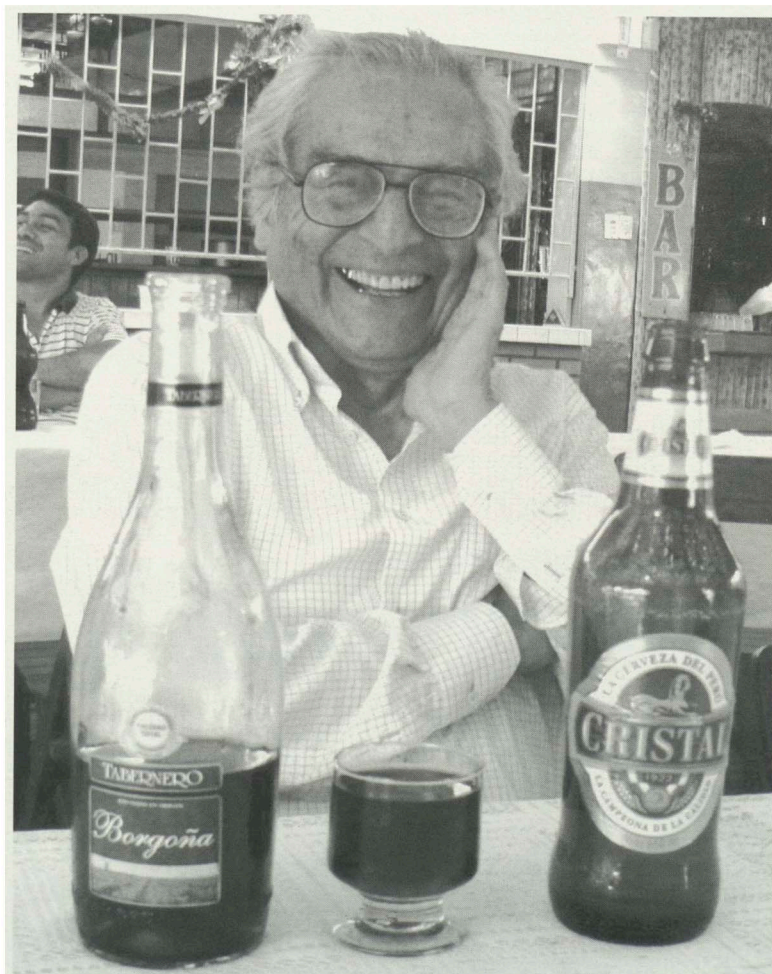






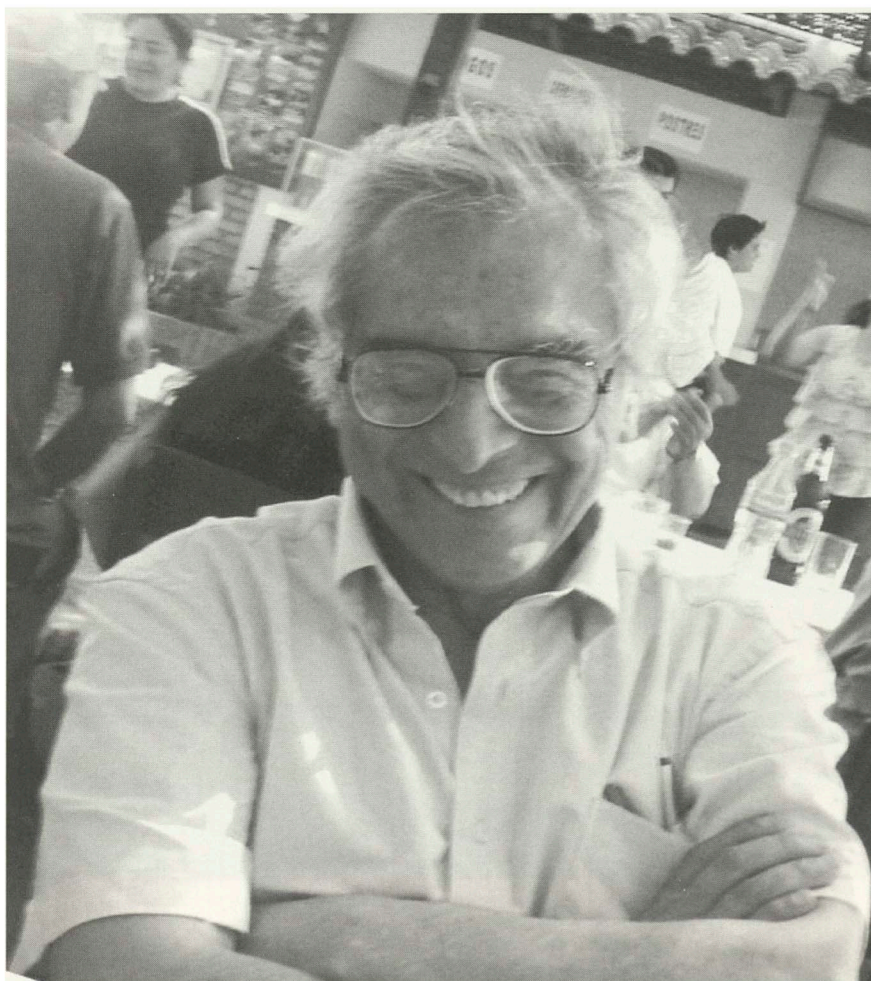
More pictures





















Reglas para la Preparación de Artículos para la Revista REVCIUNI

En la revista REVCIUNI se publican artículos de investigación actual y divulgación científica, básica o aplicada, en las áreas de Física, Matemática, Química, Ciencia de la Computación, y afines. Los artículos se reciben en el Instituto de Investigación de la Facultad de Ciencias de la UNI.

Los artículos deben de ser originales, inéditos, que no se hayan publicado previamente ni se encuentran bajo consideración para ser publicados en otras revistas. Los artículos no deben de presentar conclusiones conocidas, triviales, obvias y/o sin fundamento.

Los artículos serán recibidos por el Comité Científico el cual los enviará a uno o más árbitros para su revisión. El Comité Científico comunicará a los autores que sometieron el artículo la decisión de publicación así, como las observaciones de los árbitros. Todos los artículos serán tratados de forma confidencial hasta su publicación.

Los artículos deben de ser escritos en laTex2e tal que la redacción y el formato del artículo deben seguir las siguientes indicaciones:

- El tipo de letra es normal Roman o equivalentes.
- Los márgenes son: de los lados derecho e izquierdo 1,5cm y de arriba y abajo 2cm.
- El título debe de estar centrado y escrito con letra normal de tamaño 14pt y negritas.
- Debajo del título deben de ir los nombres completos de los autores con letra normal tamaño 10pt. Después del nombre de cada autor, deben indicarse el lugar de trabajo y el correo electrónico con letra cursiva y tamaño 10pt.
- El resumen debe escribirse en inglés y/o español, con letra normal tamaño 9pt. con un ancho del texto de 16.2cm. debe contener entre 50 y 150 palabras e indicar al final las palabras claves. Primero va el resumen en el idioma en que se redactó el artículo.
- El texto se escribe con letra normal tamaño 10pt. En dos columnas separadas en 0.7cm. Y debe ser dividido en secciones numeradas con números arábigos. El nombre de las secciones deben ser escritas en negritas tamaño 12pt. y centradas. Las subsecciones con letras negritas y centradas. Se recomienda que los artículos contengan las siguientes secciones: Introducción, Conclusiones y Agradecimientos (esta última no se numera).
- Al último va la sección sin numerar designada como Apéndice: Nombre del apéndice, en caso de haber varios apéndices van en secciones designadas como Apéndice A, Apéndice B, etc.
- Las fórmulas deben ser numeradas con números arábigos entre paréntesis en la margen derecha. La referencia de las fórmulas en el texto debe de haberse colocado entre paréntesis su número correspondiente.
- Toda letra latina que se utiliza en las fórmulas debe estar escrita en cursiva.
- Las funciones seno, coseno, logaritmo natural, y otras en esta categoría, se escriben sen, cos, ln, etc.
- Las tablas y figuras se enumeran con números arábigos. En la parte superior de la tabla e inferior de la figura, se colocará: Tabla y Figura correspondiente después el número, punto con letra negrita y la leyenda con letra cursiva, todo en tamaño 10pt.
- Las citas del texto se hacen colocando el número correspondiente de la lista de referencias entre corchetes
- Lista de referencias
 - Las referencias que se citan en el artículo es con número arábigos, en el orden de citación y va al final del artículo debajo de una línea horizontal, en dos columnas separadas en 0,7cm. El tamaño de las letras es de 9pt.
S. Frittelli, C. Kozameh, and E.T. Newman. J Matth. Phys. 36 (1995) 4975.
 - Cada entrada en la lista de referencias debe estar citada en el texto.
 - Las comunicaciones personales se citan en el texto, pero no se incluyen en la lista de referencias.
 - Cada referencia tiene el formato de párrafo francés (hanging indent) y a doble espacio.
 - Orden alfabético por apellido del primer autor.

- Obras de un mismo autor se ordenan cronológicamente por la fecha más antigua.
 - Apellidos primero, seguidos de las iniciales del nombre. Hasta tres autores.
 - Si tiene más de tres autores, se escribe el nombre del primer autor seguido de la expresión *et. al.*
 - Se utiliza el signo & antes del último autor. En español, se acepta la *y* en vez de &. En inglés, se acepta la *and* en vez de &.
- Parry, J. H., & Sherlock, P. (1976). *Historia de las Antillas*. Buenos Aires, Argentina: Kapelusz.
- En el caso de que la obra no tenga un autor, se coloca primero el título de la obra y luego la fecha.
 - Después de los nombres de los autores se coloca el nombre de la revista o libro con indicación al volumen, páginas y año.
 - Cuando la referencia es a un capítulo de un libro editado, se escribe el nombre del editor sin invertirlo, y precedido por la abreviatura Ed.
- V.I. Arnold, *Mathematical Methods of Classical Mechanics* (Ed. Springer, Berlín, 1980).

Detalles mas completos para la preparación de los artículos se encuentran en la página web <http://fc.edu.pe/revciuni>. Los artículos serán presentados previamente para su revisión en formato Portable Document Format (PDF) al e-mail: postgradofc@uni.edu.pe.

TABLE OF CONTENTS

Virgin photons do not exist <i>Holger G. Valqui</i>	1-6
Direct profiling of polarization fields in nitride semiconductors at nanometric scale using electron holography in the transmission electron microscope <i>F. A. Ponce</i>	7-15
Semiconductor Lasers <i>Carmen Gonzales</i>	16-17
Proton-fountain Electric-field-assisted Nanolithography (PEN) <i>A. La Rosa and M. Yan</i>	18-44
Perspectives in Industrial Nano-materials <i>Jose A. Alarco</i>	45-56
Transmission of electric energy without using wires <i>Ener Salinas</i>	57-65
Fission-track dating applied to Peruvian volcanic glasses <i>C. A. Tello et al.</i>	66-73
Reactivity, energetics and molecular structure: A theoretical and experimental approach <i>Juan Z. Davalos Prado</i>	74-81
Proton-fountain Electric-field-assisted Nanolithography (PEN): Fabrication of polymer nanostructures that respond to chemical and electrical stimuli. <i>Andres La Rosa, Mingdi Yan, Rodolfo Fernandez , Xiaohua Wang and Elia Zegarra</i>	82-123
Two dimensional BF gravity: A Hamilton-Jacobi analysis <i>M.C.Bertin, B.M.Pimentel and C.E.Valcarcel</i>	124-132
Higgs-Strahlung process in the $SU(3)_L \otimes U_N(1)$ Model with Heavy Leptons. <i>D. Romero and O. Pereyra</i>	133-143
Non commutative (generalized) sine-Gordon / massive Thirring correspondence, integrability and solitons <i>H. Blas and H.L. Carrion</i>	144-181
The variable mass Thirring / sine Gordon duality and continuous topological configurations <i>H. Blas and J.M. Jaramillo</i>	182-195
Breaking the Λ CDM degeneracy using CMB data <i>A. Bermi</i>	196-203
Generación de soluciones exactas de las Ecuaciones de Einstein con ayuda de euclídones <i>Paulyac Huaman, J. Anibal</i>	204-208
OA-UNI: an Astronomical Observatory at UNI <i>A. Pereyra, W. Cori, E. Meza, J. Ricra, and G. Granda</i>	209-211
Generalized AKNS System, Non-vanishing Boundary Conditions and N-Dark-Dark Solitons <i>A.de O. Assuncao, H. Blas and M. J. B. F. da Silva</i>	212-241
AKNS solitons with constant boundary conditions via dressing transformation and tau function approach <i>H.Blas and M.Zambrano</i>	242-250
Formación de Comunidades Latinoamericanas de Investigación, La experiencia de CLARA y las posibilidades de la RAAP <i>Benjamyn Marticorena</i>	251-256
Pictures of professor Valqui friends and colleagues	257-274