Lecture Notes in Physics

Editorial Board

R. Beig, Wien, Austria
J. Ehlers, Potsdam, Germany
U. Frisch, Nice, France
K. Hepp, Zürich, Switzerland
W. Hillebrandt, Garching, Germany
D. Imboden, Zürich, Switzerland
R. L. Jaffe, Cambridge, MA, USA
R. Kippenhahn, Göttingen, Germany
R. Lipowsky, Golm, Germany
H. v. Löhneysen, Karlsruhe, Germany
I. Ojima, Kyoto, Japan
H. A. Weidenmüller, Heidelberg, Germany
J. Wess, München, Germany
J. Zittartz, Köln, Germany

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Physics and Astronomy ONLINE LIBRARY
http://www.springer.de/phys/

The Editorial Policy for Proceedings

The series Lecture Notes in Physics reports new developments in physical research and teaching – quickly, informally, and at a high level. The proceedings to be considered for publication in this series should be limited to only a few areas of research, and these should be closely related to each other. The contributions should be of a high standard and should avoid lengthy redraftings of papers already published or about to be published elsewhere. As a whole, the proceedings should aim for a balanced presentation of the theme of the conference including a description of the techniques used and enough motivation for a broad readership. It should not be assumed that the published proceedings must reflect the conference in its entirety. (A listing or abstracts of papers presented at the meeting but not included in the proceedings could be added as an appendix.)

When applying for publication in the series Lecture Notes in Physics the volume's editor(s) should submit sufficient material to enable the series editors and their referees to make a fairly accurate evaluation (e.g. a complete list of speakers and titles of papers to be presented and abstracts). If, based on this information, the proceedings are (tentatively) accepted, the volume's editor(s), whose name(s) will appear on the title pages, should select the papers suitable for publication and have them refereed (as for a journal) when appropriate. As a rule discussions will not be accepted. The series editors and Springer-Verlag will normally not interfere with the detailed editing except in fairly obvious cases or on technical matters.

Final acceptance is expressed by the series editor in charge, in consultation with Springer-Verlag only after receiving the complete manuscript. It might help to send a copy of the authors' manuscripts in advance to the editor in charge to discuss possible revisions with him. As a general rule, the series editor will confirm his tentative acceptance if the final manuscript corresponds to the original concept discussed, if the quality of the contribution meets the requirements of the series, and if the final size of the manuscript does not greatly exceed the number of pages originally agreed upon. The manuscript should be forwarded to Springer-Verlag shortly after the meeting. In cases of extreme delay (more than six months after the conference) the series editors will check once more the timeliness of the papers. Therefore, the volume's editor(s) should establish strict deadlines, or collect the articles during the conference and have them revised on the spot. If a delay is unavoidable, one should encourage the authors to update their contributions if appropriate. The editors of proceedings are strongly advised to inform contributors about these points at an early stage.

The final manuscript should contain a table of contents and an informative introduction accessible also to readers not particularly familiar with the topic of the conference. The contributions should be in English. The volume's editor(s) should check the contributions for the correct use of language. At Springer-Verlag only the prefaces will be checked by a copy-editor for language and style. Grave linguistic or technical shortcomings may lead to the rejection of contributions by the series editors. A conference report should not exceed a total of 500 pages. Keeping the size within this bound should be achieved by a stricter selection of articles and not by imposing an upper limit to the length of the individual papers. Editors receive jointly 30 complimentary copies of their book. They are entitled to purchase further copies of their book at a reduced rate. As a rule no reprints of individual contributions can be supplied. No royalty is paid on Lecture Notes in Physics volumes. Commitment to publish is made by letter of interest rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume.

The Production Process

The books are hardbound, and the publisher will select quality paper appropriate to the needs of the author(s). Publication time is about ten weeks. More than twenty years of experience guarantee authors the best possible service. To reach the goal of rapid publication at a low price the technique of photographic reproduction from a camera-ready manuscript was chosen. This process shifts the main responsibility for the technical quality considerably from the publisher to the authors. We therefore urge all authors and editors of proceedings to observe very carefully the essentials for the preparation of camera-ready manuscripts, which we will supply on request. This applies especially to the quality of figures and halftones submitted for publication. In addition, it might be useful to look at some of the volumes already published. As a special service, we offer free of charge LaTeX and TeX macro packages to format the text according to Springer-Verlag's quality requirements. We strongly recommend that you make use of this offer, since the result will be a book of considerably improved technical quality. To avoid mistakes and time-consuming correspondence during the production period the conference editors should request special instructions from the publisher well before the beginning of the conference. Manuscripts not meeting the technical standard of the series will have to be returned for improvement.

For further information please contact Springer-Verlag, Physics Editorial Department II, Tiergartenstrasse 17, D-69121 Heidelberg, Germany

Series homepage – http://www.springer.de/phys/books/lnpp

Michel Planat (Ed.)

Noise, Oscillators and Algebraic Randomness

From Noise in Communication Systems to Number Theory

Lectures of a School Held in Chapelle des Bois, France, April 5–10, 1999

Springer

Editor

Michel Planat
Laboratoire de Physique et Metrologie
des Oscillateurs du CNRS
32 Avenue de l'Observatoire
25044 Besançon Cedex, France

Library of Congress Cataloging-in-Publication Data.

Noise, oscillators, and algebraic randomness : from noise in communication systems to number theory : lectures of a school held in Chapelle des Bois, France, April 5-10, 1999 / Michel Planat (ed.).
p. cm. – (Lecture notes in physics, ISSN 0075-8540 ; vol. 550)
Includes bibliographical references.
ISSN 3540675728 (alk. paper)
1. Electronic noise–Mathematical models–Congresses. 2.Oscillators, Electric–Congresses. 3. Numerical analysis–Congresses. 4. Algebra–Congresses. 5. Telecommunication–Mathematics–Congresses. I. Planat, Michel, 1951 - II. Lecture notes in physics ; 550.

TK7867.5 .N627 2000
621.382'24–dc21

00-032966

ISSN 0075-8450
ISBN 3-540-67572-8 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag is a company in the BertelsmannSpringer publishing group.
© Springer-Verlag Berlin Heidelberg 2000
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by the authors/editor
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
SPIN: 10719295 57/3144/du - 5 4 3 2 1 0

Preface

This volume presents most contributions given at the School "Noise of frequencies in oscillators and the dynamics of algebraic numbers" which was held in Chapelle des Bois, Jura (France) from 5 to 10 April 1999. The event was made possible by the full support of the thematic school program of the Centre National de la Recherche Scientifique in France.

Noise is ubiquitous in nature and in man-made systems. For example, noise in oscillators perturbs high technology devices such as time standards or digital communication systems. The understanding of its algebraic structure is thus of vital importance to properly guide the human activity.

The book addresses these topics in three parts. Several aspects of classical and quantum noise are covered in Part I, both from the viewpoint of quantum physics and from the nonlinear or fractal viewpoint. Part II is mainly concerned with noise in oscillating signals, that is phase or frequency noise and 1/f noise. From a careful analysis of the experimental noise attached to the carrier the usefulness of the number theoretical based method is demonstrated. This view is expanded in Part III, which is mathematically oriented. In conclusion, the noise concept proved to be a very attractive one gathering people from at least three scientific communities: electronic engineering, theoretical physics and number theory. They represented an original mixture of talents and the present editor acknowledges all authors for their patience and open-mindedness during the school. Most manuscripts are comprehensible to a large audience and should allow readers to discover new bridges between the fields. We ourselves have identified but a few.

The meeting was followed by a small workshop sponsored by M. Waldschmidt at the Institut Henri Poincaré in Paris on 3 and 4 December 1999: 'Théorie des nombres, bruit des fréquences et télécommunications '. The purpose here was to emphasize the newly discovered link between the Riemann zeta function and communication systems. Some papers and related material are available at the address: http://www.archetypo.web66.com, a new URL site built by Matthew Watkins and devoted to the relationship between prime number theory and physics.

Besançon, April 2000 *Michel Planat*

Contents

Introduction
Michel Planat 1

Mathemagics (A Tribute to L. Euler and R. Feynman)
Pierre Cartier 6

Part I Classical and Quantum Noise

Thermal and Quantum Noise in Active Systems
Jean-Michel Courty, Francesca Grassia, Serge Reynaud 71

Dipole at $\nu = 1$
V. Pasquier 84

Stored Ion Manipulation Dynamics of Ion Cloud and Quantum Jumps with Single Ions
Fernande Vedel 107

$1/f$ Fluctuations in Cosmic Ray Extensive Air Showers
E. Faleiro, J.M.G. Gómez 125

Stochastic Resonance and the Benefit of Noise in Nonlinear Systems
François Chapeau-Blondeau 137

Time is Money
Marcel Ausloos, Nicolas Vandewalle, Kristinka Ivanova 156

Part II Noise in Oscillators, 1/f Noise and Arithmetic

Oscillators and the Characterization of Frequency Stability: an Introduction
Vincent Giordano, Enrico Rubiola 175

Phase Noise Metrology
Enrico Rubiola, Vincent Giordano 189

Phonon Fine Structure in the $1/f$ Noise of Metals, Semiconductors and Semiconductor Devices
Mihai N. Mihaila 216

The General Nature of Fundamental $1/f$ Noise in Oscillators and in the High Technology Domain
Peter H. Handel 232

$1/f$ Frequency Noise in a Communication Receiver and the Riemann Hypothesis
Michel Planat 265

Detection of Chaos in the Noise of Electronic Oscillators by Time Series Analysis Methods
C. Eckert, M. Planat 288

Geometry and Dynamics of Numbers Under Finite Resolution
Jacky Cresson, Jean-Nicolas Dénarié 305

Diophantine Conditions and Real or Complex Brjuno Functions
Pierre Moussa, Stefano Marmi 324

Part III Algebraic Randomness

Algebraic and Analytic Randomness
Jean-Paul Allouche 345

From Symbolic Dynamics to a Digital Approach
K. Karamanos 357

Algebraic Dynamics and Transcendental Numbers
Michel Waldschmidt 372

Dynamics of Some Contracting Linear Functions Modulo 1
Yann Bugeaud, Jean-Pierre Conze 379

On the Modular Function and Its Importance for Arithmetic
Paula B. Cohen 388

On Generalized Markoff Equations and Their Interpretation
Serge Perrine 398

Introduction

Michel Planat *

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, 32 Avenue de l'Observatoire, 25044 Besançon Cedex, France

In the classical realm the best known type of noise is thermal noise. Due to thermal agitation, free electrons in a metallic conductor are moving around continuously causing collisions with the atoms and a continuous exchange of energy between the modes. This was first investigated experimentally by J.B. Johnson and H. Nyquist in 1928. Nyquist's theorem states that the power spectral density (psd) of voltage fluctuations through a resistor R at temperature T is $S_V(f) = 4kRT$, with k the Boltzman's constant. The quantum approach is also a very efficient way to understand the limitation in the accuracy of measurements performed in thermal equilibrium. The paper by J.M. Courty et al. examines the fluctuations which are expected in an operational amplifier from a quantum network approach. The ultimate sensitivity of a cold damped accelerometer designed for space applications is calculated as well.

Quantum physics is also used to understand the quantum Hall effect, that is the behaviour of charged interacting electrons in the presence of a strong magnetic field. The indivisibility of the electron charge e may be demonstrated from a measurement of the current power spectral density (psd) $S_I(f) = 2eI$ which is known as shot noise (Schottky, 1918). Such noise results from the random emission of electrons from the cathode to the anode in a diode or a semiconductor. Similarly, partition noise is added to the measurement whenever a current is distributed randomly between two electrodes. Its net effect is to introduce an extra multiplicative factor in the relation for the shot noise psd. Partition noise measurements in small size quantum conductors have recently revealed the existence of fractional charges $e\frac{p}{q}$ (p and q integers) associated to quasi particle tunneling states. The paper presented at the school by C. Glattli was published elsewhere (Phys. Rev. Lett. 79, 2526 (1997) and in "Topological aspects of low dimensional systems", Proceedings of Les Houches 1998 Summer School). The theory of the fractional quantum Hall effect is still very open to debate, as shown in the paper by V. Pasquier which adresses the problem of bosonic particles interacting repulsively at the filling factor $\nu = 1$.

In his early study of thermal noise J.B. Johnson also found a large amount of voltage noise $S_V(f) \sim KV^2/f$ at low Fourier frequencies f. From many experiments it was found that the noise intensity is inversely proportional to the number of carriers N in the sample, that is $K \sim \gamma/N$, with γ in

* planat@lpmo.univ-fcomte.fr

M. Planat (Ed.): LNP 550, pp. 1–5, 2000.
© Springer-Verlag Berlin Heidelberg 2000

the range 10^{-3} to 10^{-8}. This was generally attributed to different scattering mechanisms, by the crystal lattice or the impurities, leading to mobility fluctuations of the electrons. These findings point to a nonlinear origin of the $1/f$ noise. Fine structures revealing the interaction of electrons with bulk and surface phonons in several solid-state physical systems are given in Mihaila's paper in Part II. On the theoretical side, a quantum electrodynamical theory was developed by P. Handel in the seventies based on infrared divergent coupling of the electrons to the electromagnetic field in the scattering process. The basic result for the γ parameter involves the fine structure constant $\alpha \sim 1/137$ times the square of the ratio between the change of the velocity of the accelerated charge over the light velocity (see the introduction of Handel's paper). This so-called (by him) conventional quantum $1/f$ effect applies to small solid-state devices with K of the order 10^{-7}. For large samples K increases to $2\alpha/\pi \sim 4.6 \times 10^{-3}$ which is the value predicted by the coherent state approach of the quantum $1/f$ effect (see Fig. 2 in Handel's paper).

Charged particles can be kept free and interrogated for very long times in a miniature trap as shown in the paper by F. Vedel. Synchronized and chaotic states of the oscillating ions are studied in such a set-up. Using laser cooling with a few ions, the technique also allows the study of quantum jumps and the design of a very accurate clock.

High energy particles and nuclei reaching earth from space are called cosmic rays. Their energy spectrum is very broad (from 10^9 to 10^{20} eV); they are emitted from multiplicative cascades (cosmic showers) with a variety of disintegrations as shown in the paper by J. Gomez. At the ground level particle densities show fluctuations with a $1/f$ power spectrum in the polar angle coordinate. This a new example of the deep relationship between nonlinearity and $1/f$ noise.

The paper by F. Chapeau-Blondeau studies the intrinsically nonlinear link between signal and noise in a variety of systems. The word stochastic resonance has been coined to describe situations in which the noise can benefit the information-carrying signal. The ability to increase the signal to noise ratio from noise enhancement in a non linear information channel or an image is conclusively demonstrated.

It is not so well known that the first mathematical study of Brownian motion, which is due to Bachelier (prior to Einstein) in 1900 concerned the pricing of options in speculative markets. Anomalous (non-Gaussian) distributions are the rule in stock market data as shown in the paper by M. Ausloos.

The papers by V. Giordano (G), E. Rubiola (R), M. Planat (P), C. Eckert (E) and J. Cresson (C) in Part II are closely related. They mainly concern the understanding of the building block of an electronic oscillator, that is a resonant cavity (a quartz crystal) and a sustaining amplifier. Time and frequency metrology was born in an effort to improve the design and performance of such oscillators used as accurate clocks (G). Instruments to measure

phase noise on such oscillators have gained a high level of sophistication (R). Moreover the scheme of an oscillator is similar to the basic constituent of a communication receiver (P). The dynamics of frequency and amplitude states has the appearance of a low dimensional deterministic system (E), but the actual rules mimic analytical number theory and the properties of Riemann zeta function (P). In essence this can be understood from the physical limit of any physical measurement: the finite resolution (C).

The papers by M. N. Mihaila (M) and P. Handel (H) remind us that the question of the origin of $1/f$ noise is as old as electronics and is very universal. Nonlinear interactions between the lattice phonons and thermal phonons are clearly involved (M). Besides in the quantum $1/f$ effect, the basic nonlinear system is the charged particle interacting with the emitted field which reacts back on the source particle (M). One way to experimentally study the coupling between the oscillating particles is through nonlinear mixing and low pass filtering (P). The information-carrying oscillator of frequency f_0, when mixed with a reference oscillator of frequency f_1, produces all tones at beat frequencies $f_i = |pf_0 - qf_1|$, which after normalizing with respect to f_1 looks similar to the problem of approximating real numbers from rational numbers but with a finite resolution (C).

The observed approximations are of the diophantine type (as it is the case for resonances in celestial mechanics) and are calculated by truncating the continued fraction expansion of the frequency ratio of the oscillators at the mixer inputs. The standard map introduced by B. Chirikov in 1979 is an alternative way to describe the nonlinear coupling between two oscillators as shown in the paper by P. Moussa (M). It allows one to express the diophantine problem in terms of the Brjno function introduced by J.C. Yoccoz in 1995 to linearize a complex holomorphic map around a fixed point, and measure the radius of the associated Siegel disk at the resonance (M).

To measure the ability of continued fraction expansion (cfe) to approximate a real number one may introduce the Markoff constant A which is the asymptotic limit (when they are infinitely many terms in the cfe) of the error term modulus times the square of the partial quotient denominator. The most badly approximated numbers are $\sqrt{5}-1$ with infinitely many 1's in the cfe, $\sqrt{2}-1$ with infinitely many 2's in the cfe, $(\sqrt{221}-11)/10$ with periodicity $(2,1,1,2)$ in the cfe and so on. Getting the whole theory is a formidable task which is attempted in the paper by S. Perrine in Part III. The first-order theory was obtained by Markoff in 1880. It predicts Markoff numbers at $A_i = a_i(9a_i^2-4)^{1/2}$ where the a_i's are $1, 2, 5, 13, 29...$ and satisfy the algebraic condition $a^2+b^2+c^2 = 3abc$ and are the traces of matrices in a subgroup of index 6 in the modular group $SL(2,\mathbf{Z})$ [in topological terms it is a punctured torus as it is nicely explained in Gutzwiller's book (*Chaos in classical and quantum mechanics*, Springer, 1990)].

Frequency fluctuations of the oscillators are often characterized in the frequency domain from the power spectral density or in the time domain from

the Allan deviation (this averages the mean frequency deviation between two consecutive samples, each one measured over an integration time τ). The two measures are related; white noise (that is constant psd) corresponds to Allan deviation proportional to $\tau^{-1/2}$ and $1/f$ noise corresponds to constant Allan deviation versus τ. White noise below the thermal floor is measured using an interferometric method and correlation analysis (R). A transition from white to $1/f$ frequency noise at the resonance is observed in the electronic receiver. This transition corresponds taking bounded partial quotients in the cfe of the frequency ratio of the oscillators at the mixer inputs (P). This is interpreted in terms of the position of resolved rationals with respect to the equally spaced graduation and is equivalent to Riemann hypothesis (as expressed from the Franel-Landau sums)(P).

It is not very surprising to encounter the Riemann zeta function in physics. The ordinary Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$, with $\Re(s) > 1$ is present in the black body radiation laws. The number of photons per unit volume is proportional to $\zeta(3)$ and the energy to $\zeta(4)$. Similarly in a Bose-Einstein condensate the number of modes is proportional to $\zeta(3/2)$ and the energy to $\zeta(5/2)$. The argument of the zeta function also defines the exponent in the temperature dependence. Now the Casimir (vacuum) energy between two parallel conducting plates is essentially $\zeta(-3)$ which requires a first extension of $\zeta(s)$ to lower than 1 integer values of the argument s. This is achieved thanks to the connection of $\zeta(s)$, with s a relative integer, to Bernouilli numbers; they are defined from the algebraic expansion of the "Planck" factor $x/(e^x - 1)$ (see also Cartier's paper).

Looking at $\zeta(s)$ as a partition function (this was emphasized by B. Julia in 1994 (P)) of some "Riemann gas" with energies $\log\ i$ (instead of i in a conventional quantum harmonic oscillator), thermodynamical quantities are proportional to the inverse of the partition function so that the zeros of $\zeta(s)$ should play the dominant role in the dynamics. Since the psd of $|1/\zeta(s)|$ is a "white noise" at the critical line $s = \frac{1}{2} + it$ and transforms into a "$1/f$ noise" close to it, it is tempting to expect that $\zeta(s)$ should play a role in the unification of physics.

Can we find an algebraic definition of randomness? This is attempted in the paper by J.P. Allouche in Part III restricting the study to infinite sequences taking their values in a finite alphabet, as it is the case in digital communications. What is the alphabet in models of deterministic chaos? According to the paper by K. Karamanos the chaoticity of the symbolic sequence in the Feigenbaum bifurcation diagram (which is also a model of phase noise) has much to do with transcendance and thus with rational polynomials. This is further elaborated in Waldschmidt's paper in terms of the logarithm of Mahler's measure on such polynomials, which also expresses the topological entropy of an algebraic dynamical system.

An important ingredient in the theory of continued fraction expansions is the concept of a Farey mediant $\frac{p+p'}{q+q'}$ of two rational numbers $\frac{p}{p'}$ and $\frac{q}{q'}$. They are found recursively in the structure of the electronic receiver (P) and are intimely connected to the structure of the modular group $SL(2, \mathbf{Z})$ (C), (M). A further example is in the paper by Y. Bugeaud about a specific type of analog-to-digital converter.

Numbers with a periodic cfe beyond some level are the algebraic numbers of degree 2. They play a major role in the Markoff spectrum (see the paper by S. Perrine). A dual role is played by imaginary quadratic integers τ defined on the upper half plane $\Im(\tau) > 0$ as it is shown in the paper by P. Cohen. At such numbers the modular function $j(\tau)$ (which is the automorphic function with respect to the full modular group $SL(2, \mathbf{Z})$) takes an algebraic value, that is the associated class of elliptic curves has complex multiplication; and conversely $j(\tau)$ is transcendental if τ is not quadratic imaginary.

Mathemagics defined by P. Cartier as the symbolic methods of mathematics or operational calculus played a fundamental role in the development of physics. Heisenberg's generalization of Hamiltonian mechanics is one example. In an extensive and magistral paper P. Cartier examines the development of such formal methods from Euler to Feynman.

Mathemagics
(A Tribute to L. Euler and R. Feynman)

Pierre Cartier *

CNRS, Ecole Normale Supérieure de Paris, 45 rue d'Ulm, 75230 Paris Cedex 05
and Institut des Hautes Etudes Scientifiques, Le Bois Marie, 35 Route de
Chartres, 91440 Bures-sur-Yvette, France

1 Introduction

The implicit philosophical belief of the working mathematician is today the Hilbert-Bourbaki formalism. Ideally, one works within a **closed system**: the basic principles are clearly enunciated once for all, including (that is an addition of twentieth century science) the formal rules of logical reasoning clothed in mathematical form. The basic principles include precise definitions of all mathematical objects, and the coherence between the various branches of mathematical sciences is achieved through reduction to basic models in the universe of sets. A very important feature of the system is its **non-contradiction** ; after Gödel, we have lost the initial hopes to establish this non-contradiction by a formal reasoning, but one can live with a corresponding belief in non-contradiction. The whole structure is certainly very appealing, but the illusion is that it is eternal, that it will function for ever according to the same principles. What history of mathematics teaches us is that the principles of mathematical deduction, and not simply the mathematical theories, have evolved over the centuries. In modern times, theories like General Topology or Lebesgue's Integration Theory represent an almost perfect model of precision, flexibility and harmony, and their applications, for instance to probability theory, have been very successful.

My thesis is: **there is another way of doing mathematics, equally successful, and the two methods should supplement each other and not fight.**

This other way bears various names: symbolic method, operational calculus, operator theory ... Euler was the first to use such methods in his extensive study of infinite series, convergent as well as divergent. The calculus of differences was developed by G. Boole around 1860 in a symbolic way, then Heaviside created his own symbolic calculus to deal with systems of differential equations in electric circuitry. But the modern master was R. Feynman who used his diagrams, his disentangling of operators, his path integrals ... The method consists in stretching the formulas to their extreme consequences, resorting to some internal feeling of coherence and harmony. They are obvious pitfalls in such methods, and only experience can tell you

* cartier@ihes.fr

M. Planat (Ed.): LNP 550, pp. 6–67, 2000.
© Springer-Verlag Berlin Heidelberg 2000

that for the Dirac δ-function an expression like $x\delta(x)$ or $\delta'(x)$ is lawful, but not $\delta(x)/x$ or $\delta(x)^2$. Very often, these so-called symbolic methods have been substantiated by later rigorous developments, for instance Schwartz distribution theory gives a rigorous meaning to $\delta(x)$, but physicists used sophisticated formulas in "momentum space" long before Schwartz codified the Fourier transformation for distributions. The Feynman "sums over histories" have been immensely successful in many problems, coming from physics as well from mathematics, despite the lack of a comprehensive rigorous theory.
To conclude, I would like to offer some remarks about the word "formal". For the mathematician, it usually means "according to the standard of formal rigor, of formal logic". For the physicists, it is more or less synonymous with "heuristic" as opposed to "rigorous". It is very often a source of misunderstanding between these two groups of scientists.

2 A new look at the exponential

2.1 The power of exponentials

The multiplication of numbers started as a shorthand for repeated additions, for instance 7 times 3 (or rather "seven taken three times") is the sum of three terms equal to 7

$$7 \times 3 = \underbrace{7 + 7 + 7}_{3 \text{ times}}.$$

In the same vein 7^3 (so denoted by Viete and Descartes) means $\underbrace{7 \times 7 \times 7}_{3 \text{ factors}}$. There is no difficulty to define x^2 as xx or x^3 as xxx for any kind of multiplication (numbers, functions, matrices ...) and Descartes uses interchangeably xx or x^2, xxx or x^3.

In the exponential (or power) notation, the exponent plays the role of an operator. A great progress, taking approximately the years from 1630 to 1680 to accomplish, was to generalize a^b to new cases where the operational meaning of the exponent b was much less visible. By 1680, a well defined meaning has been assigned to a^b for a, b real numbers, $a > 0$. Rather than to retrace the historical route, we shall use a formal analogy with vector algebra. From the original definition of a^b as $a \times ... \times a$ (b factors), we deduce the fundamental rules of operation, namely

$$(a \times a')^b = a^b \times a'^b,\ a^{b+b'} = a^b \times a^{b'},\ (a^b)^{b'} = a^{bb'},\ a^1 = a. \tag{1}$$

The other rules for manipulating powers are easy consequences of the rules embodied in (1). The fundamental rules for **vector algebra** are as follows:

$$\begin{aligned}&(\mathbf{v} + \mathbf{v}').\lambda = \mathbf{v}.\lambda + \mathbf{v}'.\lambda,\ \mathbf{v}.(\lambda + \lambda') = \mathbf{v}.\lambda + \mathbf{v}.\lambda',\\ &(\mathbf{v}.\lambda).\lambda' = \mathbf{v}.(\lambda\lambda'),\ \mathbf{v}.1 = \mathbf{v}.\end{aligned} \tag{2}$$

The analogy is striking provided we compare the product $a \times a'$ of numbers to the sum $\mathbf{v} + \mathbf{v}'$ of vectors, and the exponentiation a^b to the scaling $\mathbf{v}.\lambda$ of the vector $\mathbf{v}$ by the scalar λ.

In modern terminology, to define a^b for a, b real, $a > 0$ means that we want to consider the set $\mathbf{R}_+^\times$ of real numbers $a > 0$ as a vector space over the field of real numbers $\mathbf{R}$. But to vectors, one can assign coordinates: if the coordinates of the vector $\mathbf{v}(\mathbf{v}')$ are the $v_i(v_i')$, then the coordinates of $\mathbf{v} + \mathbf{v}'$ and $\mathbf{v}.\lambda$ are respectively $v_i + v_i'$ and $v_i.\lambda$. Since we have only one degree of freedom in $\mathbf{R}_+^\times$, we should need one coordinate, that is a bijective map L from $\mathbf{R}_+^\times$ to $\mathbf{R}$ such that

$$L(a \times a') = L(a) + L(a'). \tag{3}$$

Once such a logarithm L has been constructed, one defines a^b in such a way that $L(a^b) = L(a).b$. It remains the daunting task to construct a logarithm. With hindsight, and using the tools of calculus, here is the simple definition of "natural logarithms"

$$\ln(a) = \int_1^a dt/t \quad \text{for } a > 0. \tag{4}$$

In other words, the logarithm function $\ln(t)$ is the primitive of $1/t$ which vanishes for $t = 1$. The inverse function $\exp s$ (where $t = \exp s$ is synonymous to $\ln(t) = s$) is defined for all real s, with positive values, and is the unique solution to the differential equation $f' = f$ with initial value $f(0) = 1$. The final definition of powers is then given by

$$a^b = \exp(\ln(a).b). \tag{5}$$

If we denote by e the unique number with logarithm equal to 1 (hence $e = 2.71828...$), the exponential is given by $\exp a = e^a$.

The main character in the exponential is the exponent, as it should be, in complete reversal from the original view where 2 in x^2, or 3 in x^3 are mere markers.

2.2 Taylor's formula and exponential

We deal with the expansion of a function $f(x)$ around a fixed value x_0 of x, in the form

$$f(x_0 + h) = c_0 + c_1 h + \cdots + c_p h^p + \cdots . \tag{6}$$

This can be an infinite series, or simply a finite order expansion (include then a remainder). If the function $f(x)$ admits sufficiently many derivatives, we can deduce from (6) the chain of relations

$$\begin{aligned} f'(x_0 + h) &= c_1 + 2c_2 h + \cdots \\ f''(x_0 + h) &= 2c_2 + 6c_3 h + \cdots \\ f'''(x_0 + h) &= 6c_3 + 24c_4 h + \cdots . \end{aligned}$$

By putting $h = 0$, deduce

$$f(x_0) = c_0, \quad f'(x_0) = c_1, \quad f''(x_0) = 2c_2, \ldots$$

and in general $f^{(p)}(x_0) = p!c_p$. Solving for the c_p's and inserting into (6) we get Taylor's expansion

$$f(x_0 + h) = \sum_{p \geq 0} \frac{1}{p!} f^{(p)}(x_0) h^p. \tag{7}$$

Apply this to the case $f(x) = \exp x$, $x_0 = 0$. Since the function f is equal to its own derivative f', we get $f^{(p)} = f$ for all p's, hence $f^{(p)}(0) = f(0) = e^0 = 1$. The result is

$$\exp h = \sum_{p \geq 0} \frac{1}{p!} h^p. \tag{8}$$

This is one of the most important formulas in mathematics. The idea is that this series can now be used to define the exponential of large classes of mathematical objects: complex numbers, matrices, power series, operators. For the modern mathematician, a natural setting is provided by a complete normed algebra A, with norm satisfying $||ab|| \leq ||a||.||b||$. For any element a in A, we define $\exp\ a$ as the sum of the series $\sum_{p \geq 0} a^p/p!$, and the inequality

$$||a^p/p!|| \leq ||a||^p/p! \tag{9}$$

shows that the series is absolutely convergent.

But this would not exhaust the power of the exponential. For instance, if we take (after Leibniz) the step to denote by $\mathbf{D}f$ the derivative of f, $\mathbf{D}^2 f$ the second derivative, etc... (another instance of the exponential notation!), then Taylor's formula reads as

$$f(x + h) = \sum_{p \geq 0} \frac{1}{p!} h^p \mathbf{D}^p f(x). \tag{10}$$

This can be interpreted by saying that the **shift operator** $\mathbf{T}_h$ taking a function $f(x)$ into $f(x+h)$ is equal to $\sum_{p \geq 0} \frac{1}{p!} h^p \mathbf{D}^p$, that is to the exponential $\exp h\mathbf{D}$ (question: who was the first mathematician to cast Taylor's formula in these terms?). Hence the obvious operator formula $\mathbf{T}_{h+h'} = \mathbf{T}_h.\mathbf{T}_{h'}$ reads as

$$\exp(h + h')\mathbf{D} = \exp h\mathbf{D}.\exp h'\mathbf{D}. \tag{11}$$

Notice that for numbers, the logarithmic rule is

$$\ln(a.a') = \ln(a) + \ln(a') \tag{12}$$

according to the historical aim of reducing via logarithms the multiplications to additions. By inversion, the exponential rule is

$$\exp(a + a') = \exp(a).\exp(a'). \tag{13}$$

Hence formula (10) is obtained from (12) by substituting $h\mathbf{D}$ to a and $h'\mathbf{D}$ to a'.

But life is not so easy. If we take two matrices A and B and calculate $\exp(A + B)$ and $\exp A.\exp B$ by expansion we get

$$\exp(A + B) = I + (A + B) + \frac{1}{2}(A + B)^2 + \frac{1}{6}(A + B)^3 + \cdots \tag{14}$$

$$\exp A.\exp B = I + (A + B) + \frac{1}{2}(A^2 + 2AB + B^2)$$
$$+\frac{1}{6}(A^3 + 3A^2B + 3AB^2 + B^3) + \cdots . \tag{15}$$

If we compare the terms of degree 2 we get

$$\frac{1}{2}(A + B)^2 = \frac{1}{2}(A^2 + AB + BA + B^2) \tag{16}$$

in (13) and not $\frac{1}{2}(A^2+2AB+B^2)$. Harmony is restored if A and B commute: indeed $AB = BA$ entails

$$A^2 + AB + BA + B^2 = A^2 + 2AB + B^2 \tag{17}$$

and more generally the **binomial formula**

$$(A + B)^n = \sum_{i=0}^{n} \binom{n}{i} A^i B^{n-i} \tag{18}$$

for any $n \geq 0$. By summation one gets

$$\exp(A + B) = \exp A.\exp B \tag{19}$$

if A and B commute, but not in general. The success in (10) comes from the obvious fact that hD commutes to $h'D$ since numbers commute to (linear) operators.

2.3 Leibniz's formula

Leibniz's formula for the higher order derivatives of the product of two functions is the following one

$$\mathbf{D}^n(fg) = \sum_{i=0}^{n} \binom{n}{i} \mathbf{D}^i f.\mathbf{D}^{n-i} g. \tag{20}$$

The analogy with the binomial theorem is striking and was noticed early. Here are possible explanations. For the shift operator, we have

$$\mathbf{T}_h = \exp h\mathbf{D} \tag{21}$$

by Taylor's formula and

$$\mathbf{T}_h(fg) = \mathbf{T}_h f.\mathbf{T}_h g \tag{22}$$

by an obvious calculation. Combining these formulas we get

$$\sum_{n\geq 0} \frac{1}{n!} h^n \mathbf{D}^n(fg) = \sum_{i\geq 0} \frac{1}{i!} h^i \mathbf{D}^i f. \sum_{j\geq 0} \frac{1}{j!} h^j \mathbf{D}^j g; \tag{23}$$

equating the terms containing the same power h^n of h, one gets

$$\mathbf{D}^n(fg) = \sum_{i+j=n} \frac{n!}{i!j!} \mathbf{D}^i f.\mathbf{D}^j g \tag{24}$$

that is, Leibniz's formula.

Another explanation starts from the case $n = 1$, that is

$$\mathbf{D}(fg) = \mathbf{D}f.g + f.\mathbf{D}g. \tag{25}$$

In a heuristic way it means that $\mathbf{D}$ applied to a product fg is the sum of two operators $\mathbf{D}_1$ acting on f only and $\mathbf{D}_2$ acting on g only. These actions being independent, $\mathbf{D}_1$ commutes to $\mathbf{D}_2$ hence the binomial formula

$$\mathbf{D}^n = (\mathbf{D}_1 + \mathbf{D}_2)^n = \sum_{i=0}^{n} \binom{n}{i} \mathbf{D}_1^i.\mathbf{D}_2^{n-i}. \tag{26}$$

By acting on the product fg and remarking that $\mathbf{D}_1^i.\mathbf{D}_2^j$ transforms fg into $\mathbf{D}^i f.\mathbf{D}^j g$, one recovers Leibniz's formula. In more detail, to calculate $\mathbf{D}^2(fg)$, one applies $\mathbf{D}$ to $\mathbf{D}(fg)$. Since $\mathbf{D}(fg)$ is the sum of two terms $\mathbf{D}f.g$ and $f.\mathbf{D}g$ apply $\mathbf{D}$ to $\mathbf{D}f.g$ to get $\mathbf{D}(\mathbf{D}f)g + \mathbf{D}f.\mathbf{D}g$ and to $f.\mathbf{D}g$ to get $\mathbf{D}f.\mathbf{D}g + f.\mathbf{D}(\mathbf{D}g)$, hence the sum

$$\begin{aligned}&\mathbf{D}(\mathbf{D}f).g + \mathbf{D}f.\mathbf{D}g + \mathbf{D}f.\mathbf{D}g + f.\mathbf{D}(\mathbf{D}g)\\ &\quad = \mathbf{D}^2 f.g + 2\mathbf{D}f.\mathbf{D}g + f.\mathbf{D}^2 g.\end{aligned}$$

This last proof can rightly be called "formal" since we act on the formulas, not on the objects: $\mathbf{D}_1$ transforms $f.g$ into $\mathbf{D}f.g$ but this doesn't mean that from the equality of functions $f_1.g_1 = f_2.g_2$ one gets $\mathbf{D}f_1.g_1 = \mathbf{D}f_2.g_2$ (counterexample: from $fg{=}gf$, we cannot infer $\mathbf{D}f.g = \mathbf{D}g.f$). The modern explanation is provided by the notion of tensor products: if V and W are two vector spaces (over the real numbers as coefficients, for instance), equal or distinct, there exists a new vector space $V \otimes W$ whose elements are formal

finite sums $\sum_i \lambda_i(v_i \otimes w_i)$ (with scalars λ_i and v_i in V, w_i in W); we take as basic rules the consequences of the fact that $v \otimes w$ is bilinear in v, w, but nothing more. Taking V and W to be the space $C^\infty(I)$ of the functions defined and indefinitely derivable in an interval I of $\mathbf{R}$, we define the operators $\mathbf{D}_1$ and $\mathbf{D}_2$ in $C^\infty(I) \otimes C^\infty(I)$ by

$$\mathbf{D}_1(f \otimes g) = \mathbf{D}f \otimes g, \quad \mathbf{D}_2(f \otimes g) = f \otimes \mathbf{D}g. \tag{27}$$

The two operators $\mathbf{D}_1\mathbf{D}_2$ and $\mathbf{D}_2\mathbf{D}_1$ transform $f \otimes g$ into $\mathbf{D}f \otimes \mathbf{D}g$, hence $\mathbf{D}_1$ and $\mathbf{D}_2$ commute. Define $\bar{\mathbf{D}}$ as $\mathbf{D}_1 + \mathbf{D}_2$ hence

$$\bar{\mathbf{D}}(f \otimes g) = \mathbf{D}f \otimes g + f \otimes \mathbf{D}g. \tag{28}$$

We can now calculate $\bar{\mathbf{D}}^n = (\mathbf{D}_1 + \mathbf{D}_2)^n$ by the binomial formula as in (25) with the conclusion

$$\bar{\mathbf{D}}^n(f \otimes g) = \sum_{i=0}^{n} \binom{n}{i} \mathbf{D}^i f \otimes \mathbf{D}^{n-i} g. \tag{29}$$

The last step is to go from (28) to (19). The rigorous reasoning is as follows. There is a linear operator μ taking $f \otimes g$ into $f.g$ and mapping $C^\infty(I) \otimes C^\infty(I)$ into $C^\infty(I)$; this follows from the fact that the product $f.g$ is bilinear in f and g. The formula (24) is expressed by $\mathbf{D} \circ \mu = \mu \circ \bar{\mathbf{D}}$ in operator terms, according to the diagram:

$$\begin{array}{ccc} C^\infty(I) \otimes C^\infty(I) & \xrightarrow{\mu} & C^\infty(I) \\ \bar{\mathbf{D}} \downarrow & & \downarrow \mathbf{D} \\ C^\infty(I) \otimes C^\infty(I) & \xrightarrow{\mu} & C^\infty(I). \end{array}$$

An easy induction entails $\mathbf{D}^n \circ \mu = \mu \circ \bar{\mathbf{D}}^n$, and from (28) one gets

$$\begin{aligned} &\mathbf{D}^n(fg) = \mathbf{D}^n(\mu(f \otimes g)) = \mu(\bar{\mathbf{D}}^n(f \otimes g)) \\ &= \mu(\sum_{i=0}^{n} \binom{n}{i} \mathbf{D}^i f \otimes \mathbf{D}^{n-i} g) = \sum_{i=0}^{n} \binom{n}{i} \mathbf{D}^i f.\mathbf{D}^{n-i} g. \end{aligned} \tag{30}$$

In words: first replace the ordinary product $f.g$ by the neutral tensor product $f \otimes g$, perform all calculations using the fact that $\mathbf{D}_1$ commutes to $\mathbf{D}_2$, then restore the product . in place of $\otimes$.

When the vector spaces V and W consist of functions of one variable, the tensor product $f \otimes g$ can be interpreted as the function $f(x)g(y)$ in two variables x, y; moreover $\mathbf{D}_1 = \partial/\partial x$, $\mathbf{D}_2 = \partial/\partial y$ and μ takes a function $F(x, y)$ of two variables into the one-variable function $F(x, x)$ hence $f(x)g(y)$ into $f(x)g(x)$ as it should. Formula (24) reads now

$$\frac{\partial}{\partial x}(f(x)g(x)) = (\frac{\partial}{\partial x} + \frac{\partial}{\partial y})f(x)g(y)|_{y=x}. \tag{31}$$

The previous "formal" proof goes over a familiar proof using Schwarz's theorem that $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ commute.

Starting from the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ of two vector spaces, one can iterate and obtain

$$\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3, \quad \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \otimes \mathcal{H}_4, \ldots.$$

Using once again the exponential notation, $\mathcal{H}^{\otimes n}$ is the tensor product of n copies of $\mathcal{H}$, with elements of the form $\sum \lambda.(\psi_1 \otimes ... \otimes \psi_n)$. In quantum physics, $\mathcal{H}$ represents the state vectors of a particle, and $\mathcal{H}^{\otimes n}$ represents the state vectors of a system of n independent particles of the same kind. If H is an operator in $\mathcal{H}$ representing for instance the energy of a particle, we define n operators H_i in $\mathcal{H}^{\otimes n}$ by

$$H_i(\psi_1 \otimes ... \otimes \psi_n) = \psi_1 \otimes \cdots \otimes H\psi_i \otimes \cdots \otimes \psi_n \tag{32}$$

(the energy of the i-th particle). Then $H_1, ..., H_n$ commute pairwise and $H_1 + \cdots + H_n$ is the total energy if there is no interaction. Usually, there is a pair interaction represented by an operator V in $\mathcal{H} \otimes \mathcal{H}$; then the total energy is given by $\sum_{i=1}^{n} H_i + \sum_{i<j} V_{ij}$ with

$$V_{12}(\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n) = V(\psi_1 \otimes \psi_2) \otimes \psi_3 \otimes \cdots \tag{33}$$

$$V_{23}(\psi_1 \otimes \cdots \otimes \psi_n) = \psi_1 \otimes V(\psi_2 \otimes \psi_3) \otimes \cdots \otimes \psi_n \tag{34}$$

etc... There are obvious commutation relations like

$$H_iH_j = H_jH_i$$
$$H_iV_{jk} = V_{jk}H_i \quad \text{if} \quad i, j, k \quad \text{are distinct.}$$

This is the so-called "locality principle": if two operators A and B refer to disjoint collections of particles (**a**) for A and (**b**) for B, they commute.

Faddeev and his collaborators made an extensive use of this notation in their study of quantum integrable systems. Also, Hirota introduced his so-called bilinear notation for differential operators connected with classical integrable systems (solitons).

2.4 Exponential vs. logarithm

In the case of real numbers, one usually starts from the logarithm and invert it to define the exponential (called antilogarithm not so long ago). Positive numbers have a logarithm; what about the logarithm of -1 for instance?

Things are worse in the complex domain. For a complex number z, define its exponential by the convergent series

$$\exp z = \sum_{n \geq 0} \frac{1}{n!} z^n. \tag{35}$$

From the binomial formula, using the commutativity $zz' = z'z$ one gets

$$\exp(z + z') = \exp z . \exp z' \tag{36}$$

as before. Separating real and imaginary part of the complex number $z = x + iy$ gives Euler's formula

$$\exp(x + iy) = e^x(\cos y + i \sin y) \tag{37}$$

subsuming trigonometry to complex analysis. The trigonometric lines are the "natural" ones, meaning that the angular unit is the radian (hence $\sin \delta \simeq \delta$ for small δ).

From an intuitive view of trigonometry, it is obvious that the points of a circle of equation $x^2 + y^2 = R^2$ can be uniquely parametrized in the form

$$x = R\ \cos\theta, \quad y = R\ \sin\theta \tag{38}$$

with $-\pi < \theta \leq \pi$, but the subtle point is to show that the geometric definition of $\sin\theta$ and $\cos\theta$ agree with the analytic one given by (36). Admitting this, every complex number $u \neq 0$ can be written as an exponential $\exp z_0$, where $z_0 = x_0 + iy_0$, x_0 real and y_0 in the interval $]-\pi, \pi]$. The number z_0 is called the principal determination of the logarithm of u, denoted by Ln u. But the general solution of the equation $\exp z = u$ is given by $z = z_0 + 2\pi i n$ where n is a rational integer. Hence a nonzero complex number has infinitely many logarithms. The functional property (35) of the exponential cannot be neatly inverted: for the logarithms we can only assert that $\mathrm{Ln}(u_1 \cdots u_p)$ and $\mathrm{Ln}(u_1) + \ldots + \mathrm{Ln}(u_p)$ differ by the addition of an integral multiple of $2\pi i$.

The exponential of a (real or complex) square matrix A is defined by the series

$$\exp A = \sum_{n \geq 0} \frac{1}{n!} A^n. \tag{39}$$

There are two classes of matrices for which the exponential is easy to compute:

a) Let A be diagonal $A = \mathrm{diag}(a_1, \ldots, a_n)$. Then $\exp A$ is diagonal with elements $\exp a_1, \ldots, \exp a_n$. Hence any **complex** diagonal matrix with non zero elements is an exponential, hence admits a logarithm, and even infinitely many ones.

b) Suppose that A is a special upper triangular matrix, with zeroes on the diagonal, of the type

$$A = \begin{pmatrix} 0 & a & b & c \\ & 0 & d & e \\ & & 0 & f \\ & & & 0 \end{pmatrix}.$$

Then $A^d = 0$ if A is of size $d \times d$. Hence $\exp A$ is equal to $I + B$ where B is of the form $A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \cdots + \frac{1}{(d-1)!}A^{d-1}$. Hence B is again a special upper triangular matrix and A can be recovered by the formula

$$A = B - \frac{B^2}{2} + \frac{B^3}{3} - \cdots + (-1)^d \frac{B^{d-1}}{d-1}. \tag{40}$$

This is just the truncated series for $\ln(I + B)$(notice $B^d = 0$). Hence in the case of these special triangular matrices, exponential and logarithm are inverse operations.

In general, A can be put in triangular form $A = UTU^{-1}$ where T is upper triangular. Let $\lambda_1, ..., \lambda_d$ be the diagonal elements of T, that is the eigenvalues of A. Then

$$\exp A = U.\exp T.U^{-1} \tag{41}$$

where $\exp T$ is triangular, with the diagonal elements $\exp \lambda_1, ... \exp \lambda_d$. Hence

$$\det(\exp A) = \prod_{i=1}^{d} \exp \lambda_i = \exp \sum_{i=1}^{d} \lambda_i = \exp(\mathrm{Tr}(A)). \tag{42}$$

The determinant of $\exp A$ is therefore non zero. Conversely **any complex matrix M with a nonzero determinant is an exponential**: for the proof, write M in the form $U.T.U^{-1}$ where T is composed of Jordan blocks of the form

$$T_s = \begin{pmatrix} \lambda & 1 & . & . & 0 \\ & . & . & . & \\ & 0 & & . & 1 \\ . & . & . & . & \lambda \end{pmatrix} \quad \text{with } \lambda \neq 0 \quad .$$

From the existence of the complex logarithm of λ and the study above of triangular matrices, it follows that T_s is an exponential, hence T and $M = UTU^{-1}$ are exponentials.

Let us add a few remarks:

a) A complex matrix with nonzero determinant has infinitely many logarithms; it is possible to normalize things to select one of them, but the conditions are rather artificial.

b) A real matrix with nonzero determinant is not always the exponential of a real matrix; for example, choose $M = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. This is not surprising since -1 has no real logarithm, but many complex logarithms of the form $k\pi i$ with k odd.

c) The noncommutativity of the multiplication of matrices implies that in general $\exp(A + B)$ is not equal to $\exp A.\exp B$. Here the logarithm of a product cannot be the sum of the logarithms, whatever normalization we choose.

2.5 Infinitesimals and exponentials

There are many notations in use for the higher order derivatives of a function f. Newton uses $\dot{f}, \ddot{f}, \ldots$, the customary notation is $f', f'', \ldots$. Once again, the exponential notation can be systematized, $f^{(m)}$ or $D^m f$ denoting the m-th derivative of f, for $m = 0, 1, \ldots$. This notation emphasizes that the derivation is a functional operator, hence

$$(f^{(m)})^{(n)} = f^{(m+n)}, \quad \text{or} \quad D^m(D^n f) = D^{m+n} f. \tag{43}$$

In this notation, it is cumbersome to write the chain rule for the derivative of a composite function

$$D(f \circ g) = (Df \circ g).Dg. \tag{44}$$

Leibniz's notation for the derivative is dy/dx if $y = f(x)$. Leibniz was never able to give a completely rigorous definition of the infinitesimals dx, dy, ...[1]. His explanation of the derivative is as follows: starting from x, increment it by an infinitely small amount dx; then $y = f(x)$ is incremented by dy, that is

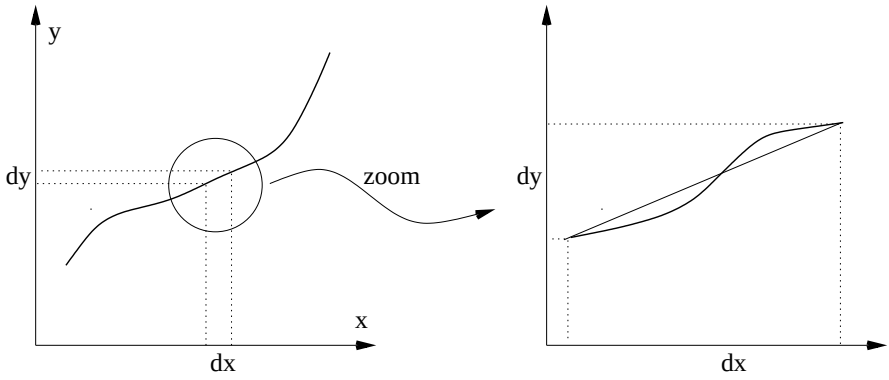

Fig. 1. Geometrical description: an infinitely small portion of the curve $y = f(x)$, after zooming, becomes infinitely close to a straight line, our function is "smooth", not fractal-like.

$$f(x + dx) = y + dy. \tag{45}$$

Then the derivative is $f'(x) = dy/dx$, hence according to (44)

$$f(x + dx) = f(x) + f'(x)dx. \tag{46}$$

[1] In modern times, Abraham Robinson has vindicated them using the tools of formal logic. There has been many interesting applications of his nonstandard analysis, but one has to admit that it remains too cumbersome to provide a viable alternative to the standard analysis. May be in the 21th century!

This cannot be literally true, otherwise the function $f(x)$ would be linear. The true formula is

$$f(x+dx) = f(x) + f'(x)dx + o(dx) \tag{47}$$

with an error term $o(dx)$ which is infinitesimal, of a higher order than dx, meaning $o(dx)/dx$ is again infinitesimal. In other words, the derivative $f'(x)$, independent of dx, is infinitely close to $\frac{f(x+dx)-f(x)}{dx}$ for all infinitesimals dx. The modern definition, as well as Newton's point of view of fluents, is a **dynamical** one: when dx goes to 0, $\frac{f(x+dx)-f(x)}{dx}$ tends to the limit $f('x)$. Leibniz's notion is **statical**: dx is a given, fixed quantity. But there is a hierarchy of infinitesimals: η is of higher order than ϵ if η/ϵ is again infinitesimal. In the formulas, equality is always to be interpreted up to an infinitesimal error of a certain order, not always made explicit.

We use these notions to describe the logarithm and the exponential. By definition, the derivative of $\ln x$ is $\frac{1}{x}$, hence

$$\frac{d\ln x}{dx} = \frac{1}{x}, \quad \text{that is } \ln(x+dx) = \ln(x) + \frac{dx}{x}.$$

Similarly for the exponential

$$\frac{d\exp x}{dx} = \exp x, \quad \text{that is } \exp(x+dx) = (\exp x)(1+dx).$$

This is a rule of compound interest. Imagine a fluctuating daily rate of interest, namely $\epsilon_1, \epsilon_2, ..., \epsilon_{365}$ for the days of a given year, every daily rate being of the order of 0.0003. For a fixed investment C, the daily reward is $C\epsilon_i$ for day i, hence the capital becomes $C + C\epsilon_1 + ... + C\epsilon_{365} = C.(1 + \sum_i \epsilon_i)$, that is approximately $C(1 + .11)$. If we reinvest every day our profit, invested capital changes according to the rule:

$$\begin{array}{ccccc} C_{i+1} = & C_i & + & C_i\epsilon_i = & C_i(1+\epsilon_i). \\ \uparrow & \uparrow & & \uparrow & \\ \text{capital} & \text{capital} & & \text{profit} & \\ \text{at day } i+1 & \text{at day } i & & \text{during day } i & \end{array}$$

At the end of the year, our capital is $C.\prod_i(1+\epsilon_i)$. We can now formulate the "bankers rule":

$$\text{if } S = \epsilon_1 + ... + \epsilon_N, \text{ then } \exp S = (1+\epsilon_1)\cdots(1+\epsilon_N). \quad (B)$$

Here N is infinitely large, and $\epsilon_1, \ldots, \epsilon_N$ are infinitely small; in our example, $S = 0.11$, hence $\exp S = 1 + S + \frac{1}{2}S^2 + \ldots$ is equal to $1.1163\ldots$: **by reinvesting daily, the yearly profit of** 11% **is increased to** 11.63%.

Formula (B) is not true without reservation. It certainly holds if all ϵ_i are of the same sign, or more generally if $\sum_i |\epsilon_i|$ is of the same order as $\sum \epsilon_i = x$.

For a counter-example, take $N = 2p^2$ with half of the ϵ_i being equal to $+\frac{1}{p}$, and the other half to $-\frac{1}{p}$ (hence $\sum_i \epsilon_i = 0$ while $\prod_i (1+\epsilon_i)$ is infinitely close to $1/e = \exp(-1)$).

To connect definition (B) of the exponential to the power series expansion $\exp S = 1 + S + \frac{1}{2!}S^2 + \cdots$ one can proceed as follows: by algebra we get

$$\prod_{i=1}^{N}(1+\epsilon_i) = \sum_{k=0}^{N} S_k, \tag{48}$$

where $S_0 = 1, S_1 = \epsilon_1 + ... + \epsilon_N = S$, and generally

$$S_k = \sum_{i_1<...<i_k} \epsilon_{i_1}...\epsilon_{i_k}. \tag{49}$$

We have to compare S_k to $\frac{1}{k!}S^k = \frac{1}{k!}(\epsilon_1 + \cdots + \epsilon_N)^k$. Developing the k-th power of S by the multinomial formula, we obtain S_k plus error terms each containing at least one of the $\epsilon_i's$ to a higher power $\epsilon_i^2, \epsilon_i^3, ...$ hence infinitesimal compared to the $\epsilon_i's$. The general **principle of compensation of errors**[2] is as follows: in an infinite sum of infinitesimals

$$\Sigma = \eta_1 + \cdots + \eta_M \tag{50}$$

subject each term to an error η_j becoming $\eta_j' = \eta_j + o(\eta_j)$ with an error $o(\eta_j)$ of higher order than η_j. Then Σ becomes

$$\Sigma' = \eta_1' + \cdots + \eta_M', \tag{51}$$

equal to Σ plus an error term $o(\eta_1) + \cdots + o(\eta_M)$. If the η_j are of the same sign, the error is $o(\Sigma)$, that is negligible compared to Σ.

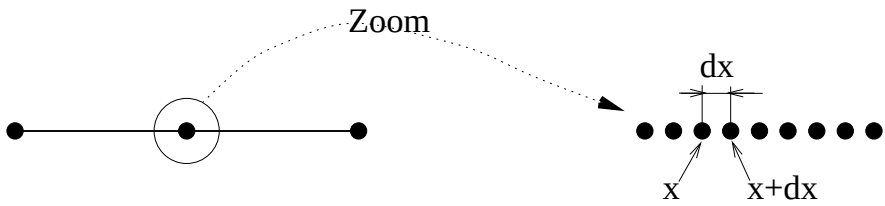

Fig. 2. Leibniz' continuum: by zooming, a finite segment of line is made of a large number of atoms of space: a fractal.

The implicit view of the continuum underlying Leibniz's calculus is as follows: a finite segment of a line is made of an infinitely large number of

[2] This terminology was coined by Lazare Carnot in 1797. Our formulation is more precise than his!

geometric atoms of space which can be arranged in a succession, each atom x being separated by dx from the next one. Hence in the definition of the logarithm

$$\ln\ a = \int_1^a \frac{dx}{x} \quad (\text{for } a > 1), \tag{52}$$

we really have $\sum_{1 \leq x \leq a} \frac{dx}{x}$. Similarly, the bankers rule (B) should be interpreted as

$$\exp\ a = \prod_{0 \leq x \leq a} (1 + dx) \ \ (\text{for } a > 0). \tag{53}$$

2.6 Differential equations

The previous formulation of the exponential suggests a method to solve a differential equation, for instance $y' = ry$. In differential form

$$dy = r(x)ydx, \tag{54}$$

that is

$$y + dy = (1 + r(x)dx)y. \tag{55}$$

The solution is

$$y(b) = \prod_{a \leq x \leq b} (1 + r(x)dx).y(a). \tag{56}$$

What is the meaning of this product? Putting $\epsilon(x) = r(x)dx$, an infinitesimal, and expanding the product as in (47), we get

$$\prod_x (1 + \epsilon(x)) = \sum_{k \geq 0} \sum_{a \leq x_1 < \ldots < x_k \leq b} \epsilon(x_1) \cdots \epsilon(x_k); \tag{57}$$

reinterpreting the multiple sum as a multiple integral, this is

$$\sum_{k \geq 0} \int \cdots \int_{\Delta_k} r(x_1) \cdots r(x_k) dx_1 \cdots dx_k. \tag{58}$$

The domain of integration Δ_k is given by the inequalities

$$a \leq x_1 \leq x_2 \leq \ldots \leq x_k \leq b. \tag{59}$$

The classical solution to the differential equation $y' = ry$ is given by

$$y(b) = (\exp \int_a^b r(x)dx).y(a). \tag{60}$$

Let us see how to go from (57) to (59). Geometrically, consider the hypercube C_k given by

$$a \leq x_1 \leq b, \cdots, a \leq x_k \leq b \tag{61}$$

in the euclidean space $\mathbf{R}^k$ of dimension k with coordinates $x_1, \ldots, x_k$. The group S_k of the permutations σ of $\{1, \ldots, k\}$ acts on $\mathbf{R}^k$, by transforming the vector $\mathbf{x}$ with coordinates $x_1, \ldots, x_k$ into the vector $\sigma.\mathbf{x}$ with coordinates $x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(k)}$. Then the cube C_k is the union of the $k!$ transforms $\sigma(\Delta_k)$. Since the function $r(x_1) \ldots r(x_k)$ to be integrated is symmetrical in the variables $x_1, \ldots, x_k$ and moreover two distinct domains $\sigma(\Delta_k)$ and $\sigma'(\Delta_k)$ overlap by a subset of dimension $< k$, hence of volume 0, we see that the integral of $r(x_1) \cdots r(x_k)$ over C_k is $k!$ times the integral over Δ_k. That is

$$\int \cdots \int_{\Delta_k} r(x_1) \cdots r(x_k) dx_1 \cdots dx_k =$$
$$\frac{1}{k!} \int_a^b dx_1 \cdots \int_a^b dx_k \; r(x_1) \cdots r(x_k) = \frac{1}{k!} (\int_a^b r(x) dx)^k.$$

Summing over k, and using the definition of an exponential by a series, we conclude

$$\sum_{k \geq 0} \int \cdots \int_{\Delta_k} r(x_1) ... r(x_k) dx_1 ... dx_k = \exp \int_a^b r(x) dx. \tag{62}$$

as promised.

The same method applies to the linear systems of differential equations. We cast them in the matrix form

$$y' = A.y, \tag{63}$$

that is the differential form

$$dy = A(x) y dx. \tag{64}$$

Here $A(x)$ is a matrix depending on the variable x, and $y(x)$ is a vector (or matrix) function of x. From (64) we get

$$y(x + dx) = (I + A(x)dx) y(x). \tag{65}$$

Formally the solution is given by

$$y(b) = \prod_{a \leq x \leq b} (I + A(x)dx).y(a). \tag{66}$$

We have to take into account the **noncommutativity** of the products $A(x)A(y)A(z) \ldots$. Explicitly, if we have chosen intermediate points

$$a = x_0 < x_1 < \ldots < x_N = b,$$

with infinitely small spacing

$$dx_1 = x_1 - x_0, dx_2 = x_2 - x_1, ..., dx_N = x_N - x_{N-1},$$

the product in (67) is

$$(I + A(x_N)dx_N)(I + A(x_{N-1})dx_{N-1}) \cdots (I + A(x_1)dx_1).$$

We use the notation $\overleftarrow{\prod}_{1 \leq i \leq N} U_i$ for a **reverse product** $U_N U_{N-1} \cdots U_1$; hence the previous product can be written as $\overleftarrow{\prod}_{1 \leq i \leq N}(I + A(x_i)dx_i)$ and we should replace $\prod$ by $\overleftarrow{\prod}$ in equation (67). The noncommutative version of equation (47) is

$$\overleftarrow{\prod}_{1 \leq i \leq N}(I + A_i) = \sum_{k=0}^{N} \sum_{i_1 > \cdots > i_k} A_{i_1} \cdots A_{i_k}. \tag{67}$$

Let us define the resolvant (or propagator) as the matrix

$$U(b, a) = \overleftarrow{\prod}_{a \leq x \leq b}(I + A(x)dx). \tag{68}$$

Hence the differential equation $dy = A(x)ydx$ is solved by $y(b) = U(b, a)y(a)$ and from (68) we get

$$U(b, a) = \sum_{k \geq 0} \int \cdots \int_{\Delta_k} A(x_k) \cdots A(x_1)dx_1 \cdots dx_k \tag{69}$$

with the factors $A(x_i)$ in **reverse order**

$$A(x_k) \cdots A(x_1) \text{ for } x_1 < \ldots < x_k. \tag{70}$$

One owes to R. Feynman and F. Dyson (1949) the following notational trick. If we have a product of factors $U_1, \cdots, U_N$, each attached to a point x_i on a line, we denote by $T(U_1 \cdots U_N)$ (or more precisely by $\overleftarrow{T}(U_1 \cdots U_N)$) the product $U_{i_1} \cdots U_{i_N}$ where the permutation $i_1 \ldots i_N$ of $1 \ldots N$ is such that $x_{i_1} > \cdots > x_{i_N}$. Hence in the rearranged product the abscisses attached to the factors increase from right to left. We argue now as in the proof of (62) and conclude that

$$\int \cdots \int_{\Delta_k} A(x_k) \cdots A(x_1)dx_1 \cdots dx_k$$
$$= \frac{1}{k!} \int_a^b dx_1 \cdots \int_a^b dx_k T(A(x_1) \cdots A(x_k)). \tag{71}$$

We can rewrite the propagator as

$$U(b, a) = T \exp \int_a^b A(x)dx, \tag{72}$$

with the following interpretation:

a) First use the series $\exp S = \sum_{k\geq 0} \frac{1}{k!} S^k$ to expand $\exp \int_a^b A(x)dx$.

b) Expand $S^k = (\int_a^b A(x)dx)^k$ as a multiple integral

$$\int_a^b dx_1 \cdots \int_a^b dx_k \ A(x_1) \cdots A(x_k).$$

c) Treat T as a linear operator commuting with series and integrals, hence

$$T \exp \ S = \sum_{k\geq 0} \frac{1}{k!} T(S^k) = \sum_{k\geq 0} \frac{1}{k!} T\{\int_a^b dx_1 \cdots \int_a^b dx_k \ A(x_1) \cdots A(x_k)\}$$
$$= \sum_{k\geq 0} \frac{1}{k!} \int_a^b dx_1 \cdots \int_a^b dx_k \ T(A(x_1) \cdots A(x_k)).$$

We give a few properties of the T (or time ordered) exponential:

a) Parallel to the rule

$$\int_a^c A(x)dx = \int_a^b A(x)dx + \int_b^c A(x)dx \quad (\text{for } a < b < c) \tag{73}$$

we get

$$T \exp \int_a^c A(x)dx = T \exp \int_b^c A(x)dx . T \exp \int_a^b A(x)dx. \tag{74}$$

Notice that, in (73), the two matrices

$$L = \int_a^b A(x)dx, \quad M = \int_b^c A(x)dx$$

don't commute, hence $\exp(L + M)$ is in general different from $\exp L . \exp M$. Hence formula (74) is not in general valid for the ordinary exponential.

b) The next formula embodies the classical method of "variation of constants" and is known in the modern litterature as a "gauge transformation". It reads as

$$S(b).T \exp \int_a^b A(x)dx.S(a)^{-1} = T \exp \int_a^b B(x)dx \tag{75}$$

with

$$B(x) = S(x)A(x)S(x)^{-1} + S'(x)S(x)^{-1}, \tag{76}$$

where $S(x)$ is an invertible matrix depending on the variable x. The general formula (75) can be obtained by "taking a continuous reverse product" $\overleftarrow{\prod}_{a\leq x\leq b}$ over the infinitesimal form

$$S(x + dx)(I + A(x)dx))S(x)^{-1} = I + B(x)dx \tag{77}$$

(for the proof, write $S(x+dx) = S(x) + S'(x)dx$ and neglect the terms proportional to $(dx)^2$). We leave it as an exercise to the reader to prove (75) from the expansion (70) for the propagator.

c) There exists a complicated formula for the T-exponential $T\exp\int_a^b A(x)\,dx$ when $A(x)$ is of the form $\frac{A_1(x)+A_2(x)}{2}$. Neglecting terms of order $(dx)^2$, we get

$$I + A(x)dx = (I + A_2(x)\frac{dx}{2})(I + A_1(x)\frac{dx}{2}) \tag{78}$$

and we can then perform the product $\overleftarrow{\prod}_{a\leq x\leq b}$. This formula is the foundation of the **multistep method** in numerical analysis: starting from the value $y(x)$ at time x of the solution to the equation $y' = Ay$, we split the infinitesimal interval $[x, x+dx]$ into two parts

$$I_1 = [x, x+\frac{dx}{2}], \quad I_2 = [x+\frac{dx}{2}, x+dx];$$

we move at speed $A_1(x)y(x)$ during I_1 and then at speed $A_2(x)y(x+\frac{dx}{2})$ during I_2. Let us just mention one corollary of this method, the so-called **Trotter-Kato-Nelson** formula:

$$\exp(L+M) = \lim_{n\to\infty}(\exp(L/n)\exp(M/n))^n. \tag{79}$$

d) If the matrices $A(x)$ pairwise commute, the T-exponential of $\int_a^b A(x)dx$ is equal to the ordinary exponential. In the general case, the following formula holds

$$T\exp\int_a^b A(x)dx = \exp V(b,a) \tag{80}$$

where $V(b,a)$ is explictly calculated using integration and iterated Lie brackets. Here are the first terms

$$V(b,a) = \int_a^b A(x)dx + \frac{1}{2}\int\int_{\Delta_2}[A(x_2), A(x_1)]dx_1dx_2 \tag{81}$$

$$+\frac{1}{3}\int\int\int_{\Delta_3}[A(x_3),[A(x_2),A(x_1)]]dx_1dx_2dx_3 \tag{82}$$

$$-\frac{1}{6}\int\int\int_{\Delta_3}[A(x_2),[A(x_3),A(x_1)]]dx_1dx_2dx_3 + \cdots \quad .$$

The higher-order terms involve integrals of order $k \geq 4$. As far as I can ascertain, this formula was first enunciated by K. Friedrichs around 1950 in his work on the foundations of Quantum Field Theory. A corollary is the **Campbell-Hausdorff formula**:

$$\exp L.\exp M =$$

$$\exp(L+M+\frac{1}{2}[L,M]+\frac{1}{12}[L,[L,M]]+\frac{1}{12}[M,[M,L]]+\cdots). \tag{83}$$

It can be derived from (80) by putting $a = 0$, $b = 2$, $A(x) = M$ for $0 \leq x \leq 1$ and $A(x) = L$ for $1 \leq x \leq 2$.

The T-exponential found lately numerous geometrical applications. If C is a curve in a space of arbitrary dimension, the line integral $\int_C A_\mu(x)dx^\mu$ is well-defined and the corresponding T-exponential

$$T \exp \int_C A_\mu(x)dx^\mu \tag{84}$$

is closely related to the parallel transport along the curve C.

3 Operational calculus

3.1 An algebraic digression: umbral calculus

We first consider the classical **Bernoulli numbers**. I claim that they are defined by the equation

$$(B+1)^n = B^n \quad \text{for } n \geq 2, \tag{1}$$

together with the initial condition $B^0 = 1$. The meaning is the following: expand $(B+1)^n$ by the binomial theorem, then replace the power B^k by B_k. Hence $(B+1)^2 = B^2$ gives $B^2 + 2B^1 + B^0 = B^2$, that is after lowering the indices $B_2 + 2B_1 + B_0 = B_2$, that is $2B_1 + B_0 = 0$. Treating $(B+1)^3 = B^3$ in a similar fashion gives $3B_2 + 3B_1 + B_0 = 0$. We write the first equations of this kind

$$\begin{aligned}
n &= 2 \quad 2B_1 + B_0 = 0 \\
n &= 3 \quad 3B_2 + 3B_1 + B_0 = 0 \\
n &= 4 \quad 4B_3 + 6B_2 + 4B_1 + B_0 = 0 \\
n &= 5 \quad 5B_4 + 10B_3 + 10B_2 + 5B_1 + B_0 = 0.
\end{aligned}$$

Starting from $B_0 = 1$ we get successively

$$B_1 = -\frac{1}{2},\ B_2 = \frac{1}{6},\ B_3 = 0,\ B_4 = -\frac{1}{30}, \ldots$$

Using the same kind of formalism, define the Bernoulli polynomials by

$$B_n(X) = (B+X)^n. \tag{2}$$

According to the previous rule, we first expand $(B+X)^n$ using the binomial theorem, then replace B^k by B_k. Hence we get explicitly

$$B_n(X) = \sum_{k=0}^{n} \binom{n}{k} B_{n-k} X^k. \tag{3}$$

Since $\frac{d}{dX}(X+c)^n = n(X+c)^{n-1}$ for any c independent of X, we expect

$$\frac{d}{dX}B_n(X) = nB_{n-1}(X). \tag{4}$$

This is easy to check on the explicit definition (3). Here is a similar calculation

$$(B+(X+Y))^n = ((B+X)+Y)^n = \sum_{k=0}^{n} \binom{n}{k} (B+X)^{n-k}Y^k,$$

from which we expect to find

$$B_n(X+Y) = \sum_{k=0}^{n} \binom{n}{k} B_{n-k}(X)Y^k. \tag{5}$$

Indeed from (4) we get

$$(\frac{d}{dX})^k B_n(X) = \frac{n!}{(n-k)!}B_{n-k}(X) \tag{6}$$

by induction on k, hence (5) follows from Taylor's formula $B_n(X+Y) = \sum_{k\geq 0} \frac{1}{k!}(\frac{d}{dX})^k B_n(X)Y^k$.

We deduce now a generating series for the Bernoulli numbers. Formally

$$\begin{aligned}(e^S-1)e^{BS} &= e^S e^{BS} - e^{BS} = e^{(B+1)S} - e^{BS} \\ &= \sum_{n\geq 0} \frac{1}{n!} S^n((B+1)^n - B^n) = S((B+1)^1 - B^1) = S.\end{aligned}$$

Since $e^{BS} = \sum_{n\geq 0} \frac{1}{n!}B^n S^n$, we expect

$$\sum_{n\geq 0} B_n S^n/n! = \frac{S}{e^S-1}. \tag{7}$$

Again this can be checked rigorously.

What is the secret behind these calculations?

We consider functions $F(B, X, \dots)$ depending on a variable B and other variables $X, \dots$. Assume that $F(B, X, \dots)$ can be expanded as a polynomial or power series in B, namely

$$F(B, X, \dots) = \sum_{n\geq 0} B^n F_n(X, \dots). \tag{8}$$

Then the "mean value" with respect to B is defined by

$$< F(B, X, ...) >= \sum_{n\geq 0} B_n F_n(X, ...), \tag{9}$$

where the B_n's are the Bernoulli numbers: this corresponds to the rule "lower the index in B^n". If $F(B, X, \dots)$ can be written as a series $\sum_i F_i(B, X, \dots)G_i(X, \dots)$ where the G_i's are independent of B, then obviously [3]

$$< F(B, X, \dots) >= \sum_i < F_i(B, X, \dots) > G_i(X, \dots). \tag{10}$$

All formal calculations are justified by this simple rule **which affords also a probabilistic interpretation** (see section 3.7).

3.2 Binomial sequences of polynomials

These are sequences of polynomials $U_0(X), U_1(X), ...$ in one variable X satisfying the following relations:

a) $U_0(X)$ is a constant;

b) for any $n \geq 1$, one gets

$$\frac{d}{dX}U_n(X) = nU_{n-1}(X). \tag{11}$$

By induction on n it follows that $U_n(X)$ is of degree $\leq n$. The binomial sequence is **normalized** if furthermore $U_0(X) = 1$, in which case every $U_n(X)$ is a monic polynomial of degree n, that is

$$U_n(X) = X^n + c_1X^{n-1} + \dots + c_n.$$

Applying Taylor's formula as above (derivation of formula (5)), one gets

$$U_n(X + Y) = \sum_{k=0}^{n} \binom{n}{k} U_{n-k}(X)Y^k. \tag{12}$$

We introduce now a numerical sequence by $u_n = U_n(0)$ for $n \geq 0$. Putting $X = 0$ in (12) and reverting from Y to X as a variable, we get

$$U_n(X) = \sum_{k=0}^{n} \binom{n}{k} u_{n-k}X^k. \tag{13}$$

Conversely, given any numerical sequence $u_0, u_1, ...$ and defining the polynomials $U_n(X)$ by (13), one derives immediately the relations

$$\frac{d}{dX}U_n(X) = nU_{n-1}(X), \quad U_n(0) = u_n. \tag{14}$$

[3] Sofar we considered only identities linear in the B_n's. If we want to treat nonlinear terms, like products $B_m.B_n$, we need to introduce two independent symbols B and B' and use the umbral rule to replace $B^m B'^n$ by $B_m B_n$. In probabilistic terms (see section 3.7), we introduce two independent random variables and take the mean value w.r.t. both simultaneously.

The exponential generating series for the constants u_n is given by

$$u(S) = \sum_{n\geq 0} u_n S^n/n!. \tag{15}$$

From (13), one obtains the exponential generating series

$$U(X,S) = \sum_{n\geq 0} U_n(X) S^n/n!$$

for the polynomials $U_n(X)$, namely in the form

$$U(X,S) = u(S)e^{XS}. \tag{16}$$

This could be expected. Writing $\partial_X,\ \partial_S \ldots$ for the partial derivatives, the basic relation $\partial_X U_n = nU_{n-1}$ translates as $(\partial_X - S)U(X,S) = 0$ or equivalently as

$$\partial_X(e^{-XS}U(X,S)) = 0. \tag{17}$$

Hence $e^{-XS}U(X,S)$ depends only on S, and putting $X = 0$ we obtain the value $U(0,S) = u(S)$.

The umbral calculus can be successfully applied to our case. Hence $U_n(X)$ can be interpreted as $\langle (X+U)^n \rangle$ provided $\langle U^n \rangle = u_n$. Similarly $u(S)$ is equal to $\langle e^{US} \rangle$ and $U(X,S)$ to $\langle e^{(X+U)S} \rangle$. The symbolic derivation of (16) is as follows

$$U(X,S) = \langle e^{(X+U)S} \rangle = \langle e^{XS}.e^{US} \rangle = e^{XS}\langle e^{US} \rangle = e^{XS}u(S).$$

We describe in more detail the three basic binomial sequences of polynomials:

a) The sequence $I_n(X) = X^n$ satisfies obviously (11). In this (rather trivial) case, we get

$$i_0 = 1,\ i_1 = i_2 = \ldots = 0,\ I(S) = 1,\ I(X,S) = e^{XS}.$$

b) **The Bernoulli polynomials** obey the rule (11)(see formula (4)). I claim that they are characterized by the further property

$$\int_0^1 B_n(x)dx = 0 \quad \text{for } n \geq 1. \tag{18}$$

Indeed, introducing the exponential generating series

$$B(X,S) = \sum_{n\geq 0} B_n(X) S^n/n!, \tag{19}$$

the requirement (18) is equivalent to the integral formula

$$\int_0^1 B(x,S)dx = 1. \tag{20}$$

According to the general theory of binomial sequences, $B(X,S)$ is of the form $b(S)e^{XS}$, hence

$$\int_0^1 B(x,S)dx = \int_0^1 b(S)e^{xS}dx = b(S)(\frac{e^S-1}{S}).$$

Solving (20) we get $b(S) = S/(e^S-1)$ and from (7) this is the exponential generating series for the Bernoulli numbers. The exponential generating series for the Bernoulli polynomials is therefore

$$B(X,S) = \frac{Se^{XS}}{e^S-1}. \tag{21}$$

Here is a short table:

$$B_0(X) = 1$$
$$B_1(X) = X - \frac{1}{2}$$
$$B_2(X) = X^2 - X + \frac{1}{6}$$
$$B_3(X) = X^3 - \frac{3}{2}X^2 + \frac{1}{2}X.$$

c) We come to the **Hermite polynomials** which form the normalized binomial sequence of polynomials characterized by

$$\int_{-\infty}^{+\infty} H_n(x)d\gamma(x) = 0 \quad \text{for } n \geq 1, \tag{22}$$

where $d_\gamma(x)$ denotes the normal probability law, that is

$$d\gamma(x) = (2\pi)^{-1/2}e^{-x^2/2}dx. \tag{23}$$

We follows the same procedure as for the Bernoulli polynomials. Hence for the exponential generating series

$$H(X,S) = \sum_{n\geq 0} H_n(X)S^n/n! = h(S)e^{XS} \tag{24}$$

we get

$$\int_{-\infty}^{+\infty} H(x,S)d\gamma(x) = 1, \tag{25}$$

that is

$$1/h(S) = \int_{-\infty}^{+\infty} e^{xS} d\gamma(x). \tag{26}$$

The last integral being easily evaluated, we conclude

$$h(S) = e^{-S^2/2}. \tag{27}$$

From this relation, we can evaluate $H(X, S)$ namely

$$H(X, S) = e^{XS - S^2/2} = e^{X^2/2} e^{-(X-S)^2/2} \tag{28}$$

and using Taylor's expansion for $e^{-(X-S)^2/2}$, we get

$$H_n(X) = (-1)^n e^{X^2/2} (\frac{d}{dX})^n e^{-X^2/2}. \tag{29}$$

In the spirit of operator calculus, use the identity

$$e^{X^2/2} \frac{d}{dX} e^{-X^2/2} = \frac{d}{dX} - X, \tag{30}$$

hence

$$H_n(X) = (X - \frac{d}{dX})^n .1. \tag{31}$$

This is tantamount to a recursion formula

$$H_{n+1}(X) = XH_n(X) - \frac{d}{dX} H_n(X) = XH_n(X) - nH_{n-1}(X). \tag{32}$$

The following table is then easily derived:

$$H_0(X) = 1$$
$$H_1(X) = X$$
$$H_2(X) = X^2 - 1$$
$$H_3(X) = X^3 - 3X$$
$$H_4(X) = X^4 - 6X^2 + 3.$$

3.3 Transformation of polynomials

We use the standard notation $\mathbf{C}[X]$ to denote the vector space of polynomials in the variable X with complex coefficients. Since the monomials X^n form a basis of $\mathbf{C}[X]$, a linear operator $\mathbf{U} : \mathbf{C}[X] \to \mathbf{C}[X]$ is completely determined

by the sequence of polynomials $U_n(X)$ defined as the image $\mathbf{U}[X^n]$ of X^n under $\mathbf{U}$. Here are a few examples:

$\mathbf{I}$	identity operator	$I_n(X) = X^n$
$\mathbf{D}$	derivation $\dfrac{d}{dX}$	$D_n(X) = nX^{n-1}$
$\mathbf{T}_c$	translation operator	$T_{c,n}(X) = (X+c)^n$.

Notice that in general $\mathbf{T}_c$ transforms a polynomial $P(X)$ into $P(X+c)$ and Taylor's formula amounts to

$$\mathbf{T}_c = e^{c\mathbf{D}}; \tag{33}$$

furthermore $\mathbf{T}_0 = \mathbf{I}$. From the definition of the derivative, one gets

$$\mathbf{D} = \lim_{c\to 0}(\mathbf{T}_c - \mathbf{I})/c. \tag{34}$$

We can reformulate the properties of binomial sequences:
- the definition $\mathbf{D}U_n(X) = nU_{n-1}(X)$ amounts to $\mathbf{UD} = \mathbf{DU}$;
- the exponential generating series $U(X,S)$ is nothing else than $\mathbf{U}[e^{XS}]$;
- formula (12), after substituting c to Y reads as

$$U_n(X+c) = \sum_{k=0}^{n} \binom{n}{k} U_{n-k}(X)c^k$$

that is

$$\mathbf{T}_c\mathbf{U}[X^n] = \sum_{k=0}^{n} \binom{n}{k} \mathbf{U}[X^{n-k}]c^k = \mathbf{U}[(X+c)^n] = \mathbf{U}\mathbf{T}_c[X^n].$$

Hence this formula expresses that $\mathbf{U}$ commutes to $\mathbf{T}_c$

$$\mathbf{T}_c\mathbf{U} = \mathbf{U}\mathbf{T}_c; \tag{35}$$

- formula (13) can be rewritten as

$$\mathbf{U}[X^n] = \sum_{k\geq 0} \frac{1}{k!} u_k \mathbf{D}^k[X^n]. \tag{36}$$

From the definition (15) of the exponential generating series, we obtain

$$\mathbf{U} = u(\mathbf{D}). \tag{37}$$

To sum up, our operators are characterized by the following equivalent properties:

a) $\mathbf{U}$ commutes to the derivative $\mathbf{D}$;

b) $\mathbf{U}$ commutes to the translation operators $\mathbf{T}_c$;

c) $\mathbf{U}$ can be expressed as a power series $u(\mathbf{D})$ in $\mathbf{D}$.

Furthermore, since $\mathbf{D}$ acts on e^{XS} by multiplication by S, then $\mathbf{U} = u(\mathbf{D})$ multiplies e^{XS} by $u(S)$, hence we recover formula (16).

3.4 Expansion formulas

As we saw before, $B_n(X)$ and $H_n(X)$ are monic polynomials and therefore the sequences $(B_n(X))_{n\geq 0}$ and $(H_n(X))_{n\geq 0}$ are two basis of the vector space $\mathbf{C}[X]$. Hence an arbitrary polynomial $P(X)$ can be expanded as a linear combination of the Bernoulli polynomials, as well as of the Hermite polynomials. Our aim is to give explicit formulas.

Consider a general binomial sequence $(U_n(X))_{n\geq 0}$ such that $u_0 \neq 0$, with exponential generating series $U(X,S) = u(S)e^{XS}$. Introduce the inverse series $v(S) = 1/u(S)$; explicitly

$$v(S) = \sum_{n\geq 0} v_n S^n/n!$$

and the coefficients v_n are defined inductively by

$$v_0 = 1/u_0,\ v_n = -\frac{1}{u_0}\sum_{k=1}^{n}\binom{n}{k} u_k v_{n-k}. \tag{38}$$

In the spirit of umbral calculus, let us define the linear form ϕ_0 on $\mathbf{C}[X]$ by $\phi_0[X^n] = v_n$. I claim that **the development of an arbitrary polynomial in terms of the U_n's is given by**

$$P(X) = \sum_{n\geq 0}\frac{1}{n!}\phi_0[\mathbf{D}^n P].U_n(X). \tag{39}$$

Before giving a proof, let us examine the three basic examples:

a) If $U_n(X) = X^n$, then $u(S) = 1$, hence $v(S) = 1$. That is $v_0 = 1$ and $v_n = 0$ for $n \geq 1$. The linear form ϕ_0 is given by $\phi_0[P] = P(0)$ and formula (39) reduces to MacLaurin's expansion

$$P(X) = \sum_{n\geq 0}\frac{1}{n!}\mathbf{D}^n P(0).X^n. \tag{40}$$

b) For the Bernoulli polynomials we know that $1/b(S)$ is equal to $(e^S - 1)/S$, hence $v_n = \frac{1}{n+1}$. The linear form ϕ_0 is defined by $\phi_0[X^n] = \frac{1}{n+1}$, that is

$$\phi_0[P] = \int_0^1 P(x)dx. \tag{41}$$

Hence

$$P(X) = \sum_{n\geq 0}\frac{1}{n!}\int_0^1 \mathbf{D}^n P(x)dx.B_n(X). \tag{42}$$

c) In the case of Hermite polynomials, we know that $1/h(S)$ is equal to $e^{S^2/2}$, hence

$$v_{2m} = \frac{(2m)!}{m!2^m}, \quad v_{2m+1} = 0. \tag{43}$$

According to (26), we get $v_n = \int_{-\infty}^{+\infty} x^n d\gamma(x)$, hence

$$\phi_0[P] = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} P(x) e^{-x^2/2} dx. \tag{44}$$

In these three cases, the formula for $\phi_0[P]$ takes a similar form, namely

$$\phi_0[P] = \int_a^b P(x) w(x) dx \tag{45}$$

with the following prescriptions:
$a = -\infty,\ b = +\infty,\ \ w(x) = \delta(x)$ in case a),
$a = 0,\ b = 1,\ \ w(x) = 1$ in case b),
$a = -\infty,\ b = +\infty,\ \ w(x) = (2\pi)^{-1/2} e^{-x^2/2}$ in case c).
The normalization $\phi_0[1] = 1$ amounts to $\int_a^b w(x)dx = 1$, that is $w(x)$ is the probability density of a random variable taking values in the interval $[a, b]$(see section 3.7).

There is a peculiarity in case c).
Namely, according to the general formula (39), an arbitrary polynomial $P(X)$ can be expanded in a series $\sum_{n\geq 0} c_n H_n(X)$ of Hermite polynomials where c_n is equal to $\frac{1}{n!} \int_{-\infty}^{+\infty} D^n P(x)\ d\gamma(x)$. Integrating by parts and taking into account the definition (29) of $H_n(X)$ we obtain

$$c_n = \int_{-\infty}^{+\infty} P(x) H_n(x) d\gamma(x). \tag{46}$$

This amounts to the **orthogonality relation**

$$\int_{-\infty}^{+\infty} H_m(x) H_n(x) d\gamma(x) = \delta_{mn} n! \tag{47}$$

for the Hermite polynomials. There is no such orthogonality relation for the Bernoulli polynomials.

One final word about the proof of (39). By linearity, it suffices to consider the case $P = U_m$, that is to prove the **biorthogonality relation**

$$\phi_0[\mathbf{D}^n U_m] = n! \delta_{mn}. \tag{48}$$

We first calculate $\phi_0[U_m]$. From formula (13), we obtain

$$\phi_0[U_m] = \sum_{k=0}^{m} \binom{m}{k} u_{m-k} \phi_0[X^k] = \sum_{k=0}^{m} \binom{m}{k} u_{m-k} v_k$$

and from (38), $\phi_0[U_m]$ is 0 for $m \geq 1$. Since $\mathbf{D}^n U_m$ is proportional to U_{m-n} according to the basic formula $\mathbf{D}U_m = mU_{m-1}$, one gets $\phi_0[\mathbf{D}^n U_m] = 0$ for $m \neq n$. Finally $\mathbf{D}^m U_m = m!$, hence $\phi_0[\mathbf{D}^m U_m] = m!$.

3.5 Signal transforms

A transmission device transforms a suitable **input** f into an **output** F. Both are evolving in time and are represented by functions of time $f(t)$ and $F(t)$[4]. We assume the device to be **linear** and in a **stationary** regime, that is there is a linear operator $\mathbf{V}$ taking $f(t)$ into $F(t)$ (linearity) and $f(t+\tau)$ into $F(t+\tau)$ for any fixed τ (stationarity).

Here are the main types of response:

Input	*Output*
$\delta(t)$	$I(t)$
e^{pt}	$\Theta(t)e^{pt}$
t^n	$V_n(t)$

In the first case, $\delta(t)$ is a Dirac singular function, that is a **pulse**, and $I(t)$ is the **impulse response**. By stationarity V transforms $\delta(t-\tau)$ into $I(t-\tau)$; an arbitrary input f can be represented as a superposition of pulses

$$f(t) = \int_{-\infty}^{+\infty} f(\tau)\delta(t-\tau)d\tau \tag{49}$$

hence by linearity the output

$$F(t) = \int_{-\infty}^{+\infty} f(\tau)I(t-\tau)d\tau = \int_{-\infty}^{+\infty} f(t-\tau)I(\tau)d\tau. \tag{50}$$

In the **non-anticipating case**, $I(t)$ is zero before the pulse $\delta(t)$ occurs, that is $I(t) = 0$ for $t < 0$. In this case the output is given by

$$F(t) = \int_{-\infty}^{t} f(\tau)I(t-\tau)d\tau. \tag{51}$$

In the case of the exponential input $f(t) = e^{pt}$, the output is equal to $\int_{-\infty}^{+\infty} e^{p\tau}I(t-\tau)d\tau = \int_{-\infty}^{+\infty} e^{p(t-\tau)}I(\tau)d\tau$ according to (50), that is to $\Theta(p)e^{pt}$ with the **spectral gain**

$$\Theta(p) = \int_{-\infty}^{+\infty} e^{-p\tau}I(\tau)d\tau. \tag{52}$$

[4] For simplicity, we restrict to the case where input and output are scalars and not vectors.

We can give an **a priori** argument: $f_p(t) = e^{pt}$ is a solution of the differential equation $\mathbf{D}f_p = pf_p$; since the operator $\mathbf{V}$ is stationary, that is commutes to the translation operator $\mathbf{T}_c$, it commutes to $\mathbf{D} = \lim_{c\to 0}(\mathbf{T}_c - \mathbf{I})/c$. Hence the output F_p corresponding to the input f_p is a solution of the differential equation $\mathbf{D}F_p = pF_p$, hence is proportional to e^{pt}.

In a similar way, the monomials t^n satisfy the cascade of differential equations

$$\mathbf{D}[t] = 1, \quad \mathbf{D}[t^2] = 2t, \quad \mathbf{D}[t^3] = 3t^2, \ldots$$

Since $\mathbf{V}$ commutes to $\mathbf{D}$ and the constants are the solutions of the differential equation $\mathbf{D}(f) = 0$, it follows that the images $V_n(t) = \mathbf{V}[t^n]$ form a binomial sequence of polynomials. Explicitly

$$V_n(t) = \int_{-\infty}^{+\infty} (t-\tau)^n I(\tau) d\tau = \sum_{k=0}^{n} \binom{n}{k} v_{n-k} t^k$$

with the constants

$$v_n = (-1)^n \int_{-\infty}^{+\infty} I(\tau)\tau^n d\tau = V_n(0). \tag{53}$$

Comparing (52) to (53) we conclude

$$\Theta(p) = \sum_{n\geq 0} v_n p^n/n!. \tag{54}$$

More generally, since e^{pt} is equal to $\sum_{n\geq 0} \frac{1}{n!} p^n t^n$, application of the linear operator $\mathbf{V}$ gives

$$\mathbf{V}[e^{pt}] = \sum_{n\geq 0} \frac{1}{n!} p^n V[t^n], \tag{55}$$

that is

$$\Theta(p)e^{pt} = \sum_{n\geq 0} \frac{1}{n!} p^n V_n(t) \tag{56}$$

Up to the change in notation (p for S, and t for x), the spectral gain $\Theta(p)$ is nothing else than the numerical exponential generating series associated to the binomial sequence $(V_n(t))_{n\geq 0}$.

Comparing with the results obtained in section 3.3, it is tempting to write the operator $\mathbf{V}$ as $\Theta(\mathbf{D})$. According to (54), $\Theta(\mathbf{D})$ can be interpreted as $\sum_{n\geq 0} v_n \mathbf{D}^n/n!$, but it is known that infinite order differential operators are not so easily dealt with. A better interpretation is obtained via Laplace or Fourier transform. Indeed since $\mathbf{D}$ multiplies e^{pt} by p, any function $F(\mathbf{D})$ ought to multiply e^{pt} by $F(p)$, and the rule $\mathbf{V}[e^{pt}] = \Theta(p)e^{pt}$ is in agreement

with the interpretation $\mathbf{V} = \Theta(\mathbf{D})$. If the input can be represented as a **Laplace transform**

$$f(t) = \int e^{pt}\phi(p)dp \tag{57}$$

then $\mathbf{V} = \Theta(\mathbf{D})$ transforms it into the output

$$F(t) = \int e^{pt}\Theta(p)\phi(p)dp. \tag{58}$$

Similarly, if the input $f(t)$ is given by its **spectral resolution** (or Fourier transform)

$$f(t) = \int_{-\infty}^{+\infty} \hat{f}(\omega)e^{i\omega t}d\omega, \tag{59}$$

then the output is given by

$$F(t) = \int_{-\infty}^{+\infty} \Theta(i\omega)\hat{f}(\omega)e^{i\omega t}d\omega. \tag{60}$$

This is **Heaviside's magic trilogy**:

symbolic	p
operator	$\mathbf{D}$
spectral	$i\omega = 2\pi i\nu$ (ν frequency, $\omega = 2\pi\nu$ pulsation)

$$\Theta(p) \longleftrightarrow \Theta(\mathbf{D}) \longleftrightarrow \Theta(2\pi i\nu)$$

Recall that in the Laplace transform, p is a complex variable, while in the Fourier transform ω and ν are real variables (see example in the next section).

3.6 The inverse problem

This is the problem of recovering the input, knowing the output. In operator terms, we have to compute the inverse $\mathbf{U}$ of the operator $\mathbf{V}$ (if it exists!). Since $\mathbf{V}$ is stationary, so is $\mathbf{U}$, and at the level of polynomial inputs and outputs, $\mathbf{U}$ corresponds to a binomial sequence of polynomials $U_n(t)$:

$$\begin{cases} \text{input} & U_n(t) \\ \text{output} & t^n. \end{cases}$$

Together with the numerical sequence $v_n = V_n(0)$, we have to consider the numerical sequence $u_n = U_n(0)$. Introducing the exponential generating series

$$u(S) = \sum_{n\geq 0} \frac{1}{n!}u_n S^n,\ v(S) = \sum_{n\geq 0} \frac{1}{n!}v_n S^n = \Theta(S), \tag{61}$$

we can write $\mathbf{U} = u(\mathbf{D})$ and $\mathbf{V} = v(\mathbf{D})$ at least when acting on polynomials. Since $\mathbf{U}$ and $\mathbf{V}$ are inverse operators, we expect the relation $u(S)v(S) = 1$, equivalent to the chain of relations

$$u_0 v_0 = 1, \ \sum_{k=0}^{n} \binom{n}{k} u_k v_{n-k} = 0 \ \text{ for } n \geq 1, \tag{62}$$

to hold. Indeed, this is easily checked (see section 3.4, formula (38)).

Since the input $U_n(t)$ corresponds to the output t^n, a Taylor-MacLaurin's expansion of the output corresponds to an expansion of the input in terms of the $U_n(t)$. Fix an epoch t_0 and use the Taylor expansion of the output

$$F(t) = \sum_{n \geq 0} \frac{1}{n!} \mathbf{D}^n F(t_0)(t - t_0)^n. \tag{63}$$

Applying the operator $\mathbf{U}$, we get

$$f(t) = \sum_{n \geq 0} \frac{1}{n!} \mathbf{D}^n F(t_0).U_n(t - t_0), \tag{64}$$

since $\mathbf{U}$ transforms $(t - t_0)^n$ into $U_n(t - t_0)$ by stationarity. The reader is invited to compare this formula to formula (39).

We give one illustrating example. Let the output be a moving average of the input

$$F(t) = \int_{t-1}^{t} f(s)ds, \tag{65}$$

corresponding to the following impulse response

The spectral gain is

$$\Theta(p) = \int_0^1 e^{-p\tau} d\tau = \frac{1 - e^{-p}}{p}. \tag{66}$$

The inverse series $v(p) = 1/\Theta(p)$ is given by

$$v(p) = \frac{p}{1 - e^{-p}}. \tag{67}$$

Since $v(-p) = \frac{p}{e^p - 1}$ is the exponential generating series of the Bernoulli numbers, the polynomials $U_n(t)$ are easily identified

$$U_n(t) = (-1)^n B_n(-t) = B_n(t) + nt^{n-1} = B_n(t + 1). \tag{68}$$

Notice that $\Theta(p)$ vanishes for $p \neq 0$ of the form $p = 2\pi in$ with an integer n; equivalently, the inverse function $v(p)$ has poles for $p \neq 0, \ p = 2\pi in$. Hence

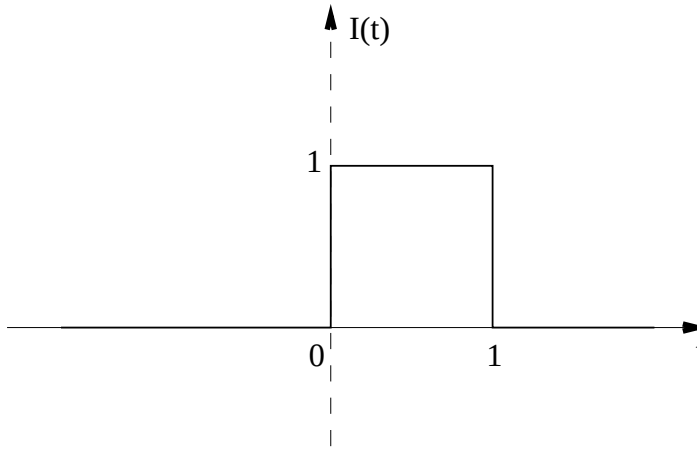

Fig. 3. The impulse response corresponding to (65)

not every output is admissible since (65) entails $\sum_n F(t+n) = \int_{-\infty}^{+\infty} f(s)ds$. That is, an output satisfies the necessary (and sufficient) condition

$$\sum_n F(t+n) = c \quad (c\text{constant}) \tag{69}$$

and the input $f(t)$ can be reconstructed from the output $F(t)$ up to the addition of a function $f_0(t)$ with

$$f_0(t) = f_0(t+1), \quad \int_0^1 f_0(t)dt = 0. \tag{70}$$

Exercise a) Derive from (64) the relation

$$f(t) = \sum_{n\geq 0} \frac{1}{n!} u_n \mathbf{D}^n F(t), \tag{71}$$

for a general transmission device, where $1/\Theta(p) = \sum_{n\geq 0} u_n p^n/n!$.

b) In the particular case (65), one gets $u_n = B_n(1)$, hence $u_n = B_n$ if $n \geq 2$, and $u_0 = 1$, $u_1 = \frac{1}{2}$.

c) Deduce from relation (7) that $B_n = 0$ for $n \geq 3$, n odd.

d) Derive the Euler-MacLaurin summation formula

$$\frac{1}{2}(f(t) + f(t-1)) = \int_{t-1}^{t} f(s)ds + \sum_{m\geq 1} \frac{1}{(2m)!} B_{2m}[\mathbf{D}^{2m-1}f(t) - \mathbf{D}^{2m-1}f(t-1)]. \tag{72}$$

3.7 A probabilistic application

We consider a random variable ξ. In general, we denote by $\langle X \rangle$ the mean value of a random variable X. We want to define a probabilistic version of the so-called Wick Powers in Quantum Field Theory.

The goal is to associate to ξ a sequence of random variables $: \xi^n :$ such that

a) **the mean value of** $: \xi^n :$ **is** 0 for $n \geq 1$;

b) **there exists a normalized binomial sequence of polynomials** $\Pi_n(X)$ **such that** $: \xi^n := \Pi_n(\xi)$.

Let $w(x)$ be the probability density associated to ξ, hence $w(x) \geq 0$ and $\int_{-\infty}^{+\infty} w(x)dx = 1$. Moreover, for any (non random) function $f(x)$ of a real variable x, the random variable $f(\xi)$ has a mean value given by

$$\langle f(\xi) \rangle = \int_{-\infty}^{+\infty} f(x)w(x)dx. \tag{73}$$

Hence the conditions a) and b) amount to

$$0 = \langle \Pi_n(\xi) \rangle = \int_{-\infty}^{+\infty} \Pi_n(x)w(x)dx \quad \text{for } n \geq 1. \tag{74}$$

Using the same method as in section 3.2, we introduce the exponential generating series $\Pi(X, S) = \sum_{n \geq 0} \Pi_n(X)S^n/n!$, hence the relation (74) translates as

$$\int_{-\infty}^{+\infty} \Pi(x, S)w(x)dx = 1. \tag{75}$$

Putting $\pi(S) = \Pi(0, S)$, hence $\Pi(X, S) = \pi(S)e^{XS}$, we derive

$$1/\pi(S) = \int_{-\infty}^{+\infty} e^{xS}w(x)dx. \tag{76}$$

We translate these relations into probabilistic jargon: replace S by p and x by ξ to get

$$1/\pi(p) = \langle e^{p\xi} \rangle \tag{77}$$

$$\Pi(\xi, p) = \sum_{n \geq 0} \frac{1}{n!} p^n : \xi^n : \tag{78}$$

$$\Pi(\xi, p) = \pi(p)e^{p\xi}. \tag{79}$$

Extending the definition of $: \ :$ by linearity to $e^{p\xi} = \sum_{n\geq 0} \frac{1}{n!} p^n \xi^n$, we rewrite (78) as $\Pi(\xi, p) =: e^{p\xi} :$. **Here is the conclusion**

$$: e^{p\xi} : \quad = \quad \frac{e^{p\xi}}{\langle e^{p\xi} \rangle}. \tag{80}$$

Let us specialize our results in the case of the binomial sequences considered so far:

a) If $\xi = 0$, then $\langle e^{p\xi} \rangle = 1$, hence $: e^{p\xi} := e^{p\xi} = 1$. That is $: \xi^n := 0$ for $n \geq 1$.

b) Suppose that ξ is **uniformly distributed** in the interval $[0, 1]$, that is $w(x) = 1$ if $0 \leq x \leq 1$, and $w(x) = 0$ otherwise. Then

$$\langle e^{p\xi} \rangle = \int_0^1 e^{px} dx = \frac{e^p - 1}{p}. \tag{81}$$

We get

$$\sum_{n\geq 0} \frac{1}{n!} p^n : \xi^n : \quad = \quad : e^{p\xi} : \quad = \quad \frac{p e^{p\xi}}{e^p - 1}, \tag{82}$$

that is

$$: \xi^n : \quad = \quad B_n(\xi) \tag{83}$$

where $B_n(X)$ is the Bernoulli polynomial of degree n. In particular

$$\begin{aligned} : \xi : \quad &= \quad \xi - \langle \xi \rangle = \xi - \frac{1}{2} \\ : \xi^2 : \quad &= \quad \xi^2 - \xi + \frac{1}{6} \\ : \xi^3 : \quad &= \quad \xi^3 - \frac{3}{2}\xi^2 + \frac{1}{2}\xi, \quad \text{etc} \ldots \end{aligned}$$

c) Assume now that ξ is **normalized**: $\langle \xi \rangle = 0$, $\langle \xi^2 \rangle = 1$, and follows a **Gaussian law**. Then $w(x) = (2\pi)^{-1/2} e^{-x^2/2}$ and

$$\langle e^{p\xi} \rangle = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} e^{px - x^2/2} dx = e^{p^2/2}. \tag{84}$$

Reasoning as above, we obtain

$$: \xi^n : \quad = \quad H_n(\xi) \tag{85}$$

where $H_n(X)$ is the Hermite polynomial of degree n. Explicitly

$$\begin{aligned} : \xi : \quad &= \quad \xi \\ : \xi^2 : \quad &= \quad \xi^2 - 1 \\ : \xi^3 : \quad &= \quad \xi^3 - 3\xi. \end{aligned}$$

To get a general formula, apply (80) to obtain the pair of relations

$$: e^{p\xi} : \quad = \quad e^{-p^2/2} e^{p\xi},\ e^{p\xi} = e^{p^2/2} : e^{p\xi} : . \tag{86}$$

Equating equal powers of p, we derive

$$: \xi^n : \quad = \quad \sum_{0 \le k \le n/2} (-1)^k \frac{n!}{2^k k!(n-2k)!} \xi^{n-2k}, \tag{87}$$

and conversely

$$\xi^n = \sum_{0 \le k \le n/2} \frac{n!}{2^k k!(n-2k)!} : \xi^{n-2k} : . \tag{88}$$

Notice that the orthogonality relation (46) for the Hermite polynomials translates in probabilistic terms as

$$\langle : \xi^m : \ : \xi^n : \rangle = m! \delta_{mn}, \tag{89}$$

hence the sequence $1, : \xi :, : \xi^2 :, \dots$ is derived from the natural sequence $1,\ \xi,\ \xi^2, \dots$ by orthogonalization.

To conclude, we can use the reflected probability density $w(-x)$ as an impulse response and define the input-output relation by

$$F(t) = \int f(t+\tau) w(\tau) d\tau, \tag{90}$$

that is

$$F(t) = \langle f(t+\xi) \rangle \tag{91}$$

in probabilistic terms. The interpretation is that the input is spoiled by random delay in transmission. Then $\mathit{\Pi}_n(t)$ is the input corresponding to the output t^n. Analytically this is expressed by

$$\int_{-\infty}^{+\infty} \mathit{\Pi}_n(t+\tau) w(\tau) d\tau = t^n \tag{92}$$

and probabilistically by

$$\langle : (\xi + t)^n : \rangle = t^n. \tag{93}$$

3.8 The Bargmann-Segal transform

Let us consider again the input-output transformation in the Gaussian case. It is then called the Bargmann-Segal transform (or B-transform), denoted by **B**. According to (90), we have

$$\mathbf{B} f(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x+z) e^{-x^2/2} dx \tag{94}$$

or

$$\mathbf{B}f(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-(z-x)^2/2} dx. \tag{94'}$$

Comparing formulas (23) and (27), one obtains

$$e^{-(z-x)^2/2} = e^{-x^2/2} \sum_{n\geq 0} H_n(x) z^n/n!, \tag{95}$$

and by integrating term by term in the expression (94′) one concludes

$$\mathbf{B}f(z) = \sum_{n\geq 0} \Gamma_n(f) z^n/n! \tag{96}$$

with

$$\Gamma_n(f) = \int_{-\infty}^{+\infty} H_n(x) f(x) d\gamma(x). \tag{97}$$

Taking into account the orthogonality property (46) namely

$$\int_{-\infty}^{+\infty} H_m(x) H_n(x) d\gamma(x) = \delta_{mn} n!,$$

one derives $\Gamma_n(f) = n! c_n$ for f given by a series $\sum_{n\geq 0} c_n H_n(x)$.
That is, **the B-transform takes** $\sum_{n\geq 0} c_n H_n(x)$ into $\sum_{n\geq 0} c_n z^n$.

To be more precise, we need to introduce some function spaces. The natural one is $L^2(d\gamma)$ consisting of the (measurable) functions $f(x)$ for which the integral $\int_{-\infty}^{+\infty} |f(x)|^2 d\gamma(x)$ is finite; the scalar product is given by

$$\langle f_1 | f_2 \rangle = \int_{-\infty}^{+\infty} \overline{f_1(x)} f_2(x) d\gamma(x). \tag{98}$$

In this space, the functions $He_n(x) := H_n(x)/(n!)^{1/2}$ (for $n = 0, 1, \dots$) form an orthonormal basis [5]. The B-transform takes this space onto the space of series $\sum c_n z^n$ with $\sum_{n\geq 0} n! |c_n|^2 < \infty$.

In its original form (94), the transformation $\mathbf{B}$ requires z to be real, but the form (94′) extends to the case of a complex number z. Indeed, from the property that $\sum_{n\geq 0} n! |c_n|^2$ is finite, it follows that the series $\sum_{n\geq 0} c_n z^n$ has an infinite radius of convergence, hence represents an entire function of the complex variable z. The space of such entire functions is denoted $\mathcal{F}(\mathbf{C})$ and called the **Fock space** (in one degree of freedom, see section 3.9).

[5] The orthonormality condition $\langle He_m | He_n \rangle = \delta_{mn}$ is nothing else than the orthogonality condition (46). But it requires a proof to show that this system is **complete**, that is that any function in the Hilbert space $L^2(d\gamma)$ can be approximated by polynomials (in the norm convergence).

The elements of $L^2(d\gamma)$ can be interpreted as the random variables of the form $X = f(\xi)$ with $\langle |X|^2 \rangle$ finite, where ξ is a normalized Gaussian random variable. We saw that $\mathbf{B}$ takes $H_n(\xi) = :\xi^n:$ into z^n. Hence it is tempting to denote by $:\ :$ the map inverse to $\mathbf{B}$, so that $:z^n: = H_n(x)$. We have **a new kind of operational calculus**

$$L^2(d\gamma) \underset{:\ :}{\overset{\mathbf{B}}{\rightleftharpoons}} \mathcal{F}(\mathbf{C}),$$

where $\mathbf{B}$ transforms a random variable $X = f(\xi)$ into the entire function

$$\mathbf{B}X(z) = e^{-z^2/2}\langle X.e^{z\xi}\rangle \tag{99}$$

and the inverse map takes an entire function $\Phi(z) = \sum_{n\geq 0} c_n z^n$ in the Fock space into the random variable $:\Phi(z) := \sum_{n\geq 0} c_n :\xi^n:$.

According to the definition (94), $\mathbf{B}$ takes the function e^{px} into $e^{pz+p^2/2}$, that is it acts as $e^{\mathbf{D}^2/2}$ where $\mathbf{D}$ is the derivation, followed by the change of variable x into z. This result can be reformulated as follows. Using the exponential generating series

$$\sum_{n\geq 0} H_n(x)p^n/n! = e^{px-p^2/2} \tag{100}$$

for the Hermite polynomials, and noting that $e^{\mathbf{D}^2/2}$ applied to $e^{pz-p^2/2}$ gives e^{px}, we conclude that $e^{\mathbf{D}^2/2}$ takes $H_n(x)$ into x^n, that is **$\mathbf{B}$ coincides on the polynomials with the differential operator $e^{\mathbf{D}^2/2}$ of infinite degree**.

We would like to conclude to the general rule

$$\mathbf{B} = e^{\mathbf{D}^2/2}, \tag{101}$$

hence

$$:\ : \ = e^{-\mathbf{D}^2/2}. \tag{102}$$

One way to substantiate these claims is to consider the heat (or diffusion equation)

$$\partial_s F(s,x) = \frac{1}{2}\partial_x^2 F(s,x) \tag{103}$$

with initial value

$$F(0,x) = f(x). \tag{104}$$

Since $\mathbf{D} = \partial_x$, the solution of equation (103) can be written formally as $F(s,x) = e^{s\mathbf{D}^2/2}f(x)$, hence $e^{\mathbf{D}^2/2}f(x)$ **represents the value for $s = 1$ of**

the solution of equation (103) which agrees for $s = 0$ with $f(x)$. But we know an explicit solution to the heat equation

$$F(s,x) = \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^{+\infty} f(x+u) e^{-u^2/2s} du. \tag{105}$$

Comparing with (94), we obtain

$$\mathbf{B}f(x) = F(1,x) \tag{106}$$

and this relation is the true expression of $\mathbf{B} = e^{\mathbf{D}^2/2}$.

The operator $e^{\mathbf{D}^2/2}$ (or $\mathbf{B}$) is **smoothing**. That is, if we simply assume that f belongs to $L^2(d\gamma)$ (that is that the integral $\int_{-\infty}^{+\infty} |f(x)|^2 e^{-x^2/2} dx$ is finite), then the function $F(1,x) = e^{\mathbf{D}^2/2} f(x)$ extends as an entire function in the complex domain. Conversely, the **Wick operator** $: \quad := e^{-\mathbf{D}^2/2}$ makes sense only for the functions $g(x)$ (for x real) which extend in the complex domain into a function $\Phi(z)$ (for z complex) belonging to the Fock space $\mathcal{F}(\mathbf{C})$.

3.9 The quantum harmonic oscillator

Let us rewrite the definition of the B-transform as an integral operator

$$\mathbf{B}f(z) = \int_{-\infty}^{+\infty} B(z,x) f(x) dx \tag{107}$$

with a kernel

$$B(z,x) = (2\pi)^{-1/2} e^{-(z-x)^2/2}. \tag{108}$$

It is often more convenient to replace the Hilbert space $L^2(d\gamma)$ by the Hilbert space $L^2(\mathbf{R})$ [6]. Defining the function $u_0(x)$ by $(2\pi)^{-1/4} e^{-x^2/4}$, we get

$$d_\gamma(x) = u_0(x)^2 dx \tag{109}$$

hence $\int |f(x)|^2 d_\gamma(x)$ is finite if and only if $\int |f(x) u_0(x)|^2 dx$ is finite. That is the multiplication by the function $u_0(x)$ gives an isometry of $L^2(d\gamma)$ onto $L^2(\mathbf{R})$. We can transfer the B-transform to $L^2(\mathbf{R})$, as the isometry $\mathbf{B}'$ of $L^2(\mathbf{R})$ onto $\mathcal{F}(\mathbf{C})$ [7] defined by $\mathbf{B}f = \mathbf{B}'(f u_0)$. Explicitly

$$\mathbf{B}'f(z) = \int_{-\infty}^{+\infty} B'(z,x) f(x) dx \tag{110}$$

[6] consisting of the (measurable) functions $\phi(x)$ such that $\int_{-\infty}^{+\infty} |\phi(x)|^2 dx$ be finite, with scalar product $\langle \phi_1 | \phi_2 \rangle = \int_{-\infty}^{+\infty} \overline{\phi_1(x)} \phi_2(x) dx$.

[7] with the scalar product $\langle \Phi_1 | \Phi_2 \rangle = \sum_{n \geq 0} n! \overline{c}_{n,1} c_{n,2}$ for $\Phi_j(z) = \sum_{n \geq 0} c_{n,j} z^n$ $(j = 1, 2)$.

with

$$B'(z,x) = u_0(x)^{-1}B(z,x) = (2\pi)^{-1/4}e^{-z^2/2+zx-x^2/4}. \tag{111}$$

Many properties are easier to describe in the Fock space. For instance the function 1 is called the **ground state** Ω, the multiplication by z is called the **creation operator**, denoted by $\mathbf{a}^*$, and the derivation $\partial_z = \frac{d}{dz}$ is the **annihilation operator**, denoted by $\mathbf{a}$. The vectors

$$e_n = \frac{1}{\sqrt{n!}}(\mathbf{a}^*)^n\Omega, \tag{112}$$

that is the functions $e_n(z) = \frac{1}{\sqrt{n!}}z^n$, form an orthonormal basis of $\mathcal{F}(\mathbf{C})$ with $e_0 = \Omega$. An easy calculation gives

$$\begin{cases} \mathbf{a}e_n = n^{1/2}e_{n-1} \text{ for } n \geq 1,\ \mathbf{a}e_0 = 0 \\ \mathbf{a}^*e_n = (n+1)^{1/2}e_{n+1}, \end{cases} \tag{113}$$

hence the matrices

$$\mathbf{a} = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & \dots \\ 0 & 0 & 0 & 0 & \dots \\ . & . & . & . & \dots \\ . & . & . & . & \dots \end{pmatrix}, \quad \mathbf{a}^* = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ \sqrt{1} & 0 & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ . & . & . & . & \dots \\ . & . & . & . & \dots \end{pmatrix}$$

in the basis $(e_n)_{n\geq 0}$; it follows that **a and $\mathbf{a}^*$ are adjoint to each other**. Moreover from the definitions $\mathbf{a}^* = z$, $\mathbf{a} = \partial_z$ follows the commutation relation

$$\mathbf{a}\mathbf{a}^* - \mathbf{a}^*\mathbf{a} = 1. \tag{114}$$

Finally the **number operator** $\mathbf{N} = \mathbf{a}^*\mathbf{a}$ is given by $\mathbf{N} = z\partial_z$, hence is diagonalized in the basis (e_n)

$$\mathbf{N}e_n = ne_n. \tag{115}$$

In the spirit of operational calculus, we transfer these results from the Fock space model to the spaces $L^2(d\gamma)$ and $L^2(\mathbf{R})$. The following table summarizes these translations (where ∂_x is $\frac{d}{dx}$):

Space	$L^2(d\gamma)$	$L^2(\mathbf{R})$	$\mathcal{F}(\mathbf{C})$
Ω	1	$(2\pi)^{-1/4}e^{-x^2/4} = u_0(x)$	1
e_n	$(n!)^{-1/2}H_n(x)$	$(n!)^{-1/2}H_n(x)u_0(x)$	$(n!)^{-1/2}z^n$
$\mathbf{a}^*$	$x - \partial_x$	$x/2 - \partial_x$	z
$\mathbf{a}$	∂_x	$x/2 + \partial_x$	∂_z
$\mathbf{N}$	$x\partial_x - \partial_x^2$	$-\partial_x^2 + x^2/4 - 1/2$	$z\partial_z$

For instance, the fact that $\mathbf{a}^*$ corresponds to $x - \partial_x$ in $L^2(d\gamma)$ is proved as follows: from the definition of $B(z,x)$ one gets

$$(x + \partial_x)B(z,x) = zB(z,x). \tag{116}$$

Multiplying by $f(x)$ and integrating by parts, we get

$$\int B(z,x)(x - \partial_x)f(x)dx = \int (x + \partial_x)B(z,x)f(x)dx$$
$$= z\int B(z,x)f(x)dx,$$

that is $\mathbf{B}((x - \partial_x)f) = z\mathbf{B}f$. The other cases are similar.

We apply these results to the **harmonic oscillator**. In classical mechanics, the harmonic oscillator is described by the **Hamiltonian** $H = \frac{p^2}{2m} + \frac{Kq^2}{2}$ in canonical coordinates p,q. The equation of motion is $\ddot{q} + \omega^2 q = 0$ with the pulsation $\omega = \sqrt{K/m}$, and the momentum $p = m\dot{q}$. To get the corresponding quantum Hamiltonian $\mathbf{H}$, replace p by the operator $\mathbf{p} = -i\hbar\partial_q$ hence

$$\mathbf{H} = -\frac{\hbar^2}{2m}\partial_q^2 + \frac{m\omega^2 q^2}{2}. \tag{117}$$

Introduce the dimensionless coordinate $x = (2m\omega/\hbar)^{1/2}q$. Then $\mathbf{H}$ can be rewritten as

$$\mathbf{H} = \hbar\omega(\mathbf{a}^*\mathbf{a} + \frac{1}{2}) \tag{118}$$

in the model $L^2(\mathbf{R})$. From the diagonalization of $\mathbf{N} = \mathbf{a}^*\mathbf{a}$, we conclude that the energy levels of the quantum harmonic oscillator (that is, the eigenvalues of $\mathbf{H}$) are given by $\hbar\omega(n + \frac{1}{2})$ with $n = 0, 1, 2, \ldots$: that is **Planck's radiation law**, with the correction $\frac{1}{2}$ giving $\frac{1}{2}\hbar\omega$ for the energy of the ground state $u_0 = \Omega$.

4 The art of manipulating infinite series

4.1 Some divergent series

Euler claimed that $S = 1-1+1-1+\dots$ is equal to $\frac{1}{2}$. Here is the purported proof:

$$\begin{array}{rl} & S = 1\ -1\ +1\ -1+\dots \\ +\ & S = \qquad 1\ -1\ +1-\dots \\ \hline & 2\ S = 1\ +0\ +0\ +0+\dots = 1 \end{array}$$

What is implicit is the use of two rules:

a) If $S = u_0 + u_1 + u_2 + \dots$, then $S = 0 + u_0 + u_1 + \dots$

b) If $S = u_0 + u_1 + u_2 + \dots$ and $S' = u_0 + u'_1 + u'_2 + \dots$, then $S + S' = (u_0 + u'_0) + (u_1 + u'_1) + (u_2 + u'_2) + \dots$.

These rules certainly hold for convergent series but to extend them to divergent series is somewhat hazardous.

Let us repeat the previous calculation in a slightly more general form:

$$\begin{array}{rl} & S = 1\ -t\ +t^2\ -t^3+\dots \\ +\ & tS = \qquad t\ -t^2\ +t^3-\dots \\ \hline & (1+t)S = 1\ +0\ +0\ +0+\dots = 1. \end{array}$$

The result is

$$1 - t + t^2 - t^3 + \dots = \frac{1}{1+t}, \tag{1}$$

the classical summation of the geometric series. If t is a real number such that $|t| < 1$, the geometric series is convergent, and the use of rules a) and b) is justified. To get Euler's result, take the limiting value $t = 1$ in (1).

What we need is the explicit description of various procedures to define rigorously the sum of certain divergent series (not all at once) and to compare these procedures. **Suppose we want to define the sum**

$$S = u_0 + u_1 + \dots . \tag{2}$$

Introduce weights $p_{0,t}, p_{1,t}, \dots$ and the weighted series

$$S_t = p_{0,t}u_0 + p_{1,t}u_1 + \dots . \tag{3}$$

If the series S_t is convergent for each value of the parameter t, and S_t approaches a limit S when t approaches some limiting value t_0, then S is the sum for this procedure [8].

The previous procedure is reasonable only when $\lim_{t\to t_0} p_{n,t} = 1$ for $n = 0, 1, \dots$. Some examples:

[8] See Knopp's book [8] for this method.

a) $p_{0,N} = p_{1,N} = \ldots = p_{N,N} = 1$, $p_{n,N} = 0$ for $n > N$ and $N = 0, 1, 2, \ldots$. Then the weighted sum amounts to the finite sum

$$S_N = u_0 + \ldots + u_N$$

(obviously convergent) and the convergence of S_N towards a limit S corresponds to the convergence of the series $u_0 + u_1 + u_2 + \ldots$ in the standard sense, with the standard sum S.

b) Put $\sigma_N = \frac{1}{N+1}(S_0 + \ldots + S_N)$; this corresponds to the weights

$$p_{n,N} = \begin{cases} 1 - \frac{n}{N+1} & \text{for } 0 \leq n \leq N \\ 0 & \text{for } n > N. \end{cases} \tag{4}$$

If σ_N converges to a limit σ, this is the **Cesaro-sum** of the series $u_0 + u_1 + u_2 + \ldots$.

c) To get the **Abel summation**, we introduce the weights $p_{n,t} = t^n$ for $n = 0, 1, 2, \ldots$ and a real parameter t with $0 < t < 1$. We take therefore the limit for $t = 1$ of the power series $\sum_{n \geq 0} u_n t^n$.

It is known that every convergent series with sum S is Cesaro-summable with the same sum $\sigma = S$. Similarly, Cesaro summation is extended by Abel summation. Euler's example is $u_n = (-1)^n$, hence

$$S_N = \begin{cases} 1 & \text{if } N > 0 \text{ is even} \\ 0 & \text{if } N > 0 \text{ is odd} \end{cases} \tag{5}$$

and therefore

$$\sigma_N = \begin{cases} \frac{1}{2} & \text{if } N \text{ is odd} \\ \frac{1}{2} + \frac{1}{2N+2} & \text{if } N \text{ is even.} \end{cases} \tag{6}$$

It follows that σ_N converges to $\sigma = \frac{1}{2}$. Hence the series $1 - 1 + 1 - 1 + \ldots$ is Cesaro-summable to $\frac{1}{2}$, and a previous calculation shows that it is Abel-summable to $\frac{1}{2}$ also.

The scope of Abel summation can be extended in various ways. For instance, if the sequence (u_n) is bounded, that is $|u_n| \leq M$ for $n = 0, 1, 2, \ldots$ with a constant M independent of n, then the power series $\sum_{n \geq 0} u_n z^n$ converges for any complex number z with $|z| < 1$ and defines therefore a holomorphic function $U(z)$ in the open disk $|z| < 1$ (see Fig. 4). If the limit $\lim_{r \to 1} U(re^{i\theta})$(for $0 \leq r < 1$) exists, it can be taken as an Abel sum for the series $\sum_{n \geq 0} u_n e^{in\theta}$.

In a slightly more general way, we can assume that the sequence (u_n) is polynomially bounded, that is

$$|u_n| \leq Cn^k$$

for all $n = 1, 2, \ldots$ and some constants $C > 0$ and $k = 1, 2, \ldots$ The radius of convergence of the series $\sum_{n \geq 0} u_n z^n$ is still at least 1, and if $U(1) =$

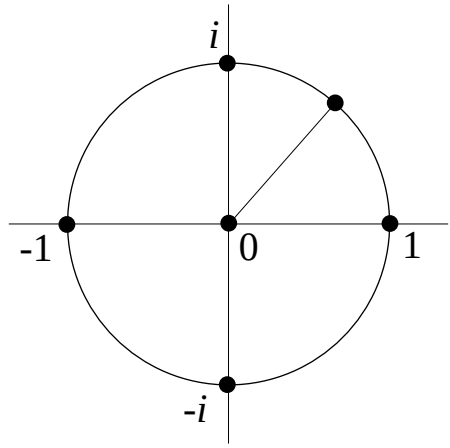

Fig. 4. The open unit disk

$\lim_{r\to 1} U(r) = \lim_{r\to 1} \sum_{n\geq 0} u_n r^n$ exists, it is the Abel sum for $u_0 + u_1 + u_2 + \ldots$

We just give one example, namely $u_n = (-1)^n n^k$ for $k = 0, 1, 2, \ldots$. We calculate

$$U(z) = \sum_{n\geq 0} u_n z^n = \sum_{n\geq 0} (-1)^n n^k z^n = \sum_{n\geq 0} n^k (-z)^n$$
$$= \sum_{n\geq 0} (z\partial_z)^k (-z)^n = (z\partial_z)^k \sum_{n\geq 0} (-z)^n = (z\partial_z)^k \frac{1}{1+z}.$$

Particular cases:

$$k = 0, \quad U(z) = \frac{1}{1+z}, \qquad U(1) = \frac{1}{2}$$
$$k = 1, \quad U(z) = \frac{-z}{(1+z)^2}, \qquad U(1) = -\frac{1}{4}$$
$$k = 2, \quad U(z) = \frac{z(z-1)}{(1+z)^3}, \qquad U(1) = 0$$
$$k = 3, \quad U(z) = \frac{-z^3 + 4z^2 - z}{(1+z)^4}, \quad U(1) = \frac{1}{8},$$

that is

$$\sum_{n\geq 0} (-1)^n = \frac{1}{2}$$

$$\sum_{n\geq 0} (-1)^n n = -\frac{1}{4}$$

$$\sum_{n\geq 0}(-1)^n n^2 = 0$$

$$\sum_{n\geq 0}(-1)^4 n^3 = \frac{1}{8}.$$

In general, we get $U(1) = (z\partial_z)^k \frac{1}{1+z}|_{z=1}$ that is, after the change of variable $z = e^u$,

$$\sum_{n\geq 0}(-1)^n n^k = \partial_u^k \frac{1}{e^u+1}|_{u=0}. \tag{7}$$

Using the exponential generating series for the Bernoulli numbers in the form

$$\frac{1}{e^u-1} = \frac{1}{u} + \sum_{k\geq 0}\frac{B_{k+1}}{(k+1)!}u^k$$

we obtain

$$\frac{1}{e^u+1} = \frac{1}{e^u-1} - \frac{2}{e^{2u}-1} = \sum_{k\geq 0}\frac{(1-2^{k+1})B_{k+1}}{(k+1)!}u^k, \tag{8}$$

hence Euler's result

$$\sum_{n\geq 0}(-1)^n n^k = \frac{(1-2^{k+1})B_{k+1}}{k+1}. \tag{9}$$

We leave it to the reader to rederive the previous cases $0 \leq k \leq 3$ using the values for B_1, B_2, B_3, B_4 given in section 3.1. We come back to this result in section 4.4.

Euler gave formulas for wildly divergent series like $\sum_{n\geq 0}(-1)^n n!$. Using the classical formula

$$n! = \int_0^\infty e^{-t}t^n dt \tag{10}$$

and assuming term by term integration, we get

$$\begin{aligned}\sum_{n\geq 0}(-1)^n n! &= \sum_{n\geq 0}(-1)^n \int_0^\infty e^{-t}t^n dt \\ &= \int_0^\infty e^{-t}\sum_{n\geq 0}(-t)^n\, dt = \int_0^\infty \frac{e^{-t}}{1+t}dt,\end{aligned}$$

the last integral being convergent. This is just the beginning of the use of Borel transform and Borel summation for divergent series.

4.2 Polynomials of infinite degree and summation of series

It is an important principle that **a polynomial can be reconstructed from its roots**. More precisely, let

$$P(z) = c_n z^n + c_{n-1} z^{n-1} + \ldots + c_1 z + c_0 \tag{11}$$

(with $c_n \neq 0$) be a polynomial of degree n with complex coefficients. If λ_1 is a root of P, that is $P(\lambda_1) = 0$, it is elementary to factorize $P(z) = (z-\lambda_1)P_1(z)$ where $P_1(z)$ is a polynomial of degree $n-1$. Continuing this process, we end up with a factorization

$$P(z) = (z-\lambda_1)\ldots(z-\lambda_m)Q(z) \tag{12}$$

where the polynomial $Q(z)$ of degree $n-m$ has no more roots. According to a highly non-trivial result, first stated by d'Alembert (1746) and proved by Gauss (1797), a polynomial without roots is a constant, hence the factorization (12) takes the form

$$P(z) = c_n(z-\lambda_1)\ldots(z-\lambda_n) \tag{13}$$

with $m = n$. By a well known calculation, one derives the following relations between coefficients and roots

$$\lambda_1 + \ldots + \lambda_n = -c_{n-1}/c_n$$
$$\sum_{i<j} \lambda_i \lambda_j = c_{n-2}/c_n, \text{ etc}\ldots$$

For our purposes, it is better to use the inverses of the roots, assumed to be nonzero. Since the logarithmic derivative transforms product into sum and annihilates constants, we derive

$$\mathbf{D}P(z)/P(z) = \sum_{i=1}^{n} \frac{1}{z-\lambda_i}. \tag{14}$$

Using the geometric series gives

$$\sum_{i=1}^{n} \frac{1}{z-\lambda_i} = -\sum_{i=1}^{n} \sum_{k\geq 0} z^k/\lambda_i^{k+1}. \tag{15}$$

Introducing the sums of inverse powers of roots

$$\gamma_k = \sum_{i=1}^{n} \lambda_i^{-k}, \tag{16}$$

we conclude from this calculation

$$z\mathbf{D}P(z) + P(z)\sum_{k\geq 1} \gamma_k z^k = 0. \tag{17}$$

Assuming for simplicity $c_0 = 1$ and equating the coefficients of equal powers of z, we obtain the following variant of **Newton's relations**

$$\gamma_k + c_1\gamma_{k-1} + \ldots + c_{k-1}\gamma_1 + kc_k = 0 \tag{18}$$

for $k \geq 1$. It is important to notice that **the degree n of $P(z)$ does not appear explicitly in the relation (18)**, which can be solved inductively

$$\gamma_1 = -c_1 \tag{19}$$
$$\gamma_2 = c_1^2 - 2c_2 \tag{20}$$
$$\gamma_3 = -c_1^3 + 3c_1c_2 - 3c_3 \tag{21}$$
$$\gamma_4 = c_1^4 - 4c_1^2c_2 + 4c_1c_3 - 4c_4 + 2c_2^2. \tag{22}$$

Around 1734, Euler undertook to calculate the sum of the series $S_2 = \sum_{n\geq 1} \frac{1}{n^2}$. This series is slowly convergent, but Euler invented efficient acceleration methods for summing series and calculated the sum $S_2 = 1.64493406\ldots$; he recognized $S_2 = \pi^2/6$. He obtained also the value of $S_4 = \sum_{n\geq 1} 1/n^4$ to be $\pi^4/90$. To establish these results rigorously, he introduced the equation $\sin x = 0$ admitting the solutions $x = 0, \pm\pi, \pm 2\pi, \pm 3\pi, \ldots$ Discarding the root $x = 0$ and using the power series expansion of $\sin x$, we are led to consider the equation

$$1 - \frac{x^2}{6} + \frac{x^4}{120} - \ldots = 0$$

with roots $\pm\pi, \pm 2\pi, \pm 3\pi, \ldots$. With the previous notations we have

$$c_1 = 0,\ c_2 = -\frac{1}{6},\ c_3 = 0,\ c_4 = \frac{1}{120}, \ldots$$
$$\gamma_2 = \sum_{n\geq 1}[\frac{1}{(\pi n)^2} + \frac{1}{(-\pi n)^2}] = 2S_2/\pi^2$$
$$\gamma_4 = \sum_{n\geq 1}[\frac{1}{(\pi n)^4} + \frac{1}{(-\pi n)^4}] = 2S_4/\pi^4.$$

Assuming that the relations (20) and (22) still hold, we get

$$2S_2/\pi^2 = \gamma_2 = -2c_2 = \frac{1}{3}$$
$$2S_4/\pi^4 = \gamma_4 = -4c_4 + 2c_2^2 = -\frac{1}{30} + \frac{1}{18} = \frac{1}{45}.$$

The sought for relations

$$S_2 = \pi^2/6,\ S_4 = \pi^4/90$$

follow immediately.

To summarize the method used by Euler:

a) first guess the value from accurate numerical work;
b) consider the function

$$\frac{\sin\ x}{x} = 1 - \frac{x^2}{6} + \frac{x^4}{120} - \dots$$

as a polynomial of infinite degree, with infinitely many roots $\pm\pi, \pm 2\pi, \pm 3\pi, \dots$;

c) since the Newton's relations (19) to (22) don't involve explicitly the degree n of the polynomial, assume their validity in the case $n = \infty$ as well, and exploit them for $P(x) = (\sin\ x)/x$.

4.3 The Euler-Riemann zeta function

We use Riemann's definition and notation

$$\zeta(s) = \sum_{n \geq 1} n^{-s}. \tag{23}$$

The series converges absolutely for any complex number s with real part $\Re(s)$ greater than 1. It has been shown by Riemann that $\zeta(s)$ can be analytically continued to the whole complex plane, the only singularity being a pole of order 1 at $s = 1$, that is $\zeta(s) - 1/(s-1)$ is an entire function. Obviously $\zeta(1) = \sum_{n \geq 1} 1/n$ is a divergent series, but $\zeta(s)$ is defined when $s \neq 1$ is an integer (positive or negative). Euler was the first to calculate $\zeta(s)$ when s is an integer.

We consider the case where s is **even and strictly positive**. Euler proved the formula

$$\zeta(2k) = \frac{2^{2k-1}\pi^{2k}|B_{2k}|}{(2k)!} \tag{24}$$

and in particular

$$\zeta(2) = \frac{2\pi^2|B_2|}{2!} = \frac{\pi^2}{6} \tag{25}$$

$$\zeta(4) = \frac{8\pi^4|B_4|}{4!} = \frac{\pi^4}{90}. \tag{26}$$

Since $\zeta(2) = \sum_{n \geq 1} 1/n^2 = S_2$ and similarly $\zeta(4) = S_4$, we recover the formulas for S_2 and S_4. The method we used in the previous section could be extended to cover the general case (24), but it is simpler to go back to the formula given for the logarithmic derivative in (14). For the function $\sin z$, the logarithmic derivative is $\cot z = \cos z / \sin z$. This function is meromorphic in the whole complex plane, with simple poles of residue 1 at each integral

multiple of π. **Euler assumed at first that, in analogy with (14),** $\cot z$ **should be equal to the sum of its polar contributions, that is**

$$\cot z = \sum_{n=-\infty}^{+\infty} \frac{1}{z - n\pi}. \tag{27}$$

Assume this relation for a moment, and derive (24). The series (27) is not absolutely convergent, but can be summed in a symmetrical way by taking $\sum_{n=-\infty}^{+\infty}$ to be $\lim_{N\to\infty} \sum_{n=-N}^{+N}$. Hence

$$\cot z - \frac{1}{z} = \sum_{n=1}^{\infty} \frac{2z}{z^2 - n^2\pi^2}. \tag{28}$$

The right-hand side can be developed using the geometric series; for $|z| < \pi$, the series involved are absolutely convergent, hence

$$\sum_{n=1}^{\infty} \frac{2z}{z^2 - n^2\pi^2} = -\sum_{n=1}^{\infty} 2z \sum_{k=1}^{\infty} \frac{(z^2)^{k-1}}{(n^2\pi^2)^k} = -2\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \frac{z^{2k-1}}{n^{2k}\pi^{2k}}$$
$$= -2\sum_{k=1}^{\infty} \frac{z^{2k-1}}{\pi^{2k}} \sum_{n=1}^{\infty} \frac{1}{n^{2k}},$$

that is

$$\cot z = \frac{1}{z} - 2\sum_{k\geq 1} \frac{\zeta(2k)}{\pi^{2k}} z^{2k-1}. \tag{29}$$

Using again the exponential generating series for the Bernoulli numbers yields

$$\cot z = i\frac{e^{2iz}+1}{e^{2iz}-1} = \frac{2i}{e^{2iz}-1} + i$$
$$= 2i\{\frac{1}{2iz} - \frac{1}{2} + \sum_{k\geq 1} \frac{B_{2k}}{(2k)!}(2iz)^{2k-1}\} + i$$

hence finally

$$\cot z = \frac{1}{z} + \sum_{k\geq 1} \frac{(-1)^k 2^{2k} B_{2k}}{(2k)!} z^{2k-1}. \tag{30}$$

To establish (24), it is enough to compare the expansions (29) and (30) for $\cot z$ and to remark that $\zeta(2k) = \sum_{n\geq 1} \frac{1}{n^{2k}}$ is the sum of a convergent series of positive numbers hence $\zeta(2k) > 0$.

Euler's proof for the expansion (27) of $\cot z$ is reproduced in many textbooks. Here is a variant which seems to have been unnoticed so far. Define

$$\Phi(z) = \cot z - \sum_{n=-\infty}^{+\infty} \frac{1}{z - n\pi}. \tag{31}$$

Examining the poles of $\cot z$, we see that $\Phi(z)$ is an entire function of the complex variable z. A simple manipulation yields the functional equation

$$\Phi(z) = \frac{1}{2}[\Phi(\frac{z}{2}) + \Phi(\frac{z+\pi}{2})]; \tag{32}$$

we have to prove that $\Phi(z) = 0$ for all z.

a) **The function Φ is bounded**: indeed, denote by C_n the set of complex numbers whose modulus is at most $(2^n + 1)\pi$. Since C_1 is a compact set and Φ is continuous, there exists a constant $M > 0$ such that $|\Phi(z)| \leq M$ for z in C_1. Assuming the estimate $|\Phi(z)| \leq M$ for z in C_n, we use the functional equation for z in C_{n+1}

$$|\Phi(z)| \leq \frac{1}{2}|\Phi(\frac{z}{2})| + \frac{1}{2}|\Phi(\frac{z+\pi}{2})| \tag{33}$$

and remark that both $z/2$ and $(z+\pi)/2$ belongs to C_n, hence $|\Phi(z/2)| \leq M$, $|\Phi((z+\pi)/2)| \leq M$; from (33) we conclude that $|\Phi(z)| \leq M$ (for z in C_{n+1}). Every complex number belongs to some set C_n, hence $|\Phi(z)| \leq M$ for all z.

b) We appeal now to **Liouville's theorem** to conclude that Φ, being a bounded entire function is a constant, hence $\Phi(z) = \Phi(0)$.

c) The function Φ is **odd**, that is $\Phi(-z) = -\Phi(z)$, hence $\Phi(0) = 0$.

Liouville's theorem, the main ingredient in this proof, was proved around 1850, a century after Euler worked on these questions. It is interesting to note that d'Alembert-Gauss theorem is an easy corollary of Liouville's theorem [**Hint**: if $P(z)$ is a polynomial without zeroes, the function $\Phi(z) = 1/P(z)$ is entire and bounded, hence a constant; that is, $P(z)$ is a constant].

4.4 Sums of powers of numbers

The other result of Euler about $\zeta(s)$ can be stated as follows

$$1^k + 2^k + 3^k + \ldots = -\frac{B_{k+1}}{k+1} \tag{34}$$

for $k = 1, 2, 3, \ldots$ It looks at first suspicious, since it gives a finite value to an infinite sum of positive numbers, obviously divergent since each term is at least 1. Euler's derivation is more or less as follows.

Formula (9) can be written as

$$1^k - 2^k + 3^k - 4^k + \ldots = -(1 - 2.2^k)\frac{B_{k+1}}{k+1}. \tag{35}$$

On the other hand, multiply the right-hand side of (34) by $1 - 2.2^k$ and rearrange. This yields

This yields

$$\begin{aligned}
&(1-2.2^k)(1^k+2^k+3^k+\dots) = \\
&\quad 1^k+2^k+3^k+4^k+5^k+6^k+\dots \\
-2(&\qquad 2^k \qquad +4^k \qquad +6^k+\dots) \\
\hline
=&\quad 1^k-2^k+3^k-4^k+5^k-6^k+\dots
\end{aligned}$$

and finally (34) is obtained from (35).

This procedure is highly questionable, but can be fixed as follows. We introduce two functions

$$\zeta(s)=\sum_{n\geq 1} n^{-s}, \quad \eta(s)=\sum_{n\geq 1}(-1)^{n-1}n^{-s}. \tag{36}$$

Provided these functions can be continued analytically to the negative integers, formula (34) and (35) read respectively as

$$\zeta(-k)=-\frac{B_{k+1}}{k+1} \tag{34$'$}$$

$$\eta(-k)=(2^{k+1}-1)\frac{B_{k+1}}{k+1} \tag{35$'$}$$

for $k=1,2,\dots$ Furthermore

$$\begin{aligned}
\eta(s) &= \sum_{\text{n odd}} n^{-s} - \sum_{\text{n even}} n^{-s} = \sum_{n\geq 1} n^{-s} - 2\sum_{\text{n even}} n^{-s} \\
&= \zeta(s) - 2\sum_{m\geq 1}(2m)^{-s}
\end{aligned}$$

and finally

$$\eta(s)=(1-2^{1-s})\zeta(s). \tag{37}$$

Our manipulation of series is justified as long as $\Re(s)>1$, but the final formula remains valid for all s for which both $\zeta(s)$ and $\eta(s)$ are regular (analytic continuation!). In particular

$$\eta(-k)=(1-2^{k+1})\zeta(-k). \tag{38}$$

Hence formulas (34$'$) and (35$'$) are equivalent, substantiating Euler's derivation.

Using the known values of the Bernoulli numbers, we deduce

$$\zeta(-2)=\zeta(-4)=\zeta(-6)=\dots=0$$

$$\zeta(-1)=-\frac{1}{12},\ \zeta(-3)=\frac{1}{120},\ \zeta(-5)=-\frac{1}{252},\dots.$$

It can also be shown that $\zeta(0) = -1/2$. Hence we get the paradoxical results:

$$\zeta(0) = 1 + 1 + 1 + \ldots = -\frac{1}{2}$$
$$\zeta(-1) = 1 + 2 + 3 + \ldots = -\frac{1}{12}.$$

Among the many methods available to construct the analytical continuation of $\zeta(s)$, we select the following one using $\eta(s)$. Indeed, from Euler's definition of the gamma function

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt \tag{39}$$

(for $\Re(s) > 1$), we deduce by a simple change of variable

$$\Gamma(s) n^{-s} = \int_0^\infty e^{-nt} t^{s-1} dt. \tag{40}$$

By summation and term by term integration

$$\Gamma(s)\eta(s) = \sum_{n\geq 1} (-1)^{n-1} \Gamma(s) n^{-s} = \sum_{n\geq 1} (-1)^{n-1} \int_0^\infty e^{-nt} t^{s-1} dt$$
$$= \int_0^\infty \{\sum_{n\geq 1} (-1)^{n-1} e^{-nt}\}.t^{s-1} dt,$$

that is

$$\Gamma(s)\eta(s) = \int_0^\infty \frac{t^{s-1}}{e^t + 1} dt. \tag{41}$$

All the calculations are justified as long as $\Re(s) > 1$. We use now a general principle, unknown to Euler, to deal with integrals of the type

$$\Phi(s) = \int_0^\infty F(t) t^{s-1} dt. \tag{42}$$

We split the integral as $\int_0^1 + \int_0^\infty$.

a) Assuming that $F(t)$ decreases at infinity faster than any power t^{-k} (for $k \geq 0$), then

$$\Phi_{1,\infty}(s) = \int_1^\infty F(t) t^{s-1} dt \tag{43}$$

extends to an entire function.

b) Assuming that $F(t)$ is differentiable to any order on the closed interval $[0, 1]$, then

$$\Phi_{0,1}(s) = \int_0^1 F(t) t^{s-1} dt \tag{44}$$

extends to a meromorphic function in the complex plane $\mathbf{C}$. The only singularities [9] are at $s = 0, -1, -2, \ldots$ with singular part $\frac{\mathbf{D}^k F(0)}{k!(s+k)}$ around $s = -k$.

Applying this principle to the definition (39) of $\Gamma(s)$ we recover the well-known fact that $\Gamma(s)$ extends as a meromorphic function, with poles at $s = 0, -1, -2, \ldots$ and singular part $\frac{(-1)^k}{k!(s+k)}$ around $s = -k$. We use now formula (41) for $\Gamma(s)\eta(s)$. Hence this function extends to a meromorphic function with poles at $s = 0, -1, \ldots$ and singular part $\frac{c_k}{s+k}$ around $s = -k$, where

$$\frac{1}{e^t + 1} = \sum_{k \geq 0} c_k t^k. \tag{45}$$

According to (8), we get

$$c_k = \frac{(1 - 2^{k+1})B_{k+1}}{(k+1)!}. \tag{46}$$

Dividing $\Gamma(s)\eta(s)$ by $\Gamma(s)$, the poles cancel; hence $\eta(s)$ extends as an entire function, and comparing the singular parts of $\Gamma(s)\eta(s)$ and $\Gamma(s)$ around $s = -k$, we find

$$\eta(-k) = (-1)^k k! c_k. \tag{47}$$

We have to distinguish several cases:
-$k = 0$ yields $\eta(0) = c_0 = -B_1 = \frac{1}{2}$;
-$k \geq 2$ is even yields $\eta(-k) = 0$ since $B_r = 0$ for $r = k + 1$ odd;
-$k \geq 1$ is odd, then $(-1)^k = -1$ and

$$\eta(-k) = \frac{(2^{k+1} - 1)B_{k+1}}{k+1}. \tag{48}$$

This is Euler's formula (35) or formula (35′). The analytical continuation of $\zeta(s)$ can now be performed by using (37), that is we define

$$\zeta(s) = \frac{\eta(s)}{1 - 2^{1-s}}. \tag{49}$$

Since $\eta(s)$ is entire, the only singularity of $\zeta(s)$ is a pole at $s = 1$, with singular part $\frac{1}{s-1}$. We can calculate now $\zeta(-k)$ from $\eta(-k)$ and get

$$\zeta(-k) = -\frac{B_{k+1}}{k+1} \tag{50}$$

for $k = 1, 2, \ldots$ as expected. Furthermore, from $\eta(0) = \frac{1}{2}$ we get the remaining value

$$\zeta(0) = -\eta(0) = -\frac{1}{2}. \tag{51}$$

[9] **Hint**: integrate by parts using $\frac{d}{dt} t^{-s} = -s t^{-s-1}$.

4.5 Variation I: Did Euler really fool himself?

Bourbaki wrote (in [2], page VI.29): "Mais la tendance au calcul formel est la plus forte, et l'extraordinaire intuition d'Euler lui-même ne l'empêche pas de tomber parfois dans l'absurde, lorsqu'il écrit par exemple $0 = \sum_{n=-\infty}^{+\infty} x^n$". Did Euler really fool himself?

To keep with our habits (after Cauchy!) denote by z a complex variable and try to evaluate the sum of $I = \sum_{n=-\infty}^{+\infty} z^n$. We break the sum into $I_+ + I_- - 1$, where

$$I_+ = \sum_{n\geq 0} z^n, \quad I_- = \sum_{n\leq 0} z^n.$$

By the geometric series, we get $I_+ = \frac{1}{1-z}$ and $I_- = \frac{1}{1-1/z}$ and simple algebra gives

$$I_+ + I_- = \frac{1}{1-z} + \frac{z}{z-1} = \frac{1-z}{1-z} = 1, \tag{52}$$

hence $I = 0$ as claimed. What is paradoxical is that there is no complex number $z \neq 0$ for which both series I_+ and I_- converge simultaneously, since $\sum_{n\geq 0} z^n$ converges for $|z| < 1$ and $\sum_{n\leq 0} z^n$ converges for $|z| > 1$. We really need analytical continuation: I_+ as a function of z extends from the convergence domain $|z| < 1$ to $\mathbf{C} - \{1\}$ as the rational function $\frac{1}{1-z}$, and one goes from $I_+$to I_- by inverting z (into $1/z$). If both I_+ and I_- are extended in this way to $\mathbf{C} - \{1\}$, the calculation (52) is perfectly valid, hence $I = 0$ in this sense.

Another method to prove $I = 0$ is to remark that multiplying I by z shifts z^n to z^{n+1}, hence rearranges the series, hence $Iz = I$, hence $I(z-1) = 0$, and by dividing by $z - 1$, we get $I = 0$ for $z \neq 1$. Nevertheless, there is some trouble. Consider the critical region $|z| = 1$ where both I_+ and I_- diverge, and use polar coordinates $z = e^{2\pi iu}$. Then I is the series

$$J(u) = \sum_{n=-\infty}^{+\infty} e^{2\pi inu}. \tag{53}$$

Playing with Fourier series, introduce a test function $f(u)$ supposed to be smooth (i.e. infinitely differentiable) and periodic $f(u+1) = f(u)$. We expand it as a Fourier series

$$f(u) = \sum_{n=-\infty}^{+\infty} c_n e^{2\pi inu} \tag{54}$$

with

$$c_n = \int_0^1 e^{-2\pi inu} f(u)du. \tag{55}$$

From (54) we get, by putting $u = 0$,

$$f(0) = \sum_{n=-\infty}^{+\infty} c_n \tag{56}$$

hence by (55) and (53)

$$f(0) = \sum_{n=-\infty}^{+\infty} \int_0^1 e^{-2\pi i n u} f(u)du = \int_0^1 \{ \sum_{n=-\infty}^{+\infty} e^{-2\pi i n u} \}.f(u)du$$

and finally

$$f(0) = \int_0^1 J(u)f(u)du. \tag{57}$$

Remove now the assumption $f(u+1) = f(u)$ by introducing a smooth function $\phi(u)$ vanishing off some finite interval and by defining

$$f(u) = \sum_{m=-\infty}^{+\infty} \phi(u+m) \tag{58}$$

(an absolutely convergent series). By an easy manipulation, one derives from (57)

$$\sum_{m=-\infty}^{+\infty} \phi(m) = \int_{m=-\infty}^{+\infty} J(u)\phi(u)du. \tag{59}$$

Using the standard Dirac's function $\delta(u)$, we get by definition

$$\phi(m) = \int_{-\infty}^{+\infty} \phi(u)\delta(u-m)du,$$

hence

$$0 = \int_{-\infty}^{+\infty} \{J(u) - \sum_{m=-\infty}^{+\infty} \delta(u-m)\}\phi(u)du. \tag{60}$$

Since the test function ϕ is arbitrary, we can omit it from (60), hence the conclusion

$$J(u) = \sum_{m=-\infty}^{+\infty} \delta(u-m). \tag{61}$$

That is, by substituting $e^{2\pi i u}$ to z, **the series** $I = \sum_{n=-\infty}^{+\infty} z^n$ **is not** 0 **but** $\sum_{m=-\infty}^{+\infty} \delta(u-m)$. So Euler was wrong, but not too much, since $\delta(u-m) = 0$ for $u \neq m$, hence $\sum_{n=-\infty}^{+\infty} z^n$ is 0 for $z \neq 1$.

Recall the other proof, using

$$I(z-1)=0; \tag{62}$$

division by $z-1$ gives $I=0$, provide $z\neq 1$, corresponding to $u\notin \mathbf{Z}$ for $z=e^{2\pi iu}$. Formula (62) is equivalent to

$$J(u)(e^{2\pi iu}-1)=0, \tag{63}$$

and this suggests a new proof of (61). Indeed, if $f(u)$ is a smooth function with isolated simple zeros u_m, then $J(u)f(u)=0$ implies that $J(u)$ is a linear combination of terms $c_m\delta(u-u_m)$. Here $f(u)=e^{2\pi iu}-1$, hence $u_m=m$ for m in $\mathbf{Z}$, that is $m=0,\pm 1,\pm 2,\dots$ hence $J(u)=\sum_{m=-\infty}^{+\infty} c_m\delta(u-m)$ for suitable coefficients c_m. But $J(u+1)=J(u)$, hence all coefficients c_m are equal to some constant c and $J(u)=c\sum_{m=-\infty}^{+\infty}\delta(u-m)$. It remains to calculate the normalization constant c. That kind of argument could be understood by Euler, but it acquires now a rigorous meaning due to Laurent Schwarz's theory of distributions (200 years after Euler!) [10].

Another version of our proof is by using contour integral (see Fig. 5). Consider a function $\Phi(z)$ holomorphic in a domain containing the annulus

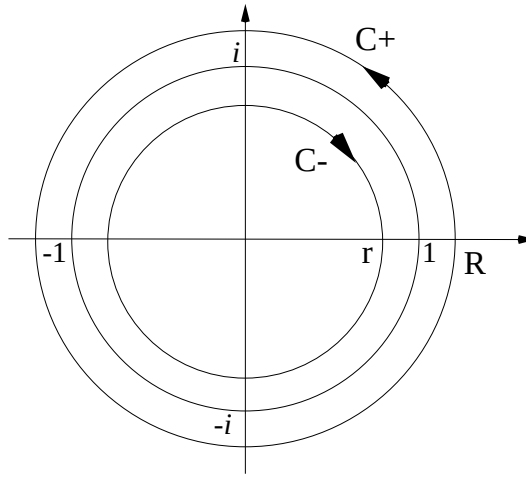

Fig. 5. Path for the contour integral

$r\leq |z|\leq R$ bounded by C_+ and C_- (beware the orientations). The rational function $R(z)=\frac{1}{z-1}$ is given by a convergent series $\sum_{n=-\infty}^{-1} z^n$ for $|z|>1$,

[10] Our final result can be expressed as $\sum_{n=-\infty}^{+\infty} e^{2\pi inu}=\sum_{m=-\infty}^{+\infty}\delta(u-m)$. It is equivalent to Poisson's summation formula.

hence for z in C_+. It follows

$$\int_{C_+} R(z)\Phi(z)dz = \sum_{n=-\infty}^{-1} \int_{C_+} z^n \Phi(z)dz. \tag{64}$$

Similarly

$$\int_{C_-} R(z)\Phi(z)dz = \sum_{n=0}^{\infty} \int_{C_-} z^n \Phi(z)dz \tag{65}$$

and using the residue formula

$$\int_{C_+} \sum_{n=-\infty}^{-1} z^n.\Phi(z)dz + \int_{C_-} \sum_{n=0}^{\infty} z^n.\Phi(z)dz = 2\pi i\Phi(1).$$

A shorthand would be

$$\sum_{n=-\infty}^{+\infty} z^n = 2\pi i\delta(z-1) \tag{66}$$

using δ-functions in the complex domain [11].

Let us go back to sums of powers and Bernoulli numbers and polynomials. A classical formula reads as follows

$$B_k(u) = -k! \sum_{n\neq 0} \frac{e^{2\pi inu}}{(2\pi in)^k}. \tag{67}$$

A complex version is as follows

$$\sum_{n\neq 0} \frac{z^n}{n^k} = -\frac{(2\pi i)^k}{k!} B_k(\frac{\log z}{2\pi i}) \tag{68}_k$$

[11] A classical formula is

$$\delta(f(u)) = \sum_m \frac{1}{|f'(u_m)|}\delta(u-u_m),$$

the summation being extended to the solutions of the equation $f(u_m) = 0$ (provided $f'(u_m) \neq 0$). In this formula $f(u)$ is a real-valued function of a real variable u. Assuming that it remains valid for $f(u) = e^{2\pi iu} - 1$ (with complex values), we derive

$$2\pi i\delta(z-1) = \sum_{m=-\infty}^{+\infty} \delta(u-m)$$

for $z = e^{2\pi iu}$. This brings together our two methods.

(by the change of variables $z = e^{2\pi i u}$) for $k = 0, 1, 2, \ldots$. For $k = 0$, this reads as Euler's "absurd formula" as

$$\sum_{n \neq 0} z^n = -1 \tag{68}_0$$

since $B_0(x) = 1$. The case $k = 1$ is

$$\sum_{n \neq 0} \frac{z^n}{n} = \pi i - \log z; \tag{68}_1$$

of course

$$\sum_{n=1}^{\infty} \frac{z^n}{n} = \log \frac{1}{1-z}$$

$$\sum_{n=-\infty}^{-1} \frac{z^n}{n} = -\sum_{m=1}^{\infty} \frac{(z^{-1})^m}{m} = -\log \frac{1}{1-1/z}$$

and $(68)_1$ amounts to

$$\log \frac{1}{1-z} - \log \frac{1}{1-1/z} = \pi i - \log z. \tag{69}$$

Since $\log(-1) = -\pi i$ and $\frac{1}{1-1/z} = \frac{z}{z-1} = \frac{z.(-1)}{1-z}$, this relation follows from $\log uv = \log u + \log v$, but some care has to be exercized with the multivalued complex logarithm (notice the ambiguity $\log(-1) = \pm\pi i$ for instance).

For the general case, notice the following. Define

$$L_k(z) = \sum_{n \neq 0} \frac{z^n}{n^k}, \quad R_k(z) = -\frac{(2\pi i)^k}{k!} B_k(\frac{\log z}{2\pi i}). \tag{70}$$

We know already that $L_0(z) = R_0(z)$ and $L_1(z) = R_1(z)$. Furthermore, it is obvious that

$$z \frac{d}{dz} L_k(z) = L_{k-1}(z), \tag{71}$$

and from the fact that the derivative of $B_k(x)$ is $kB_{k-1}(x)$, one gets

$$z \frac{d}{dz} R_k(z) = R_{k-1}(z). \tag{72}$$

So we can easily conclude that $L_2(z) - R_2(z)$ is a constant, which has to be shown to be 0 to prove $L_2(z) = R_2(z)$. Then $L_3(z) - R_3(z)$ is a constant, etc...

This line of argument can be made rigorous. Introduce the polylogarithmic functions [12]

$$\mathrm{Li}_k(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^k}. \tag{73}$$

Our formula reads now as follows:

$$\mathrm{Li}_k(z) + (-1)^k \mathrm{Li}_k(\frac{1}{z}) = -\frac{(2\pi i)^k}{k!} B_k(\frac{\log z}{2\pi i}). \tag{74}_k$$

To make sense out of it, we proceed as follows:

a) We cut the complex plane along the real interval $[0, +\infty[$, to get $\Omega_0 = \mathbf{C} - [0, +\infty[$.

b) In the cut plane, we choose the somewhat unusual branch of the logarithm $\log(re^{i\theta}) = \log r + i\theta$ for $0 < \theta < 2\pi$.

c) We define the function $\mathrm{Li}_k(z)$ by the convergent series (73) for $|z| < 1$, and verify that

$$z\frac{d}{dz}\mathrm{Li}_k(z) = \mathrm{Li}_{k-1}(z) \tag{75}$$

for $k = 1, 2, \ldots$ and $\mathrm{Li}_0(z) = \frac{z}{1-z}$. Since the cut plane $\Omega_1 = \mathbf{C} - [1, \infty[$ is simply connected, any holomorphic function in Ω_1 has a primitive, hence by (75), each $\mathrm{Li}_k(z)$ extends analytically to Ω_1.

d) For z in Ω_0, both z and $\frac{1}{z}$ are in Ω_1, hence both $\mathrm{Li}_k(z)$ and $\mathrm{Li}_k(\frac{1}{z})$ are defined for z in Ω_0, and formula $(74)_k$ is asserted for z in Ω_0.

e) The cases $k = 0$ and $k = 1$ are settled as before.

f) From (75) and the rule for the derivative of $B_k(x)$, we get that the validity of $(74)_k$ for the index k implies that of $(74)_{k+1}$ for the index $k + 1$ up to the addition of a constant. To show that it is 0 use the fact that for $k \geq 2$, the series $\sum_{n=1}^{\infty} \frac{z^n}{n^k}$ converges also for $|z| = 1$, and study the limiting value for $z \to 1$, using $B_k(0) = B_k(1)$.

So after all, Euler was right!

Putting $z = 1$ in $(74)_k$ we obtain the value of $\zeta(k) + (-1)^k\zeta(k)$. For k odd, we get $0 = 0$, but for k even, we recover the value of $\zeta(k)$ given by (24).

4.6 Variation II: Infinite products

Suppose we want to calculate $\infty! = 1.2.3\ldots$. Going to logarithms we define

$$\infty! = \exp(\sum_{n=1}^{\infty} \log n). \tag{76}$$

[12] The dilogarithm $Li_2(z)$ was known by Euler, and further developed in the 19$^{\text{th}}$ century in connection with Lobatchevski geometry. Fifteen years ago, the subject was almost forgotten, to be resurrected by geometers and mathematical physicists alike. It is now a hot subject of research.

Suppose we have a series $\sum_{n\geq 1} a_n$ with na_n bounded. The ζ-summation procedure fits our general framework in section 2: consider the convergent series $\sum_{n\geq 1} a_n n^{-\epsilon}$ for $\epsilon > 0$, and let ϵ tend to 0. To sum the series $\sum_{n\geq 1} \log n$, we should consider $\sum_{n\geq 1} n^{-\epsilon}.\log n$ but this converges for $\epsilon > 1$ only and we cannot go directly to the limit $\epsilon \to 0$. What we have to do is to consider the series $\sum_{n\geq 1} n^{-s}.\log n$ for $\Re(s) > 1$; this is obviously the derivative $-\zeta'(s)$ of the Riemann zeta function, hence it can be analytically continued to the neighborhood of 0. The regularized sum of $\sum_{n\geq 1} \log n$ is then $-\zeta'(0)$ and finally

$$\infty! = e^{-\zeta'(0)}. \tag{77}$$

From the formulas (37) and (41), one derives without much ado $\zeta'(0) = -\frac{1}{2}\log 2\pi$ (a formula more or less equivalent to Stirling's formula). Conclusion:

$$\infty! = \sqrt{2\pi}. \tag{78}$$

General rule: to normalize a divergent product $\Pi_{n\geq 1} a_n$, introduce the series $\sum_{n\geq 1} a_n^{-s} = Z(s)$, make an analytic continuation from the convergence domain $\Re(s) > \sigma_0$ to $s = 0$ and define

$$\prod_{n\geq 1}^{\text{reg}} a_n := e^{-Z'(0)}. \tag{79}$$

Generalizing slightly (78), we can use this method to prove the identity [13]

$$\prod_{n\geq 0}^{\text{reg}} (n+v) = \frac{\sqrt{2\pi}}{\Gamma(v)}. \tag{80}$$

We can compare this to the Weierstrass product expansion for the gamma function

$$\frac{1}{\Gamma(v)} = ve^{\gamma v} \prod_{n\geq 1} (1+\frac{v}{n})e^{-v/n}. \tag{81}$$

A careless, but nevertheless instructive, comparison of (80) and (81) is as follows:

$$\prod_{n\geq 1}(1+\frac{v}{n})e^{-v/n} = \prod_{n\geq 1}\frac{n+v}{n}e^{-v/n}$$
$$= \prod_{n\geq 1}(n+v)(\prod_{n\geq 1} n)^{-1}\exp(-v\sum_{n\geq 1} 1/n)$$

[13] Formally: $\frac{\infty!}{(\infty+v)!} = \Gamma(v+1)$.

and regularizing the divergent series $\sum_{n\geq 1} \frac{1}{n}$ by the Euler constant γ, we are through! Notice that the two most important properties of Γ, namely

1) the functional equation $\Gamma(v+1) = v\Gamma(v)$;

2) the function $\frac{1}{\Gamma(v)}$ of a complex variable v is entire with zeros at $0, -1, -2, \ldots$ can be read off immediately from (80).

According to our general method, the proof of (80) requires to study the function

$$\zeta(s,v) = \sum_{n\geq 0} (n+v)^{-s} \tag{82}$$

known as Hurwitz zeta function (see [9] for more details). We list a few properties:

a) **a particular case** $\zeta(s,1) = \zeta(s)$;

b) **functional equations**:

$$\zeta(s, v+1) = \zeta(s,v) - v^{-s} \tag{83}$$

$$\partial_v \zeta(s,v) = -s\zeta(s+1,v); \tag{84}$$

c) **analytic continuation**: for fixed v, $\zeta(s,v)$ can be analytically continued to the complex plane with one singularity at $s=1$, with singular part $\frac{1}{s-1}$; hence $\zeta(s,v) - \zeta(s)$ is an entire function;

d) **special values**:

$$\zeta(-k,v) = -\frac{B_{k+1}(v)}{k+1} \tag{85}$$

for $k = 0, 1, 2, \ldots$

The last relation can be written, in the spirit of Euler, as

$$v^k + (v+1)^k + (v+2)^k + \ldots = -\frac{B_{k+1}(v)}{k+1}. \tag{86}$$

As a particular case we get the surprising identity

$$v^0 + (v+1)^0 + (v+2)^0 + \ldots = \frac{1}{2} - v. \tag{87}$$

5 Conclusion: From Euler to Feynman

Feynman is the modern heir to Euler. Among his many contributions to theoretical physics, the most famous one is his use of diagrams to encode in a very compact way complicated integrals with significance in experiments in high energy physics. His method of diagrams has been generalized by various authors (Cvitanovic, Penrose,...) to provide a very flexible tool for computations in tensor analysis.

His really bold discovery is the use of integrals in function spaces (see for instance [5]), the so-called Feynman path integrals. These (so far) ill-defined integrals are powerful tools to evaluate infinite series and infinite products. We give just one example. Consider the Hilbert space $L^2(0, 2\pi)$ of functions $f(x)$ with $0 < x < 2\pi$ and $\int_0^{2\pi} |f(x)|^2 dx$ finite. The unbounded operator $\Delta = -d^2/dx^2$ can be diagonalized with eigenfunctions $e_n(x) = e^{inx}$ (for $n = 0, \pm 1, \pm 2, \dots$) corresponding to the eigenvalue n^2. Hence the characteristic determinant $\det(v - \Delta)$ is an entire function with the eigenvalues as zeros. Using our normalized products, one now defines the regularized determinant as

$$\det{}^{\mathrm{reg}}(v - \Delta) = v(\prod_{n \geq 1}^{\mathrm{reg}} (v - n^2)^2) \tag{1}$$

(0 is a simple eigenvalue, and $1^2, 2^2, \dots$ are eigenvalues of multiplicity 2). This can be evaluated by a formula due to Euler

$$\sin v = v \prod_{n \geq 1} (1 - \frac{v^2}{n^2 \pi^2}), \tag{2}$$

equivalent (via logarithmic derivatives) to the formula

$$\cot v = \frac{1}{v} + \sum_{n \geq 1} \frac{2v}{v^2 - n^2 \pi^2} \tag{3}$$

also due to Euler, and considered above.

Feynman bold step is as follows. From matrix calculus, we learn the following integral formula for a characteristic determinant

$$\det(v - A) = [\int_{\mathbf{R}^n} d^n x \ \exp -\pi(v \sum_{i=1}^{n} x_i^2 - \sum_{i,j} a_{i,j} x_i x_j)]^2 \tag{4}$$

where $d^n x$ is the volume element $dx_1 \dots dx_n$ in the Euclidean space $\mathbf{R}^n$, and $A = (a_{i,j})$ is a real symmetric, positive definite, matrix of size $n \times n$. By analogy, Feynman writes $\det(v - \Delta)$ as the square of

$$\int_{L^2(0,2\pi)} \mathcal{D}x . \exp -\pi S(x), \tag{5}$$

where the so-called action $S(x)$ is defined by

$$S(x) = v \int_0^{2\pi} x(t)^2 dt - \int_0^{2\pi} x'(t)^2 dt \tag{6}$$

(the variable in $[0, 2\pi]$ is denoted by t, the function in $L^2(0, 2\pi)$ by $x(t)$, and its derivative by $x'(t)$). The symbol $\mathcal{D}x$ is formally a volume element in the

Hilbert space $L^2(0, 2\pi)$ (infinite-dimensional generalization of the Euclidean space $\mathbf{R}^n$), sometimes written as $C \prod_t dx(t)$. Its rigorous definition is the main problem [5].

Part of these calculations have been put into a rigorous framework, but not all of them.

After all, Feynman shall be right!

Acknowledgements

Ils vont à Michel Planat et Michel Waldschmidt, pour leur impeccable organi sation du colloque de Chapelle des Bois, et pour leur conviction que ce texte verrait le jour. Leurs amicales pressions et leur aide matérielle ont grandement contribué à l'heureux dénouement.

References

1. Borel E. (1902) Leçons sur les séries à termes positifs, Gauthier-Villars
2. Bourbaki N. (1976) Fonctions d'une variable réelle (chapitres 1 à 7). Hermann, edited by Masson since 1980
3. Bourbaki N. (1984) Eléments d'histoire des mathématiques. Masson
4. Cartier P. (1966) Quantum mechanical relations and theta functions, in Algebraic groups and discontinuous subgroups. Proc. Symp. Pure Maths, vol.IX, American Mathematical Society, Providence, 361-383
5. De Witt-Morette C., Cartier P. and Folacci P. (editors) (1997) Functional Integration, Basics and Applications, Nato Asi, Series B, vol 361, Plenum Press
6. Euler L. (1755) Institutiones calculi differentialis cum ejus usu in analysi infinitorum ac doctrina serierum. Berlin, 1910
7. Hardy G. H. (1924) Orders of infinity. Cambridge Univ. Press (reprinted by Hafner, New York, 1971)
8. Knopp K. (1964) Theorie und Anwendung der unendlichen Reihen, Springer (5th edition)
9. Waldschmidt M. et al (1995) From Number Theory to Physics. Springer (see Chapter 1 by P. Cartier: An introduction to zeta functions)
10. Weil A. (1984) Number Theory, An Approach Through History, Birkhäuser

Thermal and Quantum Noise in Active Systems

Jean-Michel Courty ⋆, Francesca Grassia ⋆⋆, and Serge Reynaud⋆ ⋆ ⋆

Laboratoire Kastler Brossel †,
UPMC case 74, 4 place jussieu, F75252 Paris Cedex 05, France

Abstract. We present a quantum network approach to the treatment of thermal and quantum fluctuations in measurement devices. The measurement is described as a scattering process of input fluctuations towards output ones. We present the results obtained with this method for the treatment of a cold damped capacitive accelerometer.

1 Non-ideal Quantum Measurements

Active systems are fundamental elements in high precision measurements. Amplifiers are used either for amplifying the signal to a macroscopic level or to make the system work around its optimal working point with the help of feedback loops. With techniques such as cold damping, it is possible to manipulate actively the fluctuations and to reduce the effective noise temperature of the devices well below the operating temperature. The analysis of sensitivity limits in these devices rises many questions related to fundamental processes as well as experimental constraints. How far is it possible to reduce the measurement temperature? How are these process related to the fluctuation dissipation theorem? Are there quantum limits to this noise reduction associated with Heisenberg inequalities? How do the experimental constraints interplay with the fundamental limitations of the sensitivity?

The aim of the present paper is to address these questions with quantum network theory. This approach provides a rigorous thermodynamical framework able to withstand the constraints of a quantum analysis of the measurement. In the same time, it makes possible a realistic description of real measurement devices. Thermodynamic and quantum fluctuations are treated in the same footing. The measurement process is described as a scattering process allowing for a modular analysis of real quantum systems. Active systems such as the linear amplifier or the ideal operational amplifier are

⋆ courty@spectro.jussieu.fr
⋆⋆ grassia@spectro.jussieu.fr
⋆ ⋆ ⋆ reynaud@spectro.jussieu.fr
† UMR de l'Ecole Normale Supérieure, de l'Université Pierre et Marie Curie et du Centre National de la Recherche Scientifique

M. Planat (Ed.): LNP 550, pp. 71–83, 2000.
© Springer-Verlag Berlin Heidelberg 2000

described in this framework. Here, the approach will be illustrated by analyzing the sensitivity of a cold damped capacitive accelerometer developed for fundamental physics applications in space [1–3].

We first present the analysis of passive electrical systems in term of quantum networks. Then, we use this approach to present the quantum analysis of an operational amplifier working in the ideal limit of infinite gain, infinite input impedance and null output impedance. In the last section, we illustrate the theoretical framework with the example of a cold damped accelerometer.

2 Coupling with the Environment

Relations between fluctuations and dissipation have first been discovered by Einstein who studied the viscous damping of mechanical systems [4]. Another important application was the study of Johnson-Nyquist noise in resistive electrical elements [5]. This classical result was extended to take into account the quantum statistical properties of fluctuations [6,7]. A general approach of these relations was widely studied in the framework of linear response theory [8,9].

2.1 Dissipation and fluctuations

A first insight into the physical effect of the coupling of an electrical circuit to the environment is provided by the analysis of an antenna in an electrical resonator. When a current flows through the antenna, electromagnetic radiation is emitted and the resonator energy decreases. As far as the electric circuit is concerned, the effect of the antenna is the same as a resistance. The antenna is also able to detect electromagnetic fields. An incoming wave puts into motion the electrons in the antenna and causes an electrical current to flow in the circuit. For thermal radiation, the detection radiation leads to a random current which brings the electrical oscillator to thermal equilibrium. In the high temperature limit, it leads to the usual thermodynamic $\frac{1}{2}k_BT$ per degree of freedom, with k_B being Boltzmann constant and T the radiation temperature. In the zero temperature limit, the detected field corresponds to the vacuum fluctuations of the electromagnetic field and the induced energy of the oscillator is the zero point energy $\frac{1}{2}\hbar\omega_0$, with ω_0 the resonance frequency of the oscillator.

In figure 1 are depicted two representations for a resistance R. Figure 1a corresponds to the Thevenin representation with the noise source represented as a voltage noise generator U_n. The relation between the current I and the voltage V is

$$U = RI + U_n \tag{1}$$

Figure 1 b corresponds to a model that originates from Nyquist's analysis [5]. It consists in a semi infinite coaxial line of characteristic impedance R.

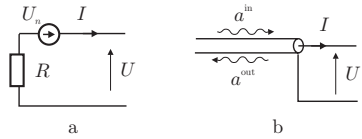

Fig. 1. Representations of a resistance R. (a) Thevenin representation with a voltage noise generator U_n. (b) Model with a semiinfinite line and propagating fields a^{in} and a^{out}

The solution of the propagation equations in the line may be written as the sum of two counterpropagating fields $I^{\text{out}}$and I^{in}

$$\begin{aligned} I(x,t) &= I^{\text{out}}\left(t+\frac{x}{c}\right) - I^{\text{in}}\left(t-\frac{x}{c}\right) \\ U(x,t) &= R\left(I^{\text{out}}\left(t+\frac{x}{c}\right) + I^{\text{in}}\left(t-\frac{x}{c}\right)\right) \end{aligned} \tag{2}$$

At the end of the line, we deduce the following relations:

$$\begin{aligned} U &= RI + 2RI^{\text{in}} = RI + U_n \\ I^{\text{out}} &= I + I^{\text{in}} \end{aligned} \tag{3}$$

The first equation corresponds to the relation (1) and leads to the identification of the noise as the input current I^{in}. The second equation describes the output fields I^{out} emitted back to the line. This output field may be used either to feed other elements of the system or to perform a measurement by extracting information on the system of interest through a line considered as the detection channel.

2.2 Treatment with quantum fields

In an infinite line, current and voltage may be treated as quantum fields propagating in a two dimensional space-time. Throughout the paper, we will consider that a function f is defined in the time domain (notation $f(t)$) or in the frequency domain (Kubo's notation $f[\omega]$) and that these two representations are related through the Fourier transform with the convention of quantum mechanics

$$f(t) = \int \frac{d\omega}{2\pi} f[\omega]\, e^{-i\omega t} \tag{4}$$

The electronics convention may be recovered by substituting j to $-i$.

Free field operators a^{in} and a^{out} can be defined as the Fourier components of $I^{\text{in}}$and I^{out}

$$\begin{aligned} I(x,t) &= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \sqrt{\frac{\hbar|\omega|}{2R}} \left(a^{\text{out}}[\omega] \exp\left[-i\omega\left(t+\frac{x}{c}\right)\right] \right. \\ &\quad \left. - a^{\text{in}}[\omega] \exp\left[-i\omega\left(t-\frac{x}{c}\right)\right] \right) \\ U(x,t) &= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \sqrt{\frac{\hbar|\omega| R}{2}} \left(a^{\text{out}}[\omega] \exp\left[-i\omega\left(t+\frac{x}{c}\right)\right] \right. \\ &\quad \left. + a^{\text{in}}[\omega] \exp\left[-i\omega\left(t-\frac{x}{c}\right)\right] \right) \end{aligned} \tag{5}$$

They are normalized so that they obey the standard commutation relations

$$\left[a^{\text{in}}[\omega], a^{\text{in}}[\omega']\right] = \left[a^{\text{out}}[\omega], a^{\text{out}}[\omega']\right] = 2\pi\ \delta(\omega+\omega')\ \varepsilon(\omega) \tag{6}$$

where $\varepsilon(\omega)$ denotes the sign of the frequency ω. This relation just means that the positive and negative frequency components correspond respectively to the annihilation a_ω and creation $a_\omega^\dagger$ operators of quantum field theory

$$a^{\text{in}}[\omega] = a_\omega \theta(\omega) + a_{-\omega}^\dagger \theta(-\omega) \tag{7}$$

$\theta(\omega)$ denotes the Heavyside function.

To characterize the fluctuations of these noncommuting operators, we use the correlation function defined as the average value of the symmetrized product. With stationary noise, the correlation function depends only on the time difference

$$\begin{aligned} &\left\langle a^{\text{in}}(t) \cdot a^{\text{in}}(t') \right\rangle = \sigma_{aa}^{\text{in}}(t-t') \\ &\left\langle a^{\text{in}}[\omega] \cdot a^{\text{in}}[\omega'] \right\rangle = 2\pi\ \delta(\omega+\omega')\ \sigma_{aa}^{\text{in}}[\omega] \end{aligned} \tag{8}$$

The dot symbol denotes a symmetrized product for quantum operators.

In the case of a thermal bath, the noise spectrum is

$$\sigma_{aa}^{\text{in}}[\omega] = \frac{1}{\exp\frac{\hbar|\omega|}{k_B T_a} - 1} + \frac{1}{2} = \frac{1}{2} \coth \frac{\hbar|\omega|}{2k_B T_a} \tag{9}$$

One recognizes the black body spectrum or the number of bosons per mode for a field at temperature T_a and a term $\frac{1}{2}$ corresponding to the quantum fluctuations. The energy per mode will be denoted in the following as an effective temperature Θ_a

$$k_B \Theta_a = \hbar|\omega|\, \sigma_{aa}^{\text{in}}[\omega] = \frac{\hbar|\omega|}{2} \coth \frac{\hbar|\omega|}{2k_B T_a} \tag{10}$$

In the high temperature limit the classical energy for an harmonic field of $k_B T_a$ per mode is recovered. In the low temperature limit, the energy $\frac{\hbar|\omega|}{2}$ corresponding to the ground state of a quantum harmonic oscillator is obtained.

Note that the term $\frac{1}{2}$ corresponding to the zero point quantum fluctuations was added by Planck so that the difference with the classical result $k_B T_a$ tends to zero in the high temperature limit [10].

These results are easily translated to obtain the expression of the Johnson Nyquist noise power

$$\sigma_{U_n U_n}[\omega] = 2R\hbar\,|\omega|\,\sigma_{aa}^{\text{in}}[\omega] = 2Rk_B\Theta_a \tag{11}$$

Our symmetric definition of the noise power spectrum leads to a factor 2 difference with the electronic convention where only positive frequencies are considered.

2.3 Quantum networks

The elementary systems described up to now as well as more complex devices to be studied later in this paper may be described by using a systematic approach which may be termed as "quantum network theory". Initially designed as a quantum extension of the classical theory of electrical networks [11], this theory was mainly developed through applications to optical systems [12,13]. It has also been viewed as a generalized quantum extension of the linear response theory which is of interest for electrical systems as well [14]. It is fruitful for analyzing non-ideal quantum measurements containing active elements [15,16].

In this quantum network approach, the various fluctuations entering the system, either by dissipative or by active elements, are described as input fields in a number of lines as depicted on 1 b.

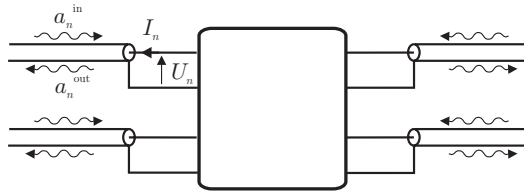

Fig. 2. Representation of an electrical circuit as a quantum network. The central box is a reactive multipole which connects noise lines corresponding to the fluctuations entering the system, either by dissipative or by active elements. For example, the upper left port n with voltage U_n and current I_n is connected to a line of impedance R_n with inward and outward fields a_n^{in} and a_n^{out}.

We first consider a passive linear network built with resistances and reactive elements like capacitances or inductances. Each resistance R_n is modeled as a semi-infinite coaxial line a_n with characteristic impedance R_n. The voltage U_n and current I_n associated with the resistance are the inward and

outward fields a_n^{in} and a_n^{out} evaluated at the end of this line

$$\begin{aligned}\mathbf{I} &= \mathbf{R}^{-\frac{1}{2}}\sqrt{\frac{\hbar\,|\omega|}{2}}\left(\mathbf{a}^{\text{out}} - \mathbf{a}^{\text{in}}\right)\\ \mathbf{U} &= \mathbf{R}^{\frac{1}{2}}\sqrt{\frac{\hbar\,|\omega|}{2}}\left(\mathbf{a}^{\text{out}} + \mathbf{a}^{\text{in}}\right)\end{aligned} \tag{12}$$

Here, $\mathbf{X} = \mathbf{I}, \mathbf{U}, \mathbf{a}^{\text{in}}, \mathbf{a}^{\text{out}}$ denotes the column vector with components X_n and $\mathbf{R}$ is the diagonal matrix formed with the characteristic impedances R_n.

Input fields corresponding to different lines commute with each other. For simplicity, we also consider that the fields entering through the various ports are uncorrelated with each other. The interaction with the reactive elements is described by a reactive impedance matrix $\mathbf{Z}$

$$\begin{aligned}\mathbf{U} &= -\mathbf{Z}\,\mathbf{I}\\ \mathbf{Z}^{\dagger} &= -\mathbf{Z}\end{aligned} \tag{13}$$

The whole network is then associated with a scattering $\mathbf{S}$ matrix, also called repartition matrix [17], describing the transformation from the input fields to the output ones

$$\begin{aligned}\mathbf{a}^{\text{out}} &= \mathbf{S}\,\mathbf{a}^{\text{in}}\\ \mathbf{S} &= \frac{\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}\mathbf{R}^{-\frac{1}{2}} - \mathbf{1}}{\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}\mathbf{R}^{-\frac{1}{2}} + \mathbf{1}}\end{aligned} \tag{14}$$

The output fields a^{out} are also free fields which obey the same commutation relations 6 as the input ones. In other words, $\mathbf{S}$ matrix is unitary. In the case of the passive network, this property is an immediate consequence of the reactive nature of the impedance matrix $\mathbf{Z}$

$$\mathbf{S}^{\dagger} = \frac{\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}^{\dagger}\mathbf{R}^{-\frac{1}{2}} - \mathbf{1}}{\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}^{\dagger}\mathbf{R}^{-\frac{1}{2}} + \mathbf{1}} = \frac{-\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}\mathbf{R}^{-\frac{1}{2}} - \mathbf{1}}{-\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}\mathbf{R}^{-\frac{1}{2}} + \mathbf{1}} = \mathbf{S}^{-\mathbf{1}} \tag{15}$$

More generally, the unitarity of the $\mathbf{S}$ matrix is required to ensure the quantum consistency of the description. In the following section, we will make use of this property to deduce general properties of amplifiers.

3 Fluctuations in Amplifiers

Quantum noise associated with linear amplifiers has been the subject of numerous works. In the line of thought initiated by early works on fluctuation-dissipation relations, active systems have been studied in the optical domain when maser and laser amplifiers were developed [18–20]. General thermodynamical constraints impose the existence of fluctuations for amplification as

well as dissipation processes. The added noise determines the ultimate performance of linear amplifiers [21,22] and plays a key role in the question of optimal information transfer in optical communication systems [23,24].

We first consider the amplification of a field, for example in long distance telecommunication systems with repeaters.

The amplification of the field a^{in} with a phase insensitive gain G is given by the following equation

$$a^{\text{out}} = Ga^{\text{in}} + B^{\text{in}} \tag{16}$$

where B^{in} is a noise added by the amplification. The gain G may be frequency dependent. The commutator of the output field is then

$$\left[a^{\text{out}}\left[\omega\right], a^{\text{out}}\left[\omega'\right]\right] = \left|G\right|^2 \left[a^{\text{in}}\left[\omega\right], a^{\text{in}}\left[\omega'\right]\right] + \left[B^{\text{in}}\left[\omega\right], B^{\text{in}}\left[\omega'\right]\right] \tag{17}$$

The unitarity of the input output transformation and the preservation of the commutation implies a non zero commutator for B^{in}

$$\begin{aligned}\left[B^{\text{in}}\left[\omega\right], B^{\text{in}}\left[\omega'\right]\right] &= \left[a^{\text{in}}\left[\omega\right], a^{\text{in}}\left[\omega'\right]\right] - \left[a^{\text{out}}\left[\omega\right], a^{\text{out}}\left[\omega'\right]\right] \\ &= \left(1 - \left|G\right|^2\right)\ 2\pi\ \delta\left(\omega + \omega'\right)\ \varepsilon\left(\omega\right)\end{aligned} \tag{18}$$

This result does not depend on the specific amplification process. For a gain larger than unity, the added noise can be represented by a free field b^{in} with the usual commutation relation (6)

$$B^{\text{in}}\left[\omega\right] = \sqrt{\left|G\right|^2 - 1} b^{\text{in}}\left[-\omega\right] = \sqrt{\left|G\right|^2 - 1}\left(b^{\text{in}}\left[\omega\right]\right)^{\dagger} \tag{19}$$

The presence of the conjugation is characteristic of amplification processes and is encountered as soon as gain is present [21,22] .

We may use this example to describe the noise analysis in a measurement process. Let us consider that the field a^{in} carries a signal A superimposed with fluctuations c^{in}

$$A = \left\langle a^{\text{in}}\right\rangle, \qquad c^{\text{in}} = a^{\text{in}} - \left\langle a^{\text{in}}\right\rangle \tag{20}$$

The input noise power Σ_{AA}^{in} corresponds to the fluctuations σ_{cc}^{in}

$$\Sigma_{AA}^{\text{in}} = \sigma_{cc}^{\text{in}} \tag{21}$$

The measurement corresponds to the output a^{out} of the amplifier

$$a^{\text{out}} = GA + Gc^{\text{in}} + \sqrt{\left|G\right|^2 - 1}\ b^{\text{in}\dagger} \tag{22}$$

To analyze the noise of this amplified signal, we define an estimator $\widehat{A}$ by normalizing the output field $a^{\text{out}}$of the amplifier so that it is the sum of A and an extra noise

$$\widehat{A} = \frac{1}{G} a^{\text{out}} = A + c^{\text{in}} + \sqrt{1 - \frac{1}{\left|G\right|^2}}\ b^{\text{in}\dagger} \tag{23}$$

The added noise Σ_{AA}^{out} is then described by a spectrum

$$\Sigma_{FF}^{\text{out}} = \sigma_{cc}^{\text{in}} + \left(1 - \frac{1}{|G|^2}\right) \sigma_{bb}^{\text{in}} \tag{24}$$

In the limit of large gain G and for thermal fluctuations, it corresponds to

$$\hbar\,|\omega|\,\Sigma_{FF}^{\text{out}} = k_B(\Theta_a + \Theta_b) \tag{25}$$

When the temperatures are equal, this corresponds to a loss of $3dB$ in the signal to noise ratio. This effect has been observed since the beginning of radiowave communications. It also sets a limit in the number of repeaters in optical fiber communications[23,24].

Most practical applications of amplifiers in measurements involve ideal operational amplifiers operating in the limits of infinite gain, infinite input impedance and null output impedance. In order to deal with the pathologies that could arise in such a system, we consider that it operates with a feedback loop which fixes its effective gain and effective impedances [25].

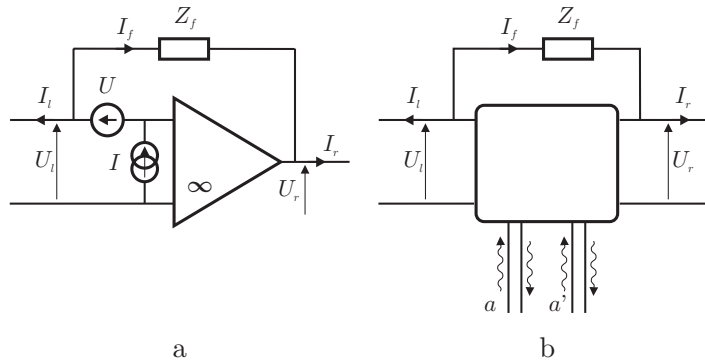

Fig. 3. Representation of the ideal operational amplifier working in the limit of infinite gain with a reactive feedback Z_f. (a) The noise sources are described as a current generator I and a voltage generator U. (b) The amplifier is represented with a left (input) port l and a right (output) port r and the noise sources are modeled as input fields in the two noise lines a and a'

We first analyze the amplifier as depicted on figure 3a where the noise sources are represented as a current generator I and a voltage generator U. By coupling two coaxial lines denoted l and r respectively on the left port and the right port of the amplifier, one realizes a measurement model. The left line comes from a monitored electrical system so that the inward field l^{in} plays the role of the signal to be measured. Meanwhile, the right line goes to an electrical meter so that the outward field r^{out} plays the role of the meter readout. In connection with the discussions of Quantum Non

Demolition measurements [26,27], l^{out} appears as the back-action field sent back to the monitored system and r^{in} represents the fluctuations coming from the readout line. A reactive impedance Z_f acts as feedback for the amplifier.

We now present the electrical equations associated with this measurement device. We first write the characteristic relations between the voltages and currents

$$\begin{aligned} U &= U_l = U_r + Z_f I_f \\ I &= I_l + I_f \end{aligned} \tag{26}$$

Here, U_p and I_p are the voltage and current at the port p, i.e. at the end of the line $p = l$ or r, while U and I are the voltage and current noise generators associated with the operational amplifier itself (see Fig.1). Z_f is the impedance feedback. All equations are implicitly written in the frequency representation and the impedances are functions of frequency. Equations (26) take a simple form because of the limits of infinite gain, infinite input impedance and null output impedance assumed for the ideal operational amplifier. We also suppose that the fields incoming through the various ports are uncorrelated with each other as well as with amplifier noises.

As already emphasized, the output fields p^{out} obey the commutation relations (6) of free fields. To make this property explicit, we use the characteristic equations (26,12) associated with the amplifier and the lines to rewrite the output fields l^{out} and r^{out} in terms of input fields l^{in}, r^{in} and of amplifier noise sources U and I

$$\begin{aligned} l^{\text{out}} &= -l^{\text{in}} + \sqrt{\frac{2}{\hbar\,|\omega|\,R_l}}U \\ r^{\text{out}} &= -r^{\text{in}} - 2\frac{Z_f}{\sqrt{R_r R_l}}l^{\text{in}} + \sqrt{\frac{2}{\hbar\,|\omega|\,R_r}}\left(\frac{R_l + Z_f}{R_l}U - Z_f I\right) \end{aligned} \tag{27}$$

We then deduce from (27) that the voltage and current fluctuations U and I obey the following commutation relations

$$\begin{aligned} &[U\,[\omega]\,, U\,[\omega']] = [I\,[\omega]\,, I\,[\omega']] = 0 \\ &[U\,[\omega]\,, I\,[\omega']] = 2\pi\ \hbar\omega\ \delta\,(\omega + \omega') \end{aligned} \tag{28}$$

Hence, voltage and current fluctuations verify Heisenberg inequalities which determine the ultimate performance of the ideal operational amplifier used as a measurement device [25].

To push this analysis further it is worth introducing new quantities a^{in} and a'^{in} as linear combinations of the noises U and I depending on a factor R having the dimension of an impedance

$$\begin{aligned} U\,[\omega] &= \sqrt{2\hbar\,|\omega|\,R}\,\left(a^{\text{in}}\,[\omega] - a'^{\text{in}}\,[-\omega]\right) \\ I\,[\omega] &= \sqrt{\frac{2\hbar\,|\omega|}{R}}\,\left(a^{\text{in}}\,[\omega] + a'^{\text{in}}\,[-\omega]\right) \end{aligned} \tag{29}$$

For an arbitrary value of R, the quantities a^{in} and a'^{in} satisfy the free field commutation relations. In other words, the voltage and current noises associated with the amplifier may be replaced by the coupling to 2 further lines a and a' and the presence of amplification requires a conjugation of fluctuations coming in one of these two lines. This representation of the amplifier as a quantum network is depicted on figure 3b.

We may then fix the parameter R to a value R_a chosen so that the fluctuations a^{in} and a'^{in} are uncorrelated. This specific value is determined by the ratio between voltage and current noise spectra

$$R_a = \sqrt{\frac{\sigma_{UU}}{\sigma_{II}}} \tag{30}$$

The 2 noise spectra σ_{UU} and σ_{II} are defined as symmetric correlation functions. The fields a^{in} and a'^{in} are thus described by temperatures T_a and $T_{a'}$. We have assumed that these fluctuations are the same for all field quadratures, i.e. that the amplifier noises are phase-insensitive. Although this assumption is not mandatory for the forthcoming analysis, we also consider for simplicity that the specific impedance R_a is constant over the spectral domain of interest.

4 The Cold Damped Accelerometer

We come to the discussion of the ultimate performance of the cold damped capacitive accelerometer designed for fundamental physics experiments in space [16].

The central element of the capacitive accelerometer is a parallelepipedic proof mass placed inside a box. The walls of these box are electrodes distant from the mass off a hundred micrometers. The proof mass is kept at the center of the cage by an electrostatic suspension. Since a three dimensional electrostatic suspension is instable, it is necessary to use an active suspension.

In the cage reference frame, an acceleration is transformed in an inertial force acting on the proof mass. The force necessary to compensate this inertial force is measured. In fact, as in most ultrasensitive measurements, the detected signal is the error signal used to compensate the effect of the measured phenomenon.

The essential elements of the accelerometer are presented in figure 4. The proof mass and the cage form two condensators. Any mass motion unbalances the differential detection bridge and provides the error signal. In order to avoid low frequency electrical noise, the electrical circuit is polarized with an AC voltage with a frequency of a hundred kilohertz. After demodulation, this signal is used for detection and as an error signal for a servo control loop which allows to keep the mass centered in its cage.

Furthermore, the derivative of this signal provides a force proportional to the mass velocity and simulates a friction force. This active friction is

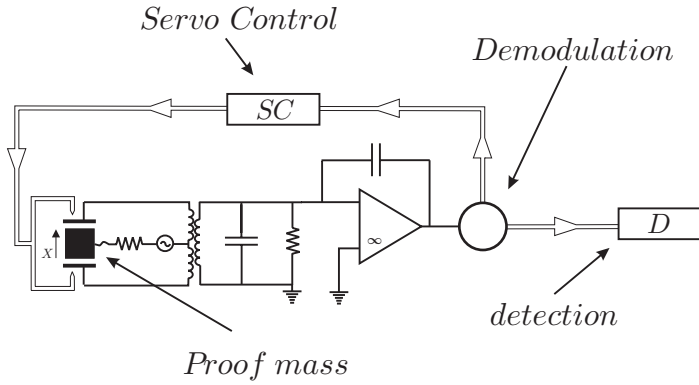

Fig. 4. Scheme of the capacitive sensor. The proof mass is placed between two electrodes formed by the inner walls of the accelerometer cage. The position dependent capacitances are polarized by an AC sinewave source which induces a mean current at frequency ω_t in the symmetrical mode. The mass displacement is read as the current induced in the antisymmetric mode. An additional capacitance is inserted to make the antisymmetric mode resonant with ω_t. The signal is detected after an ideal operational amplifier with capacitive feedback followed by a synchronous demodulation. The impedance of the detection line plays the role of a further resistance R_r. The detected signal then feds the servo loop used to keep the mass centered with respect to the cage.

called cold damping since it may be noiseless. More precisely, the effective temperature of the fluctuations of this active friction is much lower than the physical temperature of the device.

The detection is performed with the output detection signal r_1^{out}. It is a linear combination of the external force F_{ext} and of input fields in the various noise lines. We normalize this expression so that the coefficient of proportionality appearing in front of the external force F_{ext} is reduced to unity. With this normalization, we obtain a force estimator $\widehat{F}_{\text{ext}}$ which is just the sum of the true force F_{ext} to be measured and of an equivalent input force noise. In the absence of feedback, the force estimator reads [16]:

$$\widehat{F}_{\text{ext}} = F_{\text{ext}} + \sum_{\alpha} \mu_\alpha \alpha^{\text{in}} \tag{31}$$

where α^{in} denote the various input fields corresponding to the active and passive elements in the accelerometer.

When the feedback is active, the servo loop efficiently maintains the mass at its equilibrium position and the velocity is no longer affected by the external force F_{ext}. The residual motion is interpreted as the difference between the real velocity of the mass and the velocity measured by the sensor. This means that the servo loop efficiently corrects the motion of the mass except for the sensing error. However the sensitivity to external force is still present

in the correction signal. Quite remarkably, in the limit of an infinite loop gain and with the same approximations as above, the expression of the force estimator $\widehat{F}_{\text{ext}}$ is the same as in the free case [16].

The added noise spectrum Σ_{FF} is obtained as

$$\Sigma_{FF} = \sum_{\alpha} |\mu_\alpha|^2 \, \sigma^{\text{in}}_{\alpha\alpha}$$

We have evaluated the whole noise spectrum Σ_{FF} for the specific case of the instrument proposed for the μSCOPE space mission devoted to the test of the equivalence principle. Some of the main parameters of this system are listed below

$$\begin{aligned} M &= 0.27 \text{ kg} & H_m &= 1.3 \times 10^{-5} \text{ kg s}^{-1} \\ \frac{\Omega}{2\pi} &\simeq 5 \times 10^{-4} \text{ Hz} & \frac{\omega_t}{2\pi} &\simeq 10^5 \text{ Hz} \\ R_a &= 0.15 \times 10^6 \ \Omega & \Theta_a &= 1.5 \text{ K} \end{aligned} \tag{32}$$

M is the mass of the proof mass, H_m is the residual mechanical damping force, $\frac{\Omega}{2\pi}$ is the frequency of the measured mechanical motion, $\frac{\omega_t}{2\pi}$ is the operating frequency of the electrical detection circuit. R_a and Θ_a are the characteristic impedance and temperature of the amplifier.

In these conditions and at an operating temperature $\Theta_m = 300 \ K$ the added noise spectrum is dominated by the mechanical Langevin forces

$$\begin{aligned} \Sigma_{FF} &= 2 H_m k_B \Theta_m \\ &= 1.1 \times 10^{-25} \left(\text{kg m s}^{-2}\right)^2 / \text{Hz} \end{aligned} \tag{33}$$

This corresponds to a sensitivity in acceleration

$$\frac{\sqrt{\Sigma_{FF}}}{M} = 1.2 \times 10^{-12} \text{ m s}^{-2} / \sqrt{\text{Hz}} \tag{34}$$

Taking into account the integration time of the experiment, this leads to the expected instrument performance corresponding to a test accuracy of 10^{-15}.

In the present state-of-the-art instrument, the sensitivity is thus limited by the residual mechanical Langevin forces. The latter are due to the damping processes in the gold wire used to keep the proof mass at zero voltage [3]. With such a configuration, the detection noise is not a limiting factor. This is a remarkable result in a situation where the effective damping induced through the servo loop is much more efficient than the passive mechanical damping. This confirms the considerable interest of the cold damping technique for high sensitivity measurement devices.

Future fundamental physics missions in space will require even better sensitivities. To this aim, the wire will be removed and the charge of the test mass will be controlled by other means, for example UV photoemission. The mechanical Langevin noise will no longer be a limitation so that the analysis

of the ultimate detection noise will become crucial for the optimization of the instrument performance. This also means that the electromechanical design configuration will have to be reoptimized taking into account the various noise sources associated with detection [16].

References

1. A. Bernard and P. Touboul, *The GRADIO accelerometer: design and development status*, Proc. ESA-NASA Workshop on the Solid Earth Mission ARISTOTELES, Anacapri, Italy (1991).
2. P. Touboul et al., *Continuation of the GRADIO accelerometer predevelopment*, ONERA Final Report 51/6114PY, 62/6114PY ESTEC Contract (1992, 1993).
3. E. Willemenot, *Pendule de torsion à suspension électrostatique, très hautes résolutions des accéléromètres spatiaux pour la physique fondamentale*, Thèse de Doctorat de l'université Paris 11 (1997).
4. A. Einstein, *Annalen der Physik* **17** (1905) 549.
5. H. Nyquist, *Phys. Rev.* **32** (1928) 110.
6. H.B. Callen and T.A. Welton, *Phys. Rev.* **83** (1951) 34.
7. L. Landau and E.M. Lifshitz, *"Course of Theoretical Physics: Statistical Physics Part 1"* (Butterworth-Heinemann, 1980) ch. 12.
8. R. Kubo, *Rep. Prog. Phys.* **29** (1966) 255.
9. E.M. Lifshitz and L.P. Pitaevskii, *"Landau and Lifshitz, Course of Theoretical Physics, Statistical Physics Part 2"* (Butterworth-Heinemann, 1980) ch. VIII.
10. M. Planck M. 1900 *Verh. Deutsch. Phys. Ges.* **13** 138 (1911); W. Nernst *ibid.* **18** 83 (1916)
11. J. Meixner, *J. Math. Phys.* **4** (1963) 154.
12. B. Yurke and J.S. Denker, *Phys. Rev.* A **29** (1984) 1419.
13. C.W. Gardiner, *IBM J. Res. Dev.* **32** (1988) 127.
14. J-M. Courty and S. Reynaud, *Phys. Rev.* **A 46** (1992) 2766.
15. F. Grassia, *Fluctuations quantiques et thermiques dans les transducteurs électromécaniques*, Thèse de Doctorat de l'Université Pierre et Marie Curie (1998).
16. F. Grassia, J-M. Courty, S. Reynaud and P. Touboul , preprint xxx quant-ph/9904073
17. M. Feldmann, *Théorie des réseaux et systèmes linéaires,* (Eyrolles 1986)
18. H. Heffner, *Proc IRE* **50** (1962) 1604.
19. H.A. Haus and J.A. Mullen, *Phys. Rev.* **128** (1962) 2407.
20. J.P. Gordon, L.R. Walker and W.H. Louisell, *Phys. Rev.* **130** (1963) 806.
21. C.M. Caves, *Phys. Rev.* **D26** (1982) 1817.
22. R. Loudon and T.J. Shephered, *Optica Acta* **31** (1984) 1243.
23. J.P. Gordon, *Proc. IRE* (1962) 1898.
24. H. Takahasi, in *"Advances in Communication Systems"* ed. A.V. Balakrishnan (Academic, 1965) 227.
25. J-M. Courty, F. Grassia and S. Reynaud, *Europhys. Lett*, **46** (1), pp. 31-37 (1999) [preprint quant-ph/9811062].
26. V.B. Braginsky and F.Ya. Khalili, "*Quantum Measurement*" (Cambridge University Press, 1992).
27. P. Grangier, J.M. Courty and S. Reynaud, *Opt. Comm.* **89** (1992) 99.

Dipole at $\nu = 1$

V. Pasquier *

CEA/Saclay, Orme des Merisiers,
Bât. 774, F-91191 Gif-sur-Yvette Cedex, FRANCE

Abstract. I consider the problem of Bosonic particles interacting repulsively in a strong magnetic field at the filling factor $\nu = 1$. We project the system in the lowest Landau level and map the dynamics into an interacting Fermion system. We study the resulting Hamiltonian in the Hartree–Fock approximation in the case of a δ repulsive potential. The physical picture which emerges is in agreement with the proposal of N. Read that the composite Fermions behave as a gas of dipoles. We argue that the consequence of this is that the composite Fermions interact with screened short range interactions. We develop a Landau theory which we also expect to describe the physical $\nu = 1/2$ Fermionic state. The form factor, the effective mass and the conductivity are analyzed in this model [1] .

1 Introduction

There has been recently a renewed interest in the quantum Hall effect when the filling factor is a fraction with an even denominator. Willets and his collaborators [3] have observed an anomalous behavior in the surface acoustic wave propagations near $\nu = 1/2$ and $\nu = 1/4$. A remarkable outcome of their experiments is that they probe a longitudinal conductivity $\sigma_{xx}(q, \omega)$ increasing linearly with the wave vector q. Halperin, Lee and Read [4] have suggested that the system exhibits a Fermi liquid behavior at this particular value. They have developed a formalism based on the Chern–Simon theory to explain the formation of a Fermi surface which provides an explanation for the experimental observations. Subsequently, several studies have developed and improved the predictions of the Chern– Simon theory [4],[8],[9],[10], [11]. Another approach followed by Rezayi and Read [13] and Haldane et al. [18] consists in obtaining trial wave functions which enable to study numerically the properties of the system at this filling factor. In these studies the cyclotron frequency is supposed to be sufficiently large so that the only relevant excitations are confined to the lowest Landau level. The trial wave functions can be compared to the exact ground state and the overlap between the two turns

* pasquier@spht.saclay.cea.fr

[1] This lecture is issued from an unpublished paper on Bosons at filling factor 1 ("Composite fermions and confinement" March 1996). Although some results are outdated, some aspects considered here have not been discussed in the more recent literature. I give a list of more recent references on the subject for the interested reader [24-28].

M. Planat (Ed.): LNP 550, pp. 84–106, 2000.
© Springer-Verlag Berlin Heidelberg 2000

out to be extremely good [13]. In these studies the effective mass m^* which defines the Fermi velocity is generated dynamically by the interactions. This paper is an attempt to introduce a model which accounts for the success of these trial wave functions and enables to compute the physical properties of the system in the infinite cyclotron energy limit. We have tried to keep as close as possible to the microscopic reality by using a theoretical description which coincides exactly with what is simulated in the numerical studies. This is a different approach than the Chern–Simon field theory which mixes the Landau levels and requires the mass of the electron as a parameter of the theory. We do not address the physical problem of electrons in a magnetic field at $\nu = 1/2$. Instead we have considered the problem of Bosonic particles interacting repulsively in a magnetic field at a filling factor $\nu = 1$. Although it may at first look quite different, it is important to stress that the problem of formation of a Fermi sea is essentially the same as in the $\nu = 1/2$ case. If one applies the analyses of the composite Fermions or the Chern– Simon approach to such a system, one is essentially led to the same picture of Fermi sea formation as in the $\nu = 1/2$ case. We have also verified this hypotheses by performing a numerical simulation for a small system on a sphere (see Fig.1). The main reason why this is simpler theoretical problem to look at than the $\nu = 1/2$ physical problem is that the wave function one needs to start from is the Slater determinant of the lowest Landau level one body wave functions which is a much simpler object to consider than the Laughlin $\nu = 2$ wave function which one would have to use in the $\nu = 1/2$ case. We shall present an exact mapping of the Bosonic problem into a Fermionic one which has the advantage of making the appearance of a Fermi liquid natural. Let us briefly review some important features of a 2D electron gas in a strong magnetic field. Classically, due to the magnetic field each electron moves on a circular orbit at the cyclotron frequency $\omega_c = eB/m_e c$. Quantum mechanically in the absence of interactions the one particle energies are quantized ($E_n = (n + 1/2)\,\hbar\omega_c$) and the degeneracy of each energy state is obtained by dividing the area of the system by a quantum of area a_0 equal to $2\pi\hbar c/eB$. The physical properties are characterized by the filling factor $\nu = a_0\rho_e$ where ρ_e is the electron density. The amount of magnetic flux through a_0 defines the flux unit $\Phi_0 = 2\pi\hbar c/e$. One usually defines the magnetic length l equal to $\sqrt{\hbar c/eB}$.

At the filling factors where the system exhibits a quantum Hall effect, the chemical potential has a discontinuous slope. This is the case when ν is an integer where it results from the fact that ν Landau levels are fully occupied and the Pauli principle forces the next electron to be in a higher level. When the filling factor is fractional, the interactions between the electrons play a crucial role and the effect requires a much more involved explanation. The special filling factors $\nu = 1/(2p + 1)$ are well understood from the analysis of Laughlin [1] and a hierarchical extension to fractions with an odd denominator was proposed by Haldane [18]. Jain [2] has produced a mean field

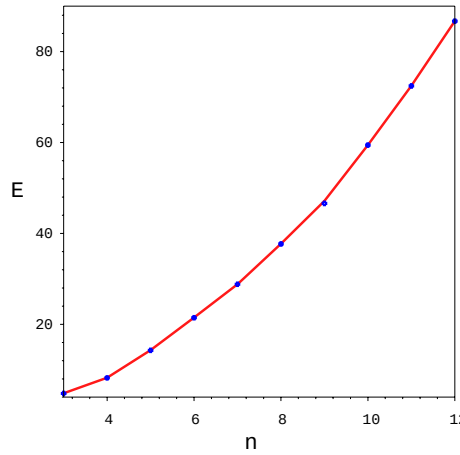

Fig. 1. The ground state energy of Bosons on a sphere at filling factor $\nu = 1$. The Bosons interact through a δ repulsive potential. We plot the energy as a function of the number of bosons n $(1 \leq n \leq 12)$. As a function of n, the energy is roughly linear by pieces with a slope breaking at each perfect square $(n = 4, 9)$. This indicates that the bosons have a similar ground state energy as a system of n free fermions on a sphere without without magnetic field.

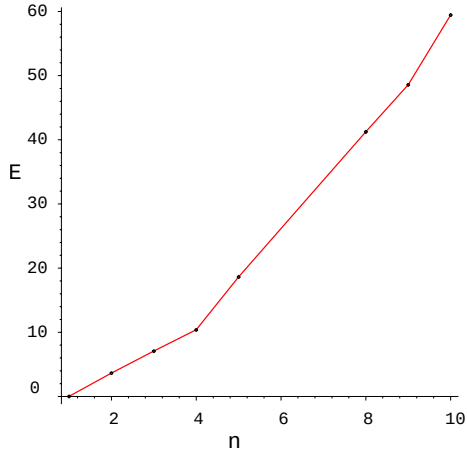

Fig. 2. We consider a system of 10 particles interacting on a sphere with a magnetic field at $\nu = 1$. the interaction potential is a δ function as in Fig.1.These particles are splitted into n bosons and $10 - n$ fermions and we plot the ground state energy as a function of n. We see that even more convincingly as in figure 1 the energy behaves a if there were n free fermions without magnetic field. The interpretation is that the boson binds to the fermionic hole to form a quasi-free bound state. The striking fact is that this feature remains true even where there is no fermion left! $(n = 10)$.

argument to predict the dominant series of fractions $\nu = p/(2p \pm 1)$ in good agreement with experiments. The mean field interpretation of Jain's trial wave functions requires attaching flux tubes to the electron which has given rise to the denomination of composite fermion to describe such a dressed electron. Halperin Lee and Read use the composite Fermions arguments to motivate the formation of a Fermi sea in the $\nu = 1/2$ case (which is the $p = \infty$ limit of Jains hierarchy). They attach two magnetic fluxes to an electron in order to cancel the magnetic field seen by the electron in the mean field approximation. This flux attachment does not modify the statistics of the electrons and if one ignores fluctuations one is led to a system of spinless Fermions in a zero magnetic field. In the case of Bosons at $\nu = 1$, we can proceed similarly by attaching one flux unit to each particle so as to cancel the exterior magnetic field. In this process the statistics is changed from Bosons to Fermions and a Fermi liquid is expected to form.

Read has interpreted the fluxes attached to the electron as physical vortices bound to it [5]. We believe that his proposal differs considerably from the mean field interpretation for the following reason. The mean field treats the composite electron as a charged particle which couples minimally to the electro-magnetic field. In Reads picture, the vortices carry a charge equal to minus one half of that of the electron so that the bound state must be viewed as a neutral particle which propagates in a constant charge background. In this case the response to an electric field depends on the internal structure of the composite object. Bakaran has independently reached this conclusion using a semi-classical analyses [7]. We are led to this picture in the $\nu = 1$ case. The essential simplification is that there is a single vortex coupled to the Boson, this vortex is a Fermion carrying the opposite charge as the Boson and we can use a second quantized formalism to analyse the model. The bound state is then a dipole whose structure was discovered long ago [16],[15],[14].

Our approach is mainly motivated by the trial wave functions of [13],[18]. The Laughlin is the wave function is the starting point to describe the low energy physics but it gives the particles the opposite statistics to the one we are interested in. In the $\nu = 1$ case for example, the exact ground state wave function for distinguishable particles interacting with a δ repulsive potential in a disc geometry is given by $\Psi(z_i) = \prod_{i<j}(z_i - z_j)\exp(-\sum_i z_i\overline{z}_i/4)$. This wave function is antisymmetric while we are interested in a Bosonic statistics. The trial wave functions are obtained by multiplying this wave function by a Slater determinant of plane waves so as to give the wave function the correct statistics. The product is then projected into the lowest Landau level. The effect of the projection is to replace the coordinate which appear in the plane waves by the guiding center coordinates [18] which displace the particles minimally from their original position. Here we advocate that the charge fluctuations induced by this displacement are the fundamental excitations.

To show this we want to use the Laughlin wave function as a vacuum and develop a low density formalism to construct the excitations one by one above

this ground state. We have looked at a model which enables us to define a Fermi momentum varying continuously between zero and the physical value $k_f = \sqrt{2}/l$. We consider a system consisting of Bosons at a filling factor ν_1 and Fermions at a filling factor ν_2 such that the total filling factor $\nu = \nu_1 + \nu_2$ is kept equal to one. We also keep the interactions equal between all the particles. We argue that the model should behave as a Fermi liquid with a Fermi momentum proportional to ν_2 regardless of the ratio ν_2/ν_1. Let us first explain why this behavior is predicted by the mean field argument. To show this, we make an analogy with the Double-Layer Quantum Hall system studied in [21] (and references therein). In a thought experiment, we consider a system consisting of a pair of 2D layers exactly on top of each other. The particles in the top layer obey the Fermi statistics and the particles in the lower layer obey the Bose statistics. Because the statistics of the particles differ from one layer to the other we are in the situation where there is no tunneling between the two layers. Although we are ultimately interested in the case where the top layer is empty we consider intermediate systems where the top layer has a filling factor ν_1 and the lower layer a filling factor ν_2 such that the total filling factor $\nu_1 + \nu_2$ is kept equal to one. The mean field argument to account for the formation of a Fermi surface proceeds as follow: Each particle regardless of which layer it belongs to is attached a unit of flux so that in the mean field approximation where the magnetic field is replaced by its spatial average the particles see no magnetic field. The mutual particle statistics needs to be modified by this flux attachment as follow. Two particles in different layers are distinguishable so we need only consider the relative statistics of the particles in the same layer. Then we must impose that if two particles of the same layer are interchanged in a clockwise manner around a curve which encloses no other particles, the wave function is changed by a factor $e^{i\pi(1+\varepsilon)}$ where $\varepsilon = 1$ or 0 according to weather the particles are Fermions (in the first layer) or Bosons to Fermion (in the second layer). Thus the attachment of a flux exchanges the statistics from Fermion to Boson in the first layer and from Boson to Fermion in the second layer. Now the particles in the first layer behave as a Bose superfluid and the particles in the second layer as a Fermi liquid so that in the low energy limit the system should behave the Fermi momentum $k_f = \sqrt{2}\nu_2 l$. To convince the reader of the validity of the conclusion we note that a trivial modification of the argument predicts that the system exhibits a quantum Hall behavior in the physical case where the two layers are filled with electrons. In this situation the quantum Hall effect is experimentally observed although the gap responsible for it is due to a nontrivial collective effect [21]. In order to test this hypotheses we have also performed a numerical evaluation of the ground state energies for a mixed system consisting of N_1 Bosons and N_2 Fermions on a sphere keeping $N_1 + N_2 = N$ fixed. The sphere has $N - 1$ quantum fluxes and the interaction between the particles is a delta function

interaction. The gross features of the spectrum are those of a system of N_2 free Fermions on a sphere with no magnetic field (see Fig. 2).

Although the preceding argument is qualitative it can be made rigorous when ν_2 is small. In this low density limit, the particles in the second Layer bind to the holes of the first layer to form exciton pairs. The main difference with the standard double layer system studied in [21] is that the excitons are Fermion. This paper is an attempt to construct a model for the Fermi-sea formation which uses these excitons as fundamental quasiparticle. In doing so we follow a suggestion made long ago by Lerner and Lozovik in a different context [15]. We obtain an exact representation of the dynamics in terms of these excitations still valid in the case where there are only Bosons (the first layer is empty). One of the outcome of this interpolation is that it makes a link between the formalism of Kallin and Halperin which is relevant at low densities ($\nu_2 \ll 1$) and the wave functions of Rezayi and Read which describe the other limit.

The next section is devoted to present the microscopic model and analyze its phenomenological consequences.

2 The Microscopic Model

We consider N particles of identical charge interacting with a repulsive force in a domain of area Ω thread by a magnetic field B. is chosen so that the flux per unit area is equal to one. We take units where $\hbar = 1$ and the magnetic length $l = \sqrt{\hbar c/eB} = 1$ in which case $\Omega = 2\pi N$. We assume that the cyclotron frequency ω_c is large compared to the interaction so that the dynamic can be restricted to the Lowest Landau Level. The one body Hamiltonian has N degenerate eigenstates, thus in the case where the particles are Fermions the only accessible state is given by the Slater determinant of the one body wave functions. This state will define the vacuum of the theory. We now discuss the case where there are two sets of particles obeying distinct statistics. The first set contains N_1 Fermions and the second set contains N_2 Bosons. We keep the sum $N_1 + N_2 = N$ fixed so that the filling factor remains equal to one. We also keep the interaction equal between all the particles. It is instructive to first look at the case of 1 Boson interacting with $N - 1$ Fermions. By performing a particle hole transformation on the Fermions, we can equivalently regard this as a Boson interacting with a hole. This problem has been studied by Kallin and Halperin [14]. A surprising outcome is that the wave functions which describe this two body state are independent of the potential and are given by the ground state eigen-functions of the free Hamiltonian [16],[15],[14]. Let us recover this result using the Lancau Gauge ($A_x = 0,\ A_y = x$ in units where the magnetic length is equal to one). We consider a cylinder of length L_y in the periodic y direction and confine the particles on a segment of length L_x in the x direction. The one-body Hamil-

tonian is given by:

$$2H = p_x^2 + (p_y - x)^2 \tag{1}$$

and the corresponding ground state wave functions are:

$$\langle \overrightarrow{x} | s \rangle = \sqrt{1/\pi L_y} e^{isy - (x-s)^2/2} \tag{2}$$

$s = 2\pi n/L_y$ denotes the momentum in the y direction and one must have $0 \leq s \leq L_x$ in the x direction. This Hamiltonian commutes with the two guiding center coordinates $R_y = p_x - y$ and $R_x = p_y$ which do not commute with each other $[R_x, R_y] = i$. In the particle hole case however the situation is quite different. The Free Hamiltonian which describes the particle-hole pair is the sum of two one body Hamiltonians where we change the sign of coupling to the potential vector for the hole:

$$H = p_{x_1}^2 + (p_{y_1} + x_1)^2 + p_{x_2}^2 + (p_{y_2} + x_2)^2 \tag{3}$$

Since the particles have exactly opposite charges, their guiding center coordinates $p_x = R_{y_1} + R_{y_2} = p_{x_1} + p_{x_2} - (y_1 - y_2)$ and $p_y = R_{x_1} + R_{x_2} = p_{y_1} + p_{y_2}$ now commute with each other and can be diagonalized simultaneously with H. The wave functions which diagonalize this generalized momentum in the lowest Landau level are given by:

$$\begin{aligned} \langle \overrightarrow{x}_1, \overrightarrow{x}_2 | \overrightarrow{p} \rangle = {} & e^{i(p_x + y_1 - y_2)(x_1 + x_2)/2} e^{ip_y(y_1 + y_2)/2} \\ & e^{-(x_1 - x_2 - p_y)^2/4} e^{-(y_1 - y_2 + p_x)^2/4} \end{aligned} \tag{4}$$

They describe a dipole propagating freely in such a way that its dipole vector given by $(p_y, -p_x)$ is perpendicular to the momentum (p_x, p_y). When we switch the interaction on, the potential commutes with the momentum p_x, p_y so that these wave functions are eigenstates of the projected potential and thus of the total projected Hamiltonian. The degeneracy is lifted and the energy of a pair depends on its momentum. When the interactions is repulsive these wave functions describe bound states which we shall denote excitons following [15]. Their energy $\varepsilon_0(\overrightarrow{p})$ will be computed later in the text.

The idea which motivates the following discussion is to subdivide the particles and the holes into pairs so as to include the interaction in each pair in the one body Hamiltonian H_0 and treat residual interaction between the different pairs as a perturbation.

2.1 Second quantized formalism

In order to study the system where several Bosons interact with the same number of holes we introduce a second quantization formalism by defining Bosonic (a_s^+) and Fermionic (b_s^+) creation operators which create the one body states in the Lowest Landau level. They obey the standard commutation

relations $\left[a_s, a_{s'}^+\right] = \delta_{s,s'}$, $\left\{b_s, b_{s'}^+\right\} = \delta_{s,s'}$. The vacuum $|0\rangle$ is the filled Landau level state which is characterized by $a_s\,|0\rangle = b_s^+\,|0\rangle = 0$. We also define the fields which create a Boson (a Fermion) at position x in the Lowest Landau Level:

$$\Phi_b^+\,(\overrightarrow{x}) = \sum_s \langle \overrightarrow{x}|s\rangle\, a_s^+$$
$$\Phi_f^+\,(\overrightarrow{x}) = \sum_s \langle \overrightarrow{x}|s\rangle\, b_s^+ \tag{5}$$

The second quantized expression for the exciton with momentum $\overrightarrow{p}$ is determined by the equation

$$\left\langle 0|\Phi_b\,(\overrightarrow{x}_1)\,\Phi_f^+\,(\overrightarrow{x}_2)\,|\,\overrightarrow{p}\right\rangle = \langle \overrightarrow{x}_1,\, \overrightarrow{x}_2|\,\overrightarrow{p}\rangle \tag{6}$$

The wave function on the right hand side is defined in (40) so, in this way we determine the state $|\,\overrightarrow{p}\rangle$ which we write in the form $|\,\overrightarrow{p}\rangle = A_{\overrightarrow{p}}^+\,|0\rangle$ where $A_{\overrightarrow{p}}^+$ is then given by:

$$A_{\overrightarrow{p}}^+ = e^{ip_x p_y/2} 1/\sqrt{N} \sum_s e^{-ip_x s} a_s^+ b_{s-p_y} \tag{7}$$

Note that the definition of $A_{\overrightarrow{p}}^+$ is not unique.

Another way to obtain the operators $A_{\overrightarrow{k}}$ consists in defining the field $A^+\,(\overrightarrow{x})$ which creates the exciton (destroys a fermion and creates a boson) at position $\overrightarrow{x}$:

$$A^+\,(\overrightarrow{x}) = 1/\sqrt{N}\Phi_b^+\,(\overrightarrow{x})\,\Phi_f\,(\overrightarrow{x})$$

This definition is similar to Read's [6] way to obtain operators that create quasi electrons and holes. One can then verify that $A_{\overrightarrow{k}}^+$ are proportional to the Fourier modes of $A^+\,(x)$:

$$A_{\overrightarrow{k}}^+ = e^{k^2/4} \int e^{-i\,\overrightarrow{k}\,\overrightarrow{x}} A^+\,(x)\,d^2x \tag{8}$$

In a similar way we also define the densities of Bosons and fermion operators:

$$\rho^b\,(\overrightarrow{x}) = \Phi_b^+\,(\overrightarrow{x})\,\Phi_b\,(\overrightarrow{x})$$
$$\rho^f\,(\overrightarrow{x}) = \Phi_f^+\,(\overrightarrow{x})\,\Phi_f\,(\overrightarrow{x}) \tag{9}$$

and their Fourier transforms as in (8). One ends up with similar expressions as in (7):

$$\rho_{\overrightarrow{p}}^b = e^{i\,\overrightarrow{p}_x\,\overrightarrow{p}_y/2} \sum_s e^{-ip_x s} a_s^+ a_{s-p_y}$$
$$\rho_{\overrightarrow{p}}^f = e^{i\,\overrightarrow{p}_x\,\overrightarrow{p}_y/2} \sum_s e^{-ip_x s} b_s^+ b_{s-p_y} \tag{10}$$

The commutations between these fields lead to a generalization of the magnetic translation algebra [17]. The relations between the ρ themselves are given by:

$$\begin{aligned}
\left[\rho^b_{\overrightarrow{p}}, \rho^b_{\overrightarrow{q}}\right] &= \left(e^{-i\overrightarrow{p}\times\overrightarrow{q}/2} - e^{i\overrightarrow{p}\times\overrightarrow{q}/2}\right)\rho^b_{\overrightarrow{p}+\overrightarrow{q}} \\
\left[\rho^f_{\overrightarrow{p}}, \rho^f_{\overrightarrow{q}}\right] &= -\left(e^{i\overrightarrow{p}\times\overrightarrow{q}/2} - e^{i\overrightarrow{p}\times\overrightarrow{q}/2}\right)\rho^f_{\overrightarrow{p}+\overrightarrow{q}} \\
\left[\rho^b_{\overrightarrow{p}}, \rho^f_{\overrightarrow{q}}\right] &= 0
\end{aligned} \tag{11}$$

where $\overrightarrow{p}\times\overrightarrow{q} = p_x q_y - p_y q_x$. The relation between the ρ and the A^+ are:

$$\begin{aligned}
\left[\rho^b_{\overrightarrow{p}}, A^+_{\overrightarrow{q}}\right] &= e^{-i\overrightarrow{p}\times\overrightarrow{q}/2}A^+_{\overrightarrow{p}+\overrightarrow{q}} \\
\left[\rho^f_{\overrightarrow{p}}, A^+_{\overrightarrow{q}}\right] &= e^{-i\overrightarrow{p}\times\overrightarrow{q}/2}A^+_{\overrightarrow{p}+\overrightarrow{q}}
\end{aligned} \tag{12}$$

Finally, the relations between $A_{\overrightarrow{p}}$ and $A^+_{\overrightarrow{q}}$ is:

$$\left\{A_{\overrightarrow{p}}, A^+_{\overrightarrow{q}}\right\} = 1/N\left(e^{-i\overrightarrow{p}\times\overrightarrow{q}/2}\rho^b_{\overrightarrow{q}-\overrightarrow{p}} + e^{i\overrightarrow{p}\times\overrightarrow{q}/2}\rho^f_{\overrightarrow{q}-\overrightarrow{p}}\right) \tag{13}$$

Instead of the fermion density operator, it will be more convenient to use its normal ordered form: $\rho^h_{\overrightarrow{p}} = e^{ip_x p_y/2}\sum_s e^{-ip_x s}b_{s-p_y}b^+_s$. One has $\rho^f_{\overrightarrow{p}} + \rho^h_{\overrightarrow{p}} = \delta_{\overrightarrow{p},\overrightarrow{0}}N$. We shall see that the physical observables can be expressed in terms of the total density operator: $\rho^t_{\overrightarrow{p}} = \rho^b_{\overrightarrow{p}} - \rho^h_{\overrightarrow{p}}$ which by construction vanishes at zero momentum: $\rho^t_{\overrightarrow{0}} = 0$. The Fourier modes $\rho^t_{\overrightarrow{p}}$ also obey the relations (11), in addition they obey the following relations with $A^+_{\overrightarrow{q}}$:

$$\left[\rho^t_{\overrightarrow{p}}, A^+_{\overrightarrow{q}}\right] = \left(e^{-i\overrightarrow{p}\times\overrightarrow{q}/2} - e^{i\overrightarrow{p}\times\overrightarrow{q}/2}\right)A^+_{\overrightarrow{p}+\overrightarrow{q}} \tag{14}$$

When we express the commutator between $A_{\overrightarrow{p}}$ and $A^+_{\overrightarrow{q}}$ in terms of ρ_h instead of ρ_f the commutator takes the more natural form:

$$\left\{A_{\overrightarrow{p}}, A^+_{\overrightarrow{q}}\right\} = \delta_{\overrightarrow{p},\overrightarrow{q}} + 1/N\left(e^{-i\overrightarrow{p}\times\overrightarrow{q}/2}\rho^b_{\overrightarrow{p}-\overrightarrow{q}} - e^{i\overrightarrow{p}\times\overrightarrow{q}/2}\rho^h_{\overrightarrow{p}-\overrightarrow{q}}\right) \tag{15}$$

Assuming that the coefficient of $1/N$ is an operator of order one, up to a $1/N$ correction this commutator is equal to the usual commutator between creation and annihilation operators. This will be our main approximation in the following.

Remark 1. The above algebra admits a compact form relevant to study the finite size system on a torus. In the appendix B we show that in this case the "Slater determinants":

$$A^+\left(\overrightarrow{p}_1\right)A^+\left(\overrightarrow{p}_2\right)...A^+\left(\overrightarrow{p}_N\right)|0\rangle \tag{16}$$

coincide with a class of trial wave functions used by Haldane et al. [18] to study the system on a finite geometry in the $\nu = 1$ case.

Remark 2. The reader familiar with the works on Double Layer systems [21] will notice a resemblance between this algebra and the algebra used in [21]. It is not surprising in view of the analogy used in the introduction. In the double layer context we can replace the operators a_s and a_s^+ in (5) by operators which anticommute instead. Since both layers carry particles which obey a Fermionic statistics, the only difference with this case is that the excitons are Bosons. If we use the same formulas (8, 9) to define the density operators and the exciton creation operator they obey the algebra considered in [21].

Remark 3. The zero mode operators $\rho^b_{\vec{0}}, \rho^f_{\vec{0}}, A^+_{\vec{0}}, A_{\vec{0}}$, form a supersymmetric $su\,(1|1)$ subalgebra which is the symmetry of the present problem analogous to the $su(2)$ symmetry in the two Layer system. A simple consequence of this will be that the two states $|\Psi\rangle$ and $A^+_{\vec{0}}\,|\Psi\rangle$ will be degenerate eigenstates of the Hamiltonian when we consider the dynamics of the model in the next section.

Holstein–Primakov representation There is a natural representation of the operators $\rho^b_{\vec{p}}\rho^h_{\vec{p}}$ and $A^+_{\vec{p}}$ in terms of creation and annihilation operators obeying Fermionic commutation relations: $\left\{c_{\vec{p}}, c^+_{\vec{q}}\right\} = \delta_{\vec{p},\vec{q}}$. It is given by:

$$\begin{aligned}
\rho^b_{\vec{p}} &= \sum_r e^{-i\,\vec{p}\times\vec{r}/2} c^+_{\vec{p}+\vec{r}} c_{\vec{r}} \\
\rho^h_{\vec{p}} &= \sum_r e^{i\,\vec{p}\times\vec{r}/2} c^+_{\vec{p}+\vec{r}} c_{\vec{r}} \\
A^+_{\vec{p}} &= c^+_{\vec{p}}
\end{aligned} \tag{17}$$

It is easy to verify that the relations (11, 12) are satisfied by this representation. To satisfy (13) one must modify the expression of $A^+_{\vec{p}}$ in terms of $c^+_{\vec{p}}$ as explained in the appendix A. Unfortunately one ends up with a non linear representation of $A^+_{\vec{p}}$ which we were not able to handle in practice. Instead we have made the assumption that the $1/N$ correction can be neglected in the commutator (15) and replace the field $A^+_{\vec{p}}$ by $c^+_{\vec{p}}$ in the following.

This approximation is similar to the RPA approximation in the sense that we replace the right hand side of (13) by its expectation value in the ground state. Since the only place where the commutator of A and A^+ appears is in the scalar product of the states, this amounts to work in a Hilbert space with a different norm than the physical one. This is essentially where this work departs from the numerical calculations which evaluate the overlap between the states in an exact way.

One can verify on a finite size system (see appendix B) that the c^+ basis of states is overcomplete. In the low density limit, the Hilbert space with n creation operators c^+ is $n!$ times bigger than the true Hilbert space. This factor originates from the fact than all the states obtained by permuting the Bosons (or the Fermions) describe the same physical state and as long as the

excitons are separated by distances large compared to their size there is no overlap between these Hilbert spaces.

The equivalence with the Fermionic Hamiltonian introduced here is similar in spirit to the representation of the Heisenberg spin chain in terms of a Bosonic Hamiltonian [19]. This would be the case here if we were studying the double Layer system. Here, since the particles of the two layers obey opposite statistics it is a Fermionic representation. We call this a generalized Holstein Primakov transformation because the expression of $A^+_{\overrightarrow{k}}$ in terms of $c^+_{\overrightarrow{k}}$ is similar to the spin operator expression of $S^+_{\overrightarrow{k}}$ in the spin chain case as shown in the appendix B.

2.2 Dynamics of the model

The Hamiltonian which governs the dynamics of the model is given by the projection of the interaction potential energy on the lowest Landau level. For obvious symmetry reasons, we consider the case where the two layers are exactly on top of each other and this we take the interaction energy to be the same between the different type of particles Bosonic or Fermionic:

$$H = \int d^2x\, d^2y\, 1/2V\left(\overrightarrow{x} - \overrightarrow{y}\right)\rho^t\left(\overrightarrow{x}\right)\rho^t\left(\overrightarrow{y}\right) \tag{18}$$

Or in Fourier modes as:

$$H = 1/2\Omega \sum_{\overrightarrow{q}} \tilde{V}\left(\overrightarrow{q}\right)\rho^t\left(\overrightarrow{q}\right)\rho^t\left(-\overrightarrow{q}\right) \tag{19}$$

where $\tilde{V}\left(\overrightarrow{q}\right) = e^{-q^2/2}\int d^2x V\left(\overrightarrow{x}\right)e^{i\overrightarrow{q}\,\overrightarrow{x}}$.

If we use the Fermionic representation of ρ^t, it can be expressed in a more conventional form of an interacting Fermion Hamiltonian:

$$\begin{aligned} H \;=\; & \sum_{\overrightarrow{k}} \varepsilon_0\left(\overrightarrow{k}\right) c^+_{\overrightarrow{k}} c_{\overrightarrow{k}} + 1/2\Omega \sum_{\overrightarrow{q},\overrightarrow{k},\overrightarrow{k}'} \tilde{V}\left(\overrightarrow{q}\right) \times [\\ & 4\,\sin\left(\overrightarrow{k} \times \overrightarrow{q}/2\right)\sin\left(\overrightarrow{k}' \times \overrightarrow{q}/2\right) c^+_{\overrightarrow{k}+\overrightarrow{q}} c^+_{\overrightarrow{k}',-\overrightarrow{q}} c_{\overrightarrow{k}',} c_{\overrightarrow{k}} \;] \end{aligned} \tag{20}$$

$\varepsilon_0\left(\overrightarrow{k}\right) = 1/2\Omega \sum_{\overrightarrow{q}} \tilde{V}\left(\overrightarrow{q}\right) 4\sin^2\left(\overrightarrow{k} \times \overrightarrow{q}/2\right)$ is the dipole dispersion relation introduced earlier in the text.

The long range part of the interaction ($q \ll 1$) has the interpretation of a dipole dipole potential where $\hat{z} \times \overrightarrow{k}, \hat{z} \times \overrightarrow{k}'$ define the dipole vectors. Note that since the dipole vector depends on the momentum of the particle the Hamiltonian has no Gallilean invariance. This is not be surprising since a Galilean transformation creates of an electric field perpendicular to the speed in the presence of a magnetic field.

We can also rewrite the interaction in the form of a current-current interaction:

$$H_{int} = 1/2\Omega \sum_{\overrightarrow{q}} q^2 \tilde{V}\left(\overrightarrow{q}\right) J_t\left(\overrightarrow{q}\right) J_t\left(-\overrightarrow{q}\right)$$

where $J_t\left(\overrightarrow{q}\right)$ is the transverse current $J_t\left(\overrightarrow{q}\right) = \sum_k \left(\overrightarrow{q} \times \overrightarrow{k}\right) c^+_{\overrightarrow{k}+\overrightarrow{q}} c_{\overrightarrow{k}}$. So, if $V(q)$ behaves as q^{-x} at small q, the coefficient of the current-current potential behaves as q^{2-x} and we expect no infrared singularity in the physical range $0 \leq x \leq 2$. This absence of divergent contribution at small momentum transfer is due to the screening of the dipole interaction. In this range of values the system becomes essentially equivalent to a neutral Fermi liquid with a short range interaction potential. In the following, we have restricted ourself to the case where $V(x) = \delta^2(x)$ because the angular integrations can be carried out analytically and in this model we do not expect the physics to depend very much on the range of the potential for $x \leq 2$.

The ground state energy E_0 can be evaluated in a Hartree–Fock approximation. Denoting $n(p)$ the ground state distribution
($n(p) = 1$ if $p < k_f, n(p) = 0$ if $p > k_f$) one obtains:

$$E_0 = 1/2\Omega \sum_{\overrightarrow{p},\overrightarrow{q}} \tilde{V}\left(\overrightarrow{q}\right) 4\sin^2\left(\overrightarrow{p} \times \overrightarrow{q}/2\right) n\left(\overrightarrow{p}\right)\left(1 - n\left(\overrightarrow{p} - \overrightarrow{q}\right)\right) \quad (21)$$

Using this expression we can determinate the appropriate Landau parameters in this approximation [22]:

$$\varepsilon(p) = 1/2\Omega \sum_{\overrightarrow{q}} \tilde{V}\left(\overrightarrow{q}\right) 4\sin^2\left(\overrightarrow{p} \times \overrightarrow{q}/2\right) \times \left(1 - n\left(\overrightarrow{p} - \overrightarrow{q}\right) - n\left(\overrightarrow{p} + \overrightarrow{q}\right)\right) \quad (22)$$

$$f\left(\overrightarrow{p}, \overrightarrow{q}\right) = -1/\Omega \tilde{V}\left(\overrightarrow{p} - \overrightarrow{q}\right) 4\sin^2\left(\overrightarrow{p} \times \overrightarrow{q}/2\right) \quad (23)$$

From this dispersion relation we can deduce the Fermi velocity at the Fermi momentum $k_f = \sqrt{2}, v_f = 7.5 \times 10^{-2}$. If we compare the effective mass $m^* = k_f/v_f$ with the "bare mass" $m_0 = 2\pi$ defined by the curvature of the dispersion relation at zero momentum $\varepsilon(p) = p^2/2m_0$ one has approximately $m^*/m_0 \approx 3$. Note that due to the lack of Galilean invariance of the theory, there is no relation between the mass and the Landau parameter F_1.

The homogeneous transport equation which follows from the Landau theory is given by [22]:

$$(-s + \cos(\theta))\,\hat{n}(\theta) + 1/2\pi \cos(\theta) \int_0^{2\pi} d\theta F(\theta - \theta')\,\hat{n}(\theta') = 0 \quad (24)$$

where the fluctuation of the quasiparticle distribution takes the form

$$\delta n\left(\overrightarrow{p}, \overrightarrow{r}, t\right) = \delta\left(\varepsilon\left(\overrightarrow{p}\right) - \mu\right) \hat{n}(\theta)\, e^{i\overrightarrow{q}\,\overrightarrow{r} - \omega t} \quad (25)$$

$s = \omega/qv_f$ and θ labels a point on the Fermi surface. The coefficient $F(\theta) = \Omega m^* f(\overrightarrow{p}, \overrightarrow{p}')/2\pi$, where $\overrightarrow{p}, \overrightarrow{p}'$ are two momenta on the Fermi surface making an angle θ with each other.

The two first Fourier modes $\left(F_n = \int d\theta/2\pi F(\theta)\right)$ of $F(\theta)$ are both less than zero, furthermore, when $p_f/\sqrt{2} = .985\ F_0$ becomes less than -1 and the system becomes unstable (The compressibility is negative). Although we have no clear interpretation for this instability, it may be related to our approximation in the following way. We know a priori that the filling factor ν_2 cannot exceed one because at this value all the particles are Bosons. The Fermionic Hamiltonian (20) which we have introduced can in principle give rise to a Fermi liquid with an arbitrary Fermi momentum and the value $p_f = \sqrt{2}$ does not seem to be distinguished among other values at first. When we consider the stability conditions however, we recover an infinite compressibility for a filling factor very close to 1 and the system becomes unstable beyond this value $\left(k_f > \sqrt{2}\right)$.

Before the Hartree–Fock approximation is taken, the Hamiltonian (20) is an exact representation of the dynamics in the sense that the physical spectrum is contained in its spectrum when we consider the finite size dynamics as in the appendix B. Since our description of the Hilbert space is overcomplete, there are eigenstates which must be disregarded in the physical theory.

At the physical point where all the particles are Bosons ($\nu_2 = 1$) one should expect that $\rho^f_{\overrightarrow{p}} |phys\rangle = 0$ on the physical Hilbert space since there is no fermion in the system: $b_s |phys\rangle = 0$. The Hartree–Fock states do not fulfill this requirement. In this paper, we use of the fact that the physical quantities depend on ρ^b or ρ^f only through the total density ρ^t. Although the quantities which depend on $\rho^{b,t}$ individually are not correctly calculated in the Hartree–Fock approximation, it is reasonable to expect than those which depend upon them through the combination ρ^t are correctly obtained. An indication that this approximation makes sense consists in evaluating the Casimir operators $\sum_{\overrightarrow{q}} \rho_{\overrightarrow{q}} \rho_{-\overrightarrow{q}}$ of the algebras generated by $\rho^b_{\overrightarrow{p}}, \rho^f_{\overrightarrow{p}}, \rho^t_{\overrightarrow{p}}$ respectively. While the first two sets of Casimir which are not related to any physical quantities are incorrectly predicted by the Hartree–Fock approximation, the third set which gives the asymptotic behavior of the form factor $S(q)$ is correctly obtained. We shall return to this question in the next section.

Remark 4. Since the interaction between two dipoles is attractive when their momentum are opposite a BCS state should form at very low temperature. In the context of the quantum Hall effect this state should probably be interpreted as a plateau. Unlike in the Chern–Simon case [23] the gap equation which one obtains is real (the dipoles are neutral). This makes suspect that the order parameter is real with nodes on the Fermi surface. Typically $\Delta(\theta) = \Delta_0 \sin(\theta - \theta_0)$ which leads to an $O(2)$ order parameter to describe the transition.

2.3 Form factor and conductivity

Form factor As an application of the preceding formalism we compute the form factor $S\left(\overrightarrow{q},\omega\right)$. We recall that the dynamical form factor is defined as:

$$S\left(q,\omega\right) = \sum_{n} \left|\left\langle n\right| \rho_{\overrightarrow{q}} \left|0\right\rangle\right|^{2} \delta\left(\omega_{0} - \omega_{n}\right) \tag{26}$$

where $\left\langle n\right| \rho_{\overrightarrow{q}} \left|0\right\rangle$ | is the matrix element of the density operator $\rho_{\overrightarrow{q}}$ between the ground state $\left|0\right\rangle$ and a complete basis denoted $\left|n\right\rangle$. [22] contains all the information one can learn about the system in a particle scattering experiment.

At a general filling factor $\nu_2 = 1$ we could in principle use either ρ^t or ρ^b since $\rho^f = 0$. Here we evaluate the form factor in a Hartree–Fock approximation which violates the condition $\rho^f = 0$ and we use ρ^t to evaluate it to be consistent with our evaluation of the energy. The computation done with ρ^b would be justified only at this specific point, it would clearly give an incorrect answer for $S\left(q\right)$ both at small and large q.

In the Fermionic representation $\rho^t\left(q\right)$ is a one body operator $\rho^t\left(q\right) = \sum_{\overrightarrow{k}} 2i \sin \left(\overrightarrow{k} \times \overrightarrow{q}/2\right) c^{+}_{\overrightarrow{k}+\overrightarrow{q}} c_{\overrightarrow{k}}$. Up to the prefactor acts on the vacuum by creating a particle hole excitation of energy:

$$\omega = \varepsilon\left(\overrightarrow{k} + \overrightarrow{q}\right) - \varepsilon\left(\overrightarrow{k}\right) \tag{27}$$

From then the Hartree–Fock calculation proceeds as usual and the expression of the dynamical form factor $S\left(\overrightarrow{q},\omega\right)$ is given by the Fermi-liquid result [22] multiplied by a factor $4\sin^2\left(\overrightarrow{k} \times \overrightarrow{q}/2\right)$ where k is determined by 27.

To test the validity of our approximation let us compare the expression of the static form factor we obtain with some theoretical predictions. The form factor $S\left(q\right)$ is defined as:

$$S\left(q\right) = 1/N \int d\omega S\left(q,\omega\right) \tag{28}$$

For q small $S\left(q\right)$ behaves as $2k_f q^3/3\pi$ instead of $2q/\pi k_f$ in a Fermi liquid. The q^3 behavior of the form factor agrees with the numerical predictions [18] (The predictions do not give the coefficient of q^3 and we shall see that the effect of the interactions is to enhance it). As for the quantum Hall effect form factor [17] there is a q^2 reduction with respect to the normal Fermi liquid behavior at small q.

For $q > 2k_f, S\left(q\right)$ goes to k_f^2 as $k_f^2\left(2 - j_1\left(qk_f\right)/qk_f\right)$ where j_1 denotes the Bessel function (we use a normalization of ρ_q for which there is no exponential behavior of $S\left(q\right)$ at large q unlike in [17]). The limiting value k_f^2 is an exact result related to the Casimir operator of the ρ algebra as shown in

the appendix B. This prediction also indicates the limitation of the present model to the $\nu = 1$ Bosonic case since in the physical $\nu = 1/2$ Fermionic case the limiting value should be $1/2$ (The general limiting value of is $1 + \nu \langle P_{ij} \rangle$ where $\langle P_{ij} \rangle$ denotes the expectation value of the permutation operator $+1$ for Bosons, -1 for Fermions [18]). This $S(q)$ does not reproduce a cusp singularity at $2k_f$ which is indicated by the simulations [18].

Macroscopic response function We can gain a better understanding of the preceding result by comparing the expression of the form factor with the response function evaluated in the quasistatic limit: $(s = \omega/v_f q \ll 1)$. In this limit one can interpret the system as a $2D$ Fermi liquid consisting of dipoles. At the Fermi surface the dipole vector of a quasiparticle with a momentum equal to k_f is given by $d_i = \varepsilon_{ij} k_j$. Let $\rho_q = \sum_k c^+_{k+q} c_k$ denote the Fourier modes of the dipole density. A scalar potential $\phi(\overrightarrow{r}, t)$ acts on the system through an interacting Hamiltonian:

$$H_e = \sum_{\overrightarrow{q}} \int d\omega \rho_{-\overrightarrow{q}} \overrightarrow{q}.\overrightarrow{d}\, \phi(\overrightarrow{q}, \omega)\, e^{-i\omega t} \tag{29}$$

where $\phi(\overrightarrow{q}, \omega)$ is the Fourier transform in space and time of $\phi(\overrightarrow{r}, t)$. In the long wavelength limit $(q \ll k_f)$ the dipole-vector can be replaced by its value at the Fermi surface and the net effect is to replace the interaction Hamiltonian by the usual coupling to a scalar potential:

$$H_e = \sum_{\overrightarrow{q}} \int d\omega \rho^t_{-\overrightarrow{q}} \phi(\overrightarrow{q}, \omega)\, e^{-i\omega t} \tag{30}$$

where $\rho^t_{\overrightarrow{k}} = \sum_{\overrightarrow{q}} \left(\overrightarrow{k} \times \overrightarrow{q} \right) c^+_{\overrightarrow{k}+\overrightarrow{q}} c_{\overrightarrow{q}}$ denotes the long wavelength limit of the total density operator. The response function is defined as:

$$\chi(q, \omega) = \left\langle \rho^t(\overrightarrow{q}, \omega) \right\rangle / \phi(\overrightarrow{q}, \omega).$$

To evaluate it we use the transport equation in the presence of the external force due tot he scalar potential. It reads:

$$(-s + \cos(\theta))\, \hat{n}(\theta) + \cos(\theta)\, \hat{n}(\theta')\, d\theta'/2\pi$$

$$= (qk_f) \sin(\theta) \cos(\theta)\, \phi(\overrightarrow{q}, \omega) \tag{31}$$

We expend the solution of this equation in powers of s and make use of the fact that: $\langle \rho^t(\overrightarrow{q}, \omega) \rangle = \sum_{\overrightarrow{p}} (\overrightarrow{p} \times \overrightarrow{q})\, \delta n(\overrightarrow{p})$. Form the first term we deduce the static response function:

$$\chi(q, 0) = -(qk_f)^2\, \nu(0)/2(1 + F_1) \tag{32}$$

where $\nu(0) = m^* \Omega / 2\pi$ is the density of states on the Fermi surface.

Note the q^2 dependence which is different from the usual Fermi liquid behavior and that F_1 appears instead of F_0. This result does not satisfy the usual compressibility sum rule because the scalar potential does not see the quasiparticles as elementary but rather as dipoles. As a result, it does not deform the Fermi sea symmetrically.

The next order gives an imaginary contribution which is related to the dynamical Form factor [22]. One deduces the following expressions:

$$S(\overrightarrow{q}, \omega) = \left(2m^* \omega / (2\pi)^2 v_f q\right) (k_f q / 1 + F_1)^2 \tag{33}$$

The first factor is the free fermion result. The factor $(k_f q)^2$ was predicted at the Hartree–Fock approximation and originates from the fact that dipoles couple much more weakly to the scalar potential as ordinary quasiparticles. Finally the many-body effects renormalize the Hartree–Fock contribution by $(1 + F_1)^{-2} \approx 4$ (instead of $(1 + F_0)^{-2}$) in the usual Fermi liquid case.

Conductivity Let us first consider the system in a constant electric field. We put the field E in the y direction by choosing the gauge $A_y = -Et$. The effect is to replace the x dependence of the one body-wave functions (2) by $x - Et$ and thus to produce a Galilean transformation in the x direction. Since the vacuum $|0\rangle$ carries the charge and moves at speed E in the x direction one recovers in this way the standard $\nu = 1$ quantum Hall conductivity: $\sigma_{xy} = e^2 / 2\pi$. To see the effect of the electric field on the dipoles it is simpler to put the electric field in the y direction by adding to the Hamiltonian (1) a potential equal to Ex. After projection onto the lowest Landau level, this adds to the Hamiltonian (20) a term equal to $\overrightarrow{E} \times \overrightarrow{K}$ where $\overrightarrow{K}$ is the total momentum operator: $\overrightarrow{K} = \sum_{\overrightarrow{k}} \overrightarrow{k} c^+_{\overrightarrow{k}} c_{\overrightarrow{k}}$. This interaction can be interpreted as a Lagrange multiplier which gives each quasiparticle an additional energy proportional to $\overrightarrow{E} \times \overrightarrow{K}$ and thus a rapidity proportional to $\left|\overrightarrow{E}\right|$ in the direction perpendicular to $\overrightarrow{E}$. Physically, we can say that although the dipoles are neutral, the electric field tends to stretch them and since their dipole vector is proportional to their momentum they acquire a rapidity by this mechanism.

We can use the arguments based on the transport equation to compute the transverse conductivity $\sigma_\perp(\overrightarrow{q})$ in the quasistatic limit in a similar way as for the response function. The electric-field is applied in the x direction and the wave vector $\overrightarrow{q}$ is perpendicular to it. We evaluate the conductivity in the direction of the electric field ($\sigma_\perp(\overrightarrow{q}) = \sigma_{xx}(\overrightarrow{q})$) in the quasistatic limit. In a normal $2D$ Fermi liquid the transverse conductivity evaluated in the Landau theory is equal to $\sigma_\perp(\overrightarrow{q}) = 2\rho_e e^2 / k_f q$.

Let J denote the quasiparticle current at the Fermi surface. In the dipole theory we define a modified current J^t in such a way that the variation of the

kinetic energy with the time is proportional to $\overrightarrow{J}^t(\overrightarrow{x}) = \overrightarrow{d}.\overrightarrow{\nabla}J(\overrightarrow{x})$ where $\overrightarrow{d}$ denotes the dipole vector. The computation of the conductivity then proceeds as for the response function and the net effect of this redefinition is to renormalize the Fermi liquid result by the factor $(k_f q)^2$. In the case of interest $\nu_2 = 1$ one has $\rho_e = 1/2\pi$ and $k_f = \sqrt{2}$ so that $\sigma_\perp(\overrightarrow{q}) = e^2\sqrt{2}q/\pi$. We see that this theory predicts an anomalous conductivity proportional to q in a straightforward way. Unfortunately, this does not produce the longitudinal conductivity $\left(\overrightarrow{E}\text{ and }\overrightarrow{q}\text{ in the }x\text{ direction}\right)$ which is observed in the experiments [3].

The physical interpretation of the above result is simple: The only quasiparticles which can radiate must travel with the group velocity of the electric field. When the group velocity is small compared to v_f they lie at the two points of the Fermi surface with a Fermi momentum perpendicular to $\overrightarrow{q}$. Since the electric field is not constant in the direction of $\overrightarrow{q}$, it couples to the dipole vector of the quasiparticles and the system radiates when $\overrightarrow{E}$ is parallel to their velocity.

2.4 Conclusion

We have introduced a microscopic model to analyze the problem of Bosonic particles in a strong magnetic field at $\nu = 1$. We have generalized the model in order to be able to vary continuously the Fermi momentum p_f and we have been able to study it in a Hartree–Fock approximation.

The present model gives a description in agreement with the dipole picture introduced by N. Read [5]. The main conclusion of our study is that the system behaves essentially as a gas of Fermionic dipoles with a dipole vector perpendicular to their momentum and a strength proportional to the momentum. As a result, the interactions are screened and the gas behaves in many respect as a neutral Fermi liquid (the same conclusion is reached in [7]). The response theory to an external field is modified by the fact that the quasiparticles couple as dipoles to the field. The main consequence is that the linear response quantities get renormalized by a factor $(k_f q)^2$ coefficient at low momentum transfer q and that the Landau parameter F_1 appears instead of F_0 in the final results. The Landau theory only relies on the hypotheses that the quasiparticles are dipoles with a dipole vector perpendicular and equal in strength to the Fermi momentum. it should therefore also be valid in the $\nu = 1/2$ case.

Since our conclusions are based on a self consistent approximation, we cannot prove them from first principles. In particular, the theory should become inconsistent when the Fermi momentum $p_f > \sqrt{2}$. A somewhat unexpected feature of our numerical study is the fact that the Landau parameter $F_0 = -1$ very close to the physical point $p_f = \sqrt{2}$.

The model differs from the Chern–Simon theory by the fact that the dynamics is projected into the Lowest Landau Level. Thus, the effective mass

depends only on the interactions. This model does not seem to predict a divergence of the effective mass. It predicts a form factor $S(q)$ with the expected qualitative behavior at small and large wave vector q. While the low q behavior is expected to be the same as in the $\nu = 1/2$ case, the large Q behavior is specific to the $\nu = 1$ Bosonic case. It is however the best test that we have that our approximation is consistent.

Curiously, we do not predict a longitudinal conductivity linear in the wave vector q but instead we compute a transverse conductivity with this behavior .

Acknowledgement
We are particularly thankful to D. Haldane for helpful discussions, we have also benefited from helpful discussions with G. Baskaran and N. Read.

Appendix A

in this appendix we recast the commutation relations (11, 12, 13) of the order parameters in a "local form" by introducing a non commutative formalism. This enables us to obtain a Fermionic representation for those fields which turns out to be a generalized Holstein–Primakov transformation. Let us introduce a plane wave basis which obey the magnetic translation group multiplication law:

$$e^{i\vec{p}\,\vec{q}}\,e^{i\vec{q}\,\vec{x}} = e^{\vec{p}\times\vec{q}/2}e^{(\vec{p}+\vec{q})\,\vec{x}} \tag{34}$$

We also impose that the Fourier modes commute at different space positions:

$$e^{i\vec{p}\,\vec{x}}\,e^{i\vec{p}\,\vec{y}} = e^{i\vec{q}\,\vec{y}}\,e^{i\vec{p}\,\vec{x}} \tag{35}$$

Given the Fourier modes $\rho(\vec{p})$ of a density operator we define $\rho(\vec{x})$ as:

$$\rho(\vec{x}) = 1/\Omega \sum_p \rho(\vec{p})\, e^{i\vec{p}\,\vec{x}} \tag{36}$$

Ω is the total area. To be consistent with our earlier definitions we do the same with the Fermion operators $A^+(\vec{x})$ but we replace Ω by $\sqrt{\Omega}$ in the definition.

$$A^+(\vec{x}) = 1/\sqrt{\Omega} \sum_p A^+(\vec{p})\, e^{i\vec{p}\,\vec{x}} \tag{37}$$

$A(\vec{x})$ is defined as the hermitian conjugated of $A^+(\vec{x})$.

We also define the delta function $\delta^2(\vec{x}) = 1/\Omega \sum_p e^{i\vec{p}\,\vec{x}}$. Because the Fourier modes do not commute, one must be careful in commuting $\delta^2(\vec{x} - \vec{y})$ through an operator $O(\vec{x})$. The rule is given by:elations

$$\delta^2(\vec{x} - \vec{y})\, O(\vec{x}) = O(\vec{y})\, \delta^2(\vec{x} - \vec{y})$$

as can be easily verified.

With this formalism, the commutation between the $\rho's$ and the $A's$ become:

$$\begin{aligned}
\left[\rho^b\left(\overrightarrow{x}\right),\rho^b\left(\overrightarrow{y}\right)\right] &= \rho^b\left(\overrightarrow{x}\right)-\rho^b\left(\overrightarrow{y}\right)\delta^2\left(\overrightarrow{x}-\overrightarrow{y}\right)\\
\left[\rho^f\left(\overrightarrow{x}\right),\rho^f\left(\overrightarrow{y}\right)\right] &= \rho^f\left(\overrightarrow{x}\right)-\rho^f\left(\overrightarrow{y}\right)\delta^2\left(\overrightarrow{x}-\overrightarrow{y}\right)\\
\left[\rho^b\left(\overrightarrow{x}\right),\rho^f\left(\overrightarrow{y}\right)\right] &= 0
\end{aligned} \tag{38}$$

$$\begin{aligned}
\left[\rho^b\left(\overrightarrow{x}\right),A^+\left(\overrightarrow{y}\right)\right] &= A^+\left(\overrightarrow{x}\right)\delta^2\left(\overrightarrow{x}-\overrightarrow{y}\right)\\
\left[\rho^f\left(\overrightarrow{x}\right),A^+\left(\overrightarrow{y}\right)\right] &= -A^+\left(\overrightarrow{y}\right)\delta^2\left(\overrightarrow{x}-\overrightarrow{y}\right)
\end{aligned} \tag{39}$$

$$\left\{A\left(\overrightarrow{x}\right),A^+\left(\overrightarrow{y}\right)\right\} = 2\pi\left(\rho^b\left(\overrightarrow{y}\right)+\rho^f\left(\overrightarrow{x}\right)\right)\delta^2\left(\overrightarrow{x}-\overrightarrow{y}\right) \tag{40}$$

A first result which can be easily obtained from these relations is that the operators $\int \rho^{b(f)^n} d^2x$ are the Casimir operators of the $\rho^{b(f)}$ algebra respectively.

Let us now consider Fermionic operators $c_{\overrightarrow{p}}, c^+_{\overrightarrow{p}}$ such tat $\left\{c_{\overrightarrow{p}}, c^+_{\overrightarrow{q}}\right\} = \delta_{\overrightarrow{p},\overrightarrow{q}}$. The field $c^+\left(\overrightarrow{x}\right)$ is defined as in 38 and $c\left(\overrightarrow{x}\right)$ is its hermitian conjugate. They verify:

$$\begin{aligned}
\left\{c\left(\overrightarrow{x}\right),c\left(\overrightarrow{y}\right)\right\} &= \left\{c^+\left(\overrightarrow{x}\right),c^+\left(\overrightarrow{y}\right)\right\} = 0\\
\left\{c\left(\overrightarrow{x}\right),c^+\left(\overrightarrow{y}\right)\right\} &= \delta^2\left(\overrightarrow{x}-\overrightarrow{y}\right)
\end{aligned} \tag{41}$$

It is then straightforward to verify that the operators $c^+\left(\overrightarrow{x}\right)c\left(\overrightarrow{x}\right)$ and $c\left(\overrightarrow{x}\right)c^+$
$\left(\overrightarrow{x}\right)$ obey the same commutation relations as $\rho^b\left(\overrightarrow{x}\right)$ and $\rho^f\left(\overrightarrow{x}\right)$ respectively and that they commute with each other. A possible realization of the operators $A\left(\overrightarrow{x}\right), A^+\left(\overrightarrow{x}\right)$ which satisfies the correct commutation relations is then:

$$\begin{aligned}
A^+\left(\overrightarrow{x}\right) &= c^+\left(\overrightarrow{x}\right)\\
A\left(\overrightarrow{x}\right) &= 2\pi c\left(\overrightarrow{x}\right)c^+\left(\overrightarrow{x}\right)c\left(\overrightarrow{x}\right) = 2\pi\rho^f\left(\overrightarrow{x}\right)c\left(\overrightarrow{x}\right)
\end{aligned} \tag{42}$$

This representation however, does not achieve the hermiticity condition A^+
$\left(\overrightarrow{x}\right) = \left(A\left(\overrightarrow{x}\right)\right)^+$. One way out is not to impose that $c\left(\overrightarrow{x}\right)$ and $c^+\left(\overrightarrow{x}\right)$ are Hermitian conjugated; another one consist in modifying the expression of $A\left(\overrightarrow{x}\right), A^+\left(\overrightarrow{x}\right)$ as follow: We show in the text section that the commutation relations are not modified if $A^+\left(\overrightarrow{x}\right) = c^+\left(\overrightarrow{x}\right)\sqrt{2\pi\rho^f\left(\overrightarrow{x}\right)}, A\left(\overrightarrow{x}\right) = \sqrt{2\pi\rho^f\left(\overrightarrow{x}\right)}c\left(\overrightarrow{x}\right)$ and the Hermiticity condition is satisfied. The square root the operator $2\pi\rho^f\left(\overrightarrow{x}\right)$ can be understood as $\sqrt{1-2\pi\rho^f\left(\overrightarrow{x}\right)}$ to define a formal series in the normal ordered operator $\rho^h\left(\overrightarrow{x}\right)$. This representation is reminiscent of the Holstein Primakov representation of the $su\left(1|1\right)$ algebra.

As for the Holstein–Primakov case the Fermionic Hilbert space must be truncated. This can be traced back to the fact that the physical scalar product of the "Slater determinants" (16) does not coincide with the natural scalar product in the Fock Hilbert space $c^{+}\left(p^{1}\right)c^{+}\left(p^{2}\right)...c^{+}\left(\overrightarrow{p}N\right)|0\rangle$. As a result there are linear combinations of slater determinants with a zero physical norm. This means that if one replaces $A^{+}_{\overrightarrow{k}}$ by their expression (7) these states vanish identically. In other words the $c^{+}_{\overrightarrow{k}}$ Hilbert space is an overcomplete description of the physical states and must be properly truncated.

Invariance of the commutation relations We show that the commutation relations of the operators $c^{+}\left(\overrightarrow{x}\right)$ and $c^{+}\left(\overrightarrow{x}\right)$(41) are not modified if one makes the substitution $c^{+}\left(\overrightarrow{x}\right) \rightarrow C\left(\overrightarrow{x}\right) = \Psi\left(\rho^{f}\left(\overrightarrow{x}\right)\right)c\left(\overrightarrow{x}\right)$,
$c^{+}\left(\overrightarrow{x}\right) \rightarrow C^{+}\left(\overrightarrow{x}\right) = c^{+}\left(\overrightarrow{x}\right)\Psi^{-1}\left(\rho^{f}\left(\overrightarrow{x}\right)\right)$. In the following ρ stands for ρ^{f} and we denote the vector $\overrightarrow{a}$ as a. We assume that $\Psi\left(\rho\right)$ can be represented as $\Psi\left(\rho\right) = \int due^{up}f\left(u\right)$. In which case $\Psi^{-1}\left(\rho\right) = \int due^{up}g\left(u\right)$ where $duf\left(u\right)g\left(t-u\right) = \delta\left(t\right)$. If one substitutes $\Psi\left(\rho\right)$ by its expression in the commutator, the result follows if the following commutator depends only on the sum $\lambda + \mu$:

$$B\left(\lambda,\mu\right) = \left\{e^{\lambda\rho^{f}\left(x\right)}c\left(x\right), c^{+}\left(y\right)e^{\mu\rho^{f}\left(y\right)}\right\} = \tilde{B}\left(\lambda+\mu\right) \tag{43}$$

Here $\tilde{B}\left(\lambda\right)$ is an operator function which verifies $\tilde{B}\left(0\right) = \delta\left(x-y\right)$.

To show (44) it is sufficient to show that $dB/d\lambda - dB/d\mu = O$. This last expression can be rewritten as:

$$\begin{aligned} & e^{\lambda\rho(x)}\left(\left\{\rho\left(x\right)c\left(x\right), c^{+}\left(y\right)e^{\mu\rho(y)}\right\} - \left\{c\left(x\right), c^{+}\left(y\right)e^{\mu\rho(y)}\rho\left(y\right)\right\}\right) \\ & - \left(\left\{e^{\lambda\rho(x)}, c^{+}\left(y\right)e^{\mu\rho(y)}\right\}\rho\left(x\right) - \left\{e^{\lambda\rho(x)}, c^{+}\left(y\right)\rho\left(y\right)e^{\mu\rho(y)}\right\}\right)c\left(x\right) \\ & = e^{\lambda\rho(x)}C + Dc\left(x\right) \end{aligned} \tag{44}$$

We now show that $C = D = 0$.

$$\begin{aligned} C = & \left(\left\{\rho\left(x\right)c\left(x\right)c^{+}\left(y\right)\right\} - \left\{c\left(x\right), c^{+}\left(y\right)\rho\left(y\right)\right\}\right)e^{\mu\rho(y)} \\ & -c^{+}\left(y\right)\left(\left\{\rho\left(x\right)c\left(x\right), e^{\mu\rho(y)}\right\} - \rho\left(y\right)\left\{c\left(x\right), e^{\mu\rho(y)}\right\}\right) \end{aligned} \tag{45}$$

The first line of the above expression can easily be seen to vanish. If we call $\beta\left(\mu\right)$ the coefficient of $c^{+}\left(y\right)$, one can show along the same lines that it obeys $d\beta/d\mu = \rho\left(y\right)\beta\left(\mu\right)$ and it therefore vanishes since $\beta\left(0\right) = 0$. Thus $C = 0$.

To show that $D = 0$ one considers D as a function of λ. $D\left(0\right) = 0$ and one has:

$$\begin{aligned} & dD/d\lambda = \rho\left(x\right)D\left(\lambda\right) \\ & + \left(\left\{\rho\left(x\right), c^{+}\left(y\right)e^{\mu\rho(y)}\right\} - \left\{\rho\left(x\right), c^{+}\left(y\right)e^{\mu\rho(y)}\rho\left(y\right)\right\}\right)e^{\lambda\rho(x)} \end{aligned} \tag{46}$$

The coefficient of $e^{\lambda\rho(x)}$ can be evaluated directly and vanishes identically therefore $D = 0$ as well. This proves the result.

Appendix B

In this appendix we extend the formalism to the case of periodic boundary conditions. This enables us to relate the approach of this paper to the trial-wave functions considered by Haldane [18]. Let us consider the case of a periodic box of length $L_x L_y = 2\pi N$ where N is an integer so that exactly N flux tubes thread the box.

The lowest Landau level wave functions are obtained by taking the sum of the wave functions $\langle x|s\rangle$ defined in (2) for $s = s_0 + kL_x, k \in Z$ in order to make them quasi-periodic in the x direction.

$$\langle \overrightarrow{x}|s_0\rangle = \sum_{s=s_0+kL_x} \sqrt{1/\pi L_y} e^{isy-(x-s)^2/2} \tag{47}$$

Now $s_0 = 2\pi n_0/L_y$ where n_0 takes the values $0, 1, ...N-1$ and $\langle x + L_x y|s\rangle = e^{iL_x y}\langle x, y|s\rangle$. Let us define the variables $z = L_x(y - ix)/2\pi, \tau = iL_x^2/2\pi$ and $a = n_0/N$. The periodic wave functions rewrite:

$$\langle \overrightarrow{x}|s_0\rangle = e^{-\pi Im(z)^2)/Im(\tau)}\theta_{a,0}(z,\tau) \tag{48}$$

where $\theta_{a,b}$ for $a, b \in Z/N$ is the theta function with characteristics a, b [20] defined as:

$$\theta_{a,b}(z,\tau) = \sum_{n\in Z} \exp\left(\pi i (a+n)^2 \tau + 2\pi i (n+a)(z+b)\right) \tag{49}$$

They obey the following periodicity conditions:

$$\begin{aligned} \theta_{a,0}(z+N) &= \theta_{a,0}(z) \\ \theta_{a,0}(z+\tau) &= e^{-i\pi\tau-2i\pi z}\theta_{a,0}(z) \end{aligned} \tag{50}$$

Unlike those defined in (2), the wave functions $\langle \overrightarrow{x}|s_0\rangle$ are not eigenstates of the operator $R_x = p_y$ but they are eigenstates of $\gamma_1 = e^{2\pi i R_y/L_y}$ with the eigenvalues $e^{2\pi i a}$. Similarly, the operator $\gamma_2 = e^{2\pi i R_y/L_y}$ where $R_y = p_x - y$ acts on the wave functions by shifting a by $-1/N$. One recovers the action of the Heisenberg group $\left(\gamma_1^N = \gamma_2^N = 1, \gamma_1\gamma_2 = e^{-2i\pi/N}\gamma_2\gamma_1\right)$ on the theta functions given by:

$$\begin{aligned} \gamma_1\theta(z) &= \theta(z+1) \\ \gamma_2\theta(z) &= e^{-i\pi\tau/N^2-2i\pi z/N}\theta(z-\tau/N) \end{aligned} \tag{51}$$

In these notations, the momentum $\overrightarrow{p} = 2\pi(n_1/L_x, n_2/L_y)$ so that the one body operator $e^{ip_y R_y}e^{ip_x R_x} = \gamma_2^{n_2}\gamma_1^{n_1}$. Assuming the particles are Bosons, the

second quantized expression of this operator is given by $\sum_{s=2\pi n/L_y} e^{ip_x s} a^+_{s-p_y} a_s$ where n takes the values $0, ..., N-1$ and a_s, a^+_s are Bosonic creation and annihilation operators. This operator coincides up to a phase with the projected density operator $\rho^b_{\overrightarrow{p}}$ defined in (10). The same holds for $\rho^f_{\overrightarrow{p}}$ if the particles are Fermions and Fermionic creation operators $b^+_s, (b_s)$. One can verify that the two sets $\rho^b_{\overrightarrow{p}}$, and $\rho^f_{\overrightarrow{p}}$ generate two commuting $SU(N)$ lie algebras when we consider their commutators.

The Fermionic ground state which is the only possible state when the Lowest Landau level is occupied by N Fermionic particles is given by the slater determinant of the one-body wave-functions $\Psi_0(\overrightarrow{x}_i) = \mathrm{Det}(\langle \overrightarrow{x}_i | j/N \rangle)$.

Consider the case where all the particles obey a Bosonic statistics ($N_2 = N$). Given a set of N momenta $\overrightarrow{p}_i$, one can construct a symmetric wave function given by:

$$\Psi_{\{\overrightarrow{p}_i\}}(\overrightarrow{x}_i) = \sum_{\sigma \in S_N} (-)^\sigma \prod_j e^{-ip_x \overrightarrow{R}_\sigma} \Psi_0(\overrightarrow{x}_i) \tag{52}$$

where the operator $e^{i\overrightarrow{p}\overrightarrow{R}_i}$ acts on the coordinate $\overrightarrow{x}_i$. Here $e^{i\overrightarrow{p}\overrightarrow{R}}$ stands for the ordered expression $e^{ip_x p_y/2} e^{ip_y R_y} e^{ip_x R_x}$ defined above. These trial-wave functions have been considered by Haldane [18] to study the Fermi-sea formation in the $\nu = 1$ case. If we define the operators $A^+_{\overrightarrow{p}} = e^{-ip_x p_y/2} \sum_s e^{ip_x s} a^+_{s-p_y} b_s$, the second-quantized expression for these wave functions is given by $\prod_{\overrightarrow{p}_i} A^+_{\overrightarrow{p}_i} |0\rangle$. This establishes the link between the formalism used in this paper and the trial wave functions studied in [18].

Casimir operator

The quadratic Casimir operator of the ρ^t algebra is given by

$$C_2 = \sum_{\overrightarrow{q}} \rho^t_{\overrightarrow{q}} \rho^t_{-\overrightarrow{q}} = \sum_{\overrightarrow{q}} e^{i\overrightarrow{q}(\overrightarrow{R}_i - \overrightarrow{R}_j)} \tag{53}$$

The action of this operator on a many-body wave functions of M particles is given by

$$C_2 = \sum_{i,j=1}^{M} \sum_{n_1,n_2=1}^{M} (\gamma_1^{n_1} \gamma_2^{n_2})_i \left(\gamma_2^{-n_2} \gamma_1^{-n_1}\right)_j \tag{54}$$

where the subscripts i, j means that the operator acts on the coordinate z_i, z_j respectively. When we evaluate this operator on a wave function the product of the two factors when $i = j$ is equal to 1 and one can verify that for $i \neq j$ the sum oven n_1, n_2 adds up to N times the permutation of the coordinates z_i and z_j. Thus $C_2 = N^2 M + N \sum_{i \neq j} P_{i,j}$. If the system consists of N_1 Fermions and N_2 Bosons, only two possible representations of the permutation group appear in the decomposition. They are characterized by the

two young tableaux $\left(N_2, 1^{N_1}\right)$ and $\left(N_2+1, 1^{N_1-1}\right)$. In the representations $\left(\lambda_1, 1^{\lambda_2}\right)$ the operator $\sum_{i\neq j} P_{i,j}$ takes the value $(\lambda_1+\lambda_2)(\lambda_1-\lambda_2-1)$ so that the Casimir C_2 takes 2 possible value which can be evaluated.

Let us consider the thermodynamic limit $M = N_1+N_2 = N \to \infty, N_2/N = \nu_2$. For q large $S(q)(28)$ has a limit given by $1/N^3 C_2 = 2\nu_2$. This result can be compared with the Hartree–Fock approximation which gives $S(q) \to k_f^2$. Since $k_f^2 = 2\nu_2$ the two results agree with each other although we have no good explanation for this fact.

References

1. R.B. Laughlin, Phys. Rev. Lett.**50,** 1395 (1983).
2. J.K. Jain, Phys. Rev. Lett. **63**, 199 (1989).
3. R.L. Willet, M.A. Paalanen, R.R. Ruel, K.W. West, L.N. Pfeiffer and D.J. Bishop, Phys. Rev. Lett. **54,** 112 (1990).
4. B.A. Halperin, P.A. Lee and N. Read, Phys. Rev. B**47**, 7312 (1993).
5. N. Read, Semicon. Sci. Tech. **9**, 1859 (1994).
6. N. Read, Phys. Rev. Lett. **62**, 86 (1989).
7. G. Baskaran, Physica B**212**, 320 (1995).
8. H.J. Kwon, A. Houghton and J.B. Marston, Phys. Rev. Lett. **73**, 284 (1994).
9. Y.B. Kim, W.G. Wen, P.A. Lee and A. Furusaki, Phys. Rev. B**50**, 17917 (1994).
10. B.L. Altshuler, L.B. Ioffe and A.J. Millis, Phys. Rev. B**50**, 14084 (1994).
11. C. Nayak and F. Wilczed, Nucl. Phys. B**417**, 359 (1994).
12. D.C. Tsui, H. Störmer and A. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).
13. E.H. Rezayi and N. Read, Phys. Rev. Lett. **72**, 900 (1994).
14. C. Kallin and B. Halperin, Phys. Rev. B**30**, 5655 (1984).
15. I.V. Lerner and Yu.E. Lozovik, Zh. Eksp. Theor. Fiz **78**, 1167 (1978) [Sov. Phys. JETP**51**, 588 (1980)].
16. L.P. Gor'kov and I.E. Dzyoloshinskii, Sov. Phys. JETP**26,** 449 (1968).
17. S.M. Girvin, A.H. MacDonald and P.M. Platzman, Phys. Rev. B**33**, 2481 (1986).
18. F.D.M. Haldane, Private communication.
19. F. Dyson, Phys. Rev. **102**, 1217 (1956).
20. D. Mumford "Tata Lectures on Theta" Birkhäuser (1983).
21. K. Moon, H. Mori, Kun Yang, S.M. Girvin, A.H. MacDonald, L. Zheng, D. Yoshioka and shou-Cheng Zhang, Phys. Rev. B**51**, 5138 (1995).
22. D. Pines and P. Nozières "The Theory of Quantum Liquids" Vol.1, W.A. Benjamin, Inc. (1966).
23. M. Greiter, X.G. Wen and F. Wilczek, Phys. Rev. Lett. *66*, 3205 (1991).
24. R. Shankar and G. Murthy, Phys. Rev. Lett. 79, 4437 (1997).
25. V. Pasquier and F.D.M. Haldane, Nucl. Phys. B**51**, 719 (1998).
26. N. Read, Phys. Rev. B**58**, 16262 (1998).
27. D.H. Lee, Phys. Rev. Lett. 80, 4745 (1998).
28. A. Stern, B.I. Halperin, F. von Oppen, S. Simon, cond-mat/9812135.

Stored Ion Manipulation Dynamics of Ion Cloud and Quantum Jumps with Single Ions

Fernande Vedel *

Physique des Interactions Ioniques et Moléculaires (UMR 6633 CNRS - UAM1), Université de Provence, Centre de St-Jérôme, Case C21, F - 13397 Marseille Cedex 20

Abstract. Ion storage is a powerful tool for keeping charged particles for very long times in a perturbation-free environment, a perfect system for atomic physics or frequency metrology. Due to the existing anharmonicies in the confining field, equations of the ion motion are governed by non-linear dynamics and new frequencies built on the classic rules of the frequency dynamics appears.

In addition with laser cooling, the technique allows one to develop precise investigations on microscopic systems and then very fundamental illustrations in quantum optics. Moreover, storing single ion are now a "common" to propose new frequency standards in the optical domain; however the frequency locking on the clock transition presents the originality to use the quantum jumps detection. The paper will essentially present some highlights in connection with the research undertaken at PIIM.

1 Introduction

Due to the long storage times possible, the very few collisions with neutral background gas, the high degree of spatial localisation and excellent detection sensitivities, measurements on trapped ions allow very high resolution spectroscopy (such as hyperfine structure and radiative lifetime measurements), the proposal of novel frequency standards and the observation of several quantum effects [1]. This paper is obviously limited to the presentation of some aspects among the numerous applications more or less connected to the studies carried on in Marseille.

The motion equation in the ideal case (Mathieu equation) of confined ion in a Paul trap and then the properties of this ion motion in the real cases, where limit conditions of the trap, as well as collisions, space charge, ... will modify the equation of motion will be first discussed. Ions could be considered as an ensemble of similar oscillators with non-linear properties which raise frequency locking and new motion frequencies, as well as instabilities or amplitude increases at this frequencies. Experimental illustrations of these characteristics will be described.

Because the ion motion is oscillating, it is possible to reach situations where the amplitude of ion motion is much less than the wavelength of an

* fern@frmrs12.u-3mrs.fr

M. Planat (Ed.): LNP 550, pp. 107–124, 2000.
© Springer-Verlag Berlin Heidelberg 2000

atomic transition (Lamb-Dicke regime). The first order Doppler effect is suppressed. Performing frequency standard in the microwave regions using the ion storage are then realizable.

Laser cooling techniques permit to extend this conditions in the optical domain. The storage of a chain of few ions or of a single particle quasi at rest can also be used for developing new frequency standards also in the optical domain. The last part of this contribution will be devoted to present shortly the progress in this area.

2 Dynamics of Stored Ions

2.1 Ion motion in a pure quadrupole field

Ideally, the trajectory of the ion is governed by three independent pure Mathieu equations as for the z-component of the motion [2]:

$$\frac{dz^2}{dt^2} + \frac{\Omega^2}{4}(a_z - 2q_z \cos \Omega t)z = 0 \tag{1}$$

For a given e/m ratio, the stability domain depends on the parameters a_z and q_z which are direct functions of the dc and ac trapping voltages U_{dc} and V_{ac}, of the r.f. frequency Ω of the ac voltage and of the dimension parameters of the trap r_0, z_0 [2]. The corresponding solution is :

$$z(t) = u_z \left(A_z(t) \cos(\omega_z t + \phi_z) + B_z(t) \sin(\omega_z t + \phi_z)\right) \tag{2}$$

where $\mathrm{A}_z(\mathrm{t})$ and $\mathrm{B}_z(\mathrm{t})$ are the sums of the harmonics of the r.f. frequency Ω, u_z and ϕ_z the motion constants, ω_z is the secular frequency which is related to the parameter β_z such that $\omega_z = \beta_z \Omega/2$. β_z is computed by an iterative technique from a_z and q_z. The frequency spectrum is formed by the ensemble of Ω and sums of $\Omega \pm \omega_z$ shows that the ion motion could be considered as the superposition of two contributions: the micro-motion at Ω and the macro-motion at ω_z.

Refined 3D-modelisation of the ion and Monte-Carlo simulations allows to measure, with T.O.F. techniques the kinetic energy of the ion cloud with a good accuracy [3,4]. Actually, this information is required for applications in physico-chemistry and to ion/molecule collision studies

The spectral properties of the ion motion will influence the shape of absorption or emission profiles, which normally is monitored by the Doppler effect (first order and second order). Actually, the frequency of the radiation (ω_0 in the laboratory frame) as seen by the ion is modulated by the first Doppler effect and the time dependence of the field becomes in $\cos(\omega_0 + \frac{\omega_0 v}{\omega_z c} \sin \omega_z t)$, where v is the maximum amplitude of the velocity. This modulation can be expressed in Bessel functions developments and that consequently the ion see

the radiation with discrete components spaced symmetrically about ω_0 and separated by the secular frequency. These components will be resolved when the first-order shift is less than the secular frequency, that is finally when the amplitude of the motion is smaller than the wavelength divided by π. This last condition is so-called "the Lamb–Dicke regime". In this case the first-Doppler profile become discret [5,6].

2.2 Ion motion in experimental conditions

In real cases, the ion motion is governed by a more complex equation than 1. Investigating the fundamental aspect of the properties of the Mathieu equation with additional terms is important for developping new tools for non linear physics. On the other hand, a greater knowledge of the dynamics of the ion cloud is always sought, for instance, for refining ion detection methods, mass-selective trapping of injected ion species, and the determination of the optimum conditions for prolonged ion storage at quasi-rest.

The ion motion must be described by:

$$\begin{aligned} \frac{d^2z}{dt^2} + \frac{\Omega^2}{4}(a_z - 2q_z\cos\Omega t)z &= \Gamma\frac{dz}{dt} + \sum_{i=1}^{n} A\frac{z-z_i}{r^3} + q_e\frac{\partial V_e(\omega,z)}{\partial z} \\ &+ \ f(x,y,z) \\ r^2 &= (x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2 \end{aligned} \tag{3}$$

where the viscosity Γ in the first term of the right-hand side represents collisions with buffer gas, or laser cooling interaction, the second term describes the coulombic interactions, the third one could be an alternating electrical excitation and finally $f(x,y,z)$ takes into account the imperfections of the rf driving field.

Frequencies are defined within a sharp but not null interval, therefore integer ratios between them are experimentally unreachable. However, when the ratio of the frequencies involved is close to a rational number, this ratio becomes rational due to the presence of nonlinearities, that will result in frequency locking. Frequency locking is understood in terms of the excitation of an oscillator with a periodic force [7]. The eigenfrequencies of the oscillator will tend to values corresponding to rational ratios between these frequencies and those of the excitation and will lead to resonance.

In Fig. 1 is shown, as an example, the resulting frequency locking, as computed in the space charge case [8]. Actually, Poincaré sections computed when increasing the importance of the additional space charge term do not continue to be well defined ellipses, but spiral successively in an increasing or decreasing way limited by two ellipses, according to the KAM theorem. Then for a given threshold of the perturbation amplitude, points are concentrated in discrete islands instead of a quasiuniform distribution. When the

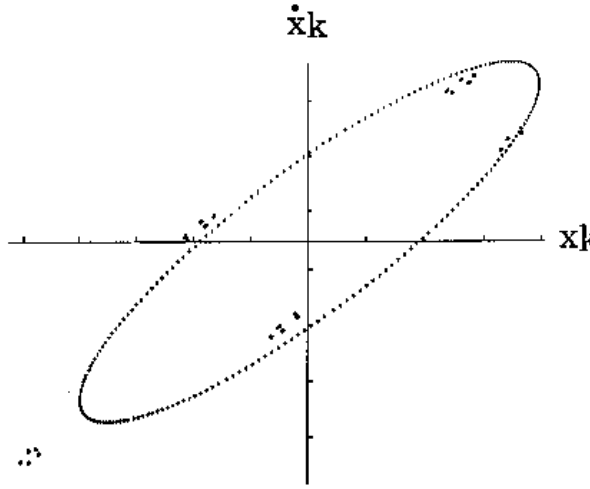

Fig. 1. Example of frequency locking. Poincaré section (or stroboscopic phase space representation for one component: $x_k, \dot{x}_k$, at rf-phase $= \pi/2$) of the trajectory of an rf stored ion is computed in the presence of strong coupling of the degrees of freedom. Here, the perturbation, modeling space charge effect, is shaped approximately as $Kn \int 2k(\alpha) x exp\left(-k(\alpha)\left(x^2 + y^2\right) + k'(\alpha) z^2\right) d\alpha$, n being the ion number, K and k geometrical factors. The pure ellipse corresponds to the case of "n =0"; the islands demonstrate that the secular frequency becomes a multiple of the rf frequency and arises for a given value of n depending on the ion temperature

perturbation becomes larger, resonances can also appear for ratios different from an integer value, following the Arnold tongue scheme, in the same way that a viscosity term allows forced resonances not only for one value but for an interval around this value [10].

Following the pioneering work of Kotowski [11] for the theoretical aspect of the modified Mathieu equation, and of Dawson [12] on its applications to the quadrupole mass filter, experimental [13,14], as well as theoretical and numerical [15] studies have been carried out subsequently. A dramatic variation of the ion storage capability was observed as the working point was moved within the stability diagram that led to the discovery of the so-called black holes and black canyons [16,17]. These defects in trapping capability were explained by the presence of couplings between the macro- and micro-motions, which were already observed previously by Paul in the early experiments for the 2D-case (linear mass filter [18]).Trapping variations due to axial-radial motion couplings besides those coming from the well known macro-micro ones can be also observed.

As mentioned in [15] resonant couplings between the different degrees of freedom can be predicted from the following equations:

$$n_x \frac{\beta_x}{2} + n_z \frac{\beta_z}{2} = \aleph \qquad N = \mid n_x \mid + \mid n_z \mid \tag{4}$$

with n_x or $n_z = 0, \pm 1, \pm 2, \ldots$ and $\aleph = 0, 1, 2, \ldots$

The case $\aleph = 0$ corresponds to radial-axial couplings, while $\aleph \neq 0$ describes couplings between the macro- and the micromotion. Up to now the case $\aleph = 0$ was considered to be rather unlikely for observation [15]. It induces necessarily n_x and n_z with opposite signs. In this radial-axial coupling case, energy is alternately exchanged between the radial and the axial components, that will lead at each time one of the motion amplitude to be increased and even larger than the trap dimensions, and then allowing ion loss by neutralization on the electrodes ("ion evaporation"). For $\aleph \neq 0$ energy could be exchanged but it is above all supplied by the rf driving field. In both cases of couplings, resonances will increase the amplitude of motion or will render it unstable. These resonances are expected for integer n_x, n_z and $\aleph$.

The radial-axial couplings reveal new regions of the stability diagram where the storage is less efficient [19]; these regions come in addition to those corresponding to the well-known regions of less efficient storage due to couplings between the macro- and the micro-motions.

2.3 Experimental observations

The number of the stored N^+ ions for different trapping conditions was measured by varying the dc voltage U_{dc} in step for different fixed V_{ac} voltages.

Fig. 2 evidences that the number of ions depends dramatically on the working point. The minima indicate that a significant part of the ions are lost during the storage time. From the values of U_{dc} and V_{ac} at which these storage minima were observed, it is possible to get approximate values of the axial and the radial secular frequencies ω_z and $\omega_{x,y}$.

More precisely, the frequency spectrum of the ion motion is obtained by applying a tickle as shown in Fig. 3 which establishes pertinently the presence of harmonics [21]. Indeed, the spectra is then formed by parametric contributions, $2\omega_z/n$, $2\omega_x/n$, $n\omega_z + p\omega_x$, the occurence and the intensities of each peaks depend on the strenght and the shape of the anharmonicities.

Figure 4 the broadening due to the individual behavior of the ions and the effect of increase of the magnitude of perturbation (a) the existence of collective oscillation.

In the spectrum from Fig.5 frequency was found to be in coincidence with the half of the axial frequency. The other ion storage minima happen for rational values of β_z (macro-micro-motion coupling case), here 2/5, 1/3 and 2/7 (x, y, and z), which is significant of distortions of the 5th, 6th and 7th orders, respectively; the 6th order showing a defect in the reflection symmetry (cf. eq. 4), as foreseen in [13–15]. Lines corresponding to N <5

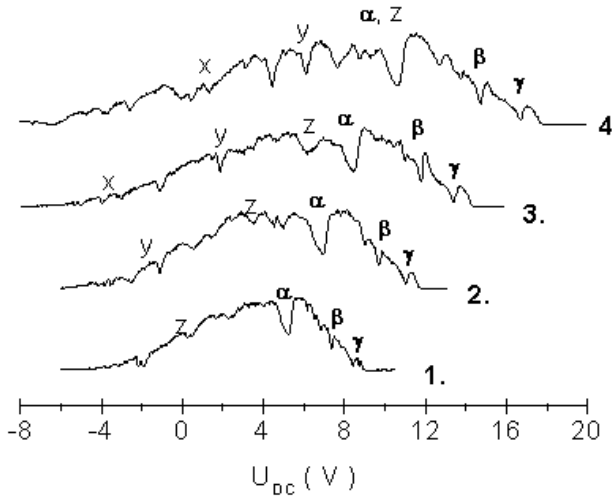

Fig. 2. Comparison of the variation of the number of stored N^+ ions for four ac voltages; 1: $V_{\rm ac}$=80$V_{\rm rms}$, 2: $V_{\rm ac}$=90$V_{\rm rms}$, 3: $V_{\rm ac}$=100$V_{\rm rms}$, 4: $V_{\rm ac}$=110$V_{\rm rms}$. The valleys are assigned as x: β_z =2/5, y: β_z=1/3, z: β_z =2/7, $\alpha : \beta_z/\beta_x = 1$, $\beta : \beta_z/\beta_x = 2/3$, $\gamma : \beta_z/\beta_x = 1/2$. Note that the loci of x ,y ,z , show constant values of β_z, the α, β, γ intersect as it would be expected because α, β, γ are constant ratios

could not be observed under our Vac conditions (as well as similar properties for β_x).

2.4 Simplified motion equation for radial-axial couplings

In order to understand the observations, we can establish new equations of motion obtained when considering the pseudopotential-well model introduced by H. Dehmelt [20] when the static potential boundaries contains higher order terms. These equations are easier to understand as exact equations and could therefore give rapidly qualitative descriptions. The major defect of such models is to suppress the presence of the rf motion, that is not crucial here, since we are not interested in macro-micro couplings. Detailed presentation of this developments are given in [22].

Assuming the cylindrical symmetry, the potential up to the fourth order is given by an expansion of the spherical harmonics Φ_n :

$$\Phi(x,y,z,t) = \Phi_0(t) + \Phi_0(t) \sum_{i=1}^{i=4} a_i \Phi_i(r,z) \qquad r = \sqrt{x^2+y^2} \tag{5}$$

with

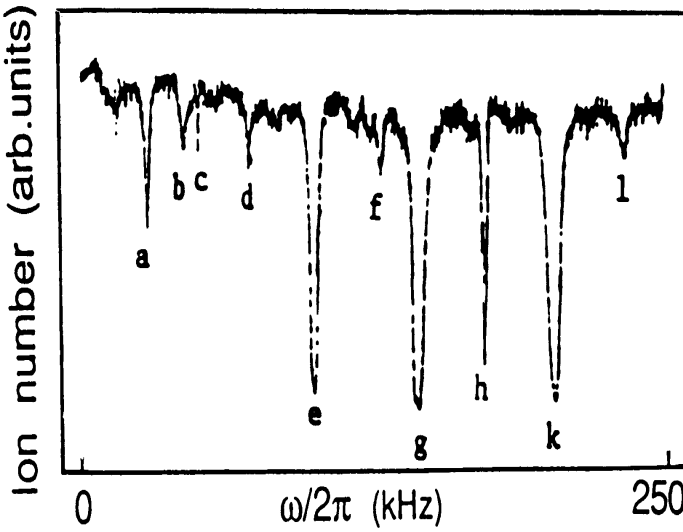

Fig. 3. Ion ejection signal after the interuption of the confinement voltage monitored after a delay with the quadrupolar excitation, a: $\omega_x - \omega_z$, b: $2\omega_x - \omega_z$,c: $\omega_z/2$, d: ω_x, e: ω_z, f: $2\omega_z - \omega_x$, g: $2\omega_x$,h: $\omega_x + \omega_z$, k: $2\omega_z$, l : $3\omega_z - \omega_x$, at the upper limit, $2\omega_x + \omega_z$

$$\Phi_0(t) = \frac{U_{\mathrm{dc}} + V_{\mathrm{ac}} \cos(\Omega t)}{2} \tag{6}$$

$\Phi(x, y, z, t)$ can be written as a sum of two functions: $\mathcal{U}$ depending only on the space coordinates and $\mathcal{V}$ depending on the space and on the time, versus a geometrical function $\mathcal{G}$.

$$\mathcal{U}(x, y, z) = \frac{U_{\mathrm{dc}}}{2} \left(1 + \mathcal{G}\left(x, y, z\right)\right) \tag{7}$$

$$\mathcal{V}(x, y, z) = \frac{V_{\mathrm{ac}} \cos(\Omega t)}{2} \left(1 + \mathcal{G}\left(x, y, z\right)\right) \tag{8}$$

$$\mathcal{G}(x, y, z) = \sum_{i=1}^{n} A_i \Phi_i(\sqrt{x^2 + y^2}, z); \quad A_i = \frac{a_i}{r_0^i} \tag{9}$$

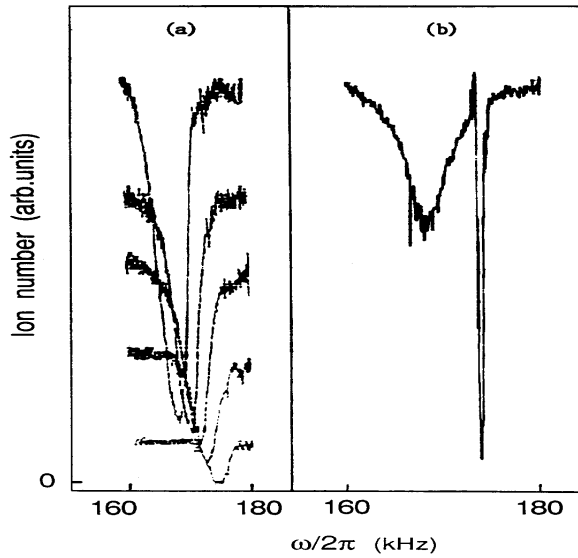

Fig. 4. (a) Axial resonance for N^+ with parametric excitation $2\omega_z$ as a function of ion number . Top to bottom: ion number decreasing by shortening the ionization pulse, (b) detailed $2\omega_z$ profile due to the space charge, a very narrow response, which occurs at a greater frequency indicates a strong collective behaviour

The stored particle is then in a steady field, suffering a force, which oscillates at a high frequency (Ω) which will induce small oscillations in the motion. Therefore, the pseudopotential method [10],[19] can be applied. Averaging the fast motion on the rf period, the slow motion equation is given by:

$$\ddot{z} = -e\frac{d\mathcal{U}}{dz} - \frac{1}{m\Omega^2}\overline{f_z\frac{\partial f_z}{\partial z}} = -e\frac{dU_{eff}}{dz} \qquad f_z = -e\frac{d\mathcal{V}}{dz} = -e\frac{V_{\rm ac}\cos(\Omega t)}{2}\frac{d\mathcal{G}}{dz} \tag{10}$$

Expanding the terms up to third order, 10 becomes:

$$m\ddot{z} + \omega_z^2\left(1 - \frac{\alpha_3}{\omega_z^2}\left(x^2+y^2\right)\right)z + \alpha_{2z}z^2 + \alpha_{3z}z^3 = -\alpha_1 + \alpha_{2r}\left(x^2+y^2\right) \tag{11}$$

$$\omega_z^2 = \frac{e}{2m}\left(2A_2U_{\rm dc} + \left(2A_2^2 + 3A_1A_3\right)\frac{eV_{\rm ac}^2}{2m\Omega^2}\right) \tag{12}$$

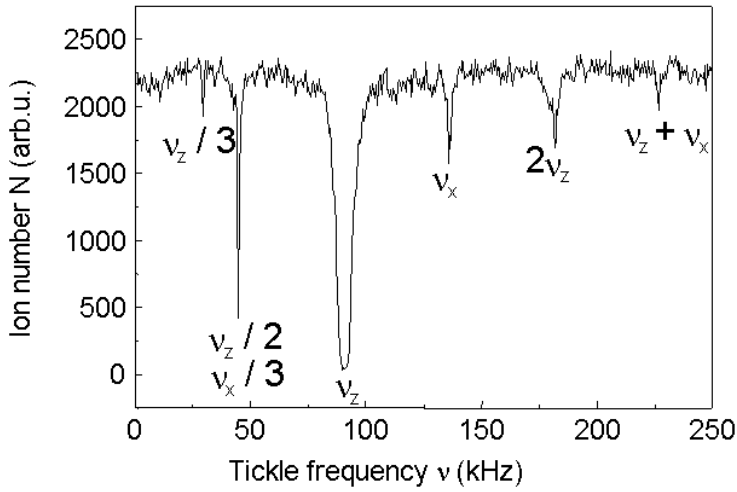

Fig. 5. Experimental frequency spectrum of N^+ ion motion for working conditions inside the valley previously assigned to $\beta_z/\beta_x = 2/3$, $\nu_z = \omega_z/2\pi$

and for one of the radial components:

$$\ddot{x} + \omega_x^2 \left(1 + \frac{\alpha_{xz}}{\omega_x^2} - \frac{\alpha_{x2y}}{\omega_x^2} y^2 + \frac{\alpha_{x2z}}{\omega_x^2} z^2\right) x + \alpha_{3x} x^3 = 0 \tag{13}$$

$$\omega_x^2 = \frac{e}{2m} \left(A_2 \frac{eV_{\mathrm{ac}}^2}{4m\Omega^2} - U_{\mathrm{dc}}\right) A_2 \tag{14}$$

where the coefficients $\alpha_{n\xi m\xi'}$ are functions of U_{dc}, V_{ac}, A_i and Ω^2

The shape of eqs. 11 and 13 show couplings between the different components of the motion. From these equations, the effects of the couplings can be foreseen. Actually, the oscillating motion of each component, due to the confinement potential, will act as a forced oscillator at the secular frequency of the involved components ω_x, ω_z). For given relations between ω_x and ω_z, resonances could occur. These resonances will lead to momentum exchange, that will increase the amplitude of the motion of one component, while the other(s) will decrease. From these alternating variations, one can imagine that the amplitude will be great enough to render the ion cloud larger (while its density will decrease), or to permit ions to hit the walls of the trap. The relations between ω_x and ω_z, for which resonances are possible can be found from the shape of the equations themselves.

In the most general case, symmetry of reflection is not rigorously respected and A_{2i+1} are not zero. Eq. (11), e.g., can be written in terms of a forced Duffing equation. From the resolution of the general equation, resonances occur if $\omega_z/\omega_x = \text{n}$ or $\frac{1}{n}$. The properties of the parametric resonance teaches that the width or the strength of the resonance is weaker when n increases, since it varies versus h^n, when we identify the term of z in eq. (11) to $\omega_z^2\,(1 + h\cos 2\omega_x t)$. However, because α_{2z} or $\alpha_2 > \alpha_{3z}$ or α_3, the particular solution, which depends on the time through x and y, should be rather considered and leads only to resonances such as $n = 2\frac{p}{q}$. The respect of the symmetry of reflection, that is $\alpha_1 = \alpha_{2z} = \alpha_{2r} = 0$, raises new kinds of resonances such as $\omega_z/\omega_x = \frac{2}{3}n$, which then could be considered as possible values. Similar deductions could be done from Eq. (11) for the ratio ω_x/ω_z which is of the form $\omega_x/\omega_z = 2\frac{p'}{q'}$. However, because of the shape of the x coefficient, the approach seems more complex. The found ratios come from $p = 1$, $q = 2, 3, 4$ and $p' = 1$, $q' = 2, 3$.

The general construction for finding resonances as the Farey tree algorithm [6,18], valid when two frequencies appears in a non linear equation, can also be used. The involved ratios as 1, 1/2, 2/3 appear on the first, second and third line of the Farey tree.

3 Application to Frequency Standards

3.1 The radiofrequency domain

Using an atomic transition of a stored ion (passive clock) to control a local oscillator was proposed by Dehmelt [23] in the early stage of the ion storage developments. Because the ion storage technique allows suppression of the first-order Doppler broadning and does not limit the interrogation time, the realisation must permit to satisfy the best possible to the fundamental two criteria for a standard: accuracy and stability. While fulfilling the first criterion needs a severe analysis of the frequency shift due to the imperfections of the systems, the second one is controlled by measuring the two-sample variance or the Allan deviation defined by [24]:

$$\sigma(\tau) = \frac{1}{2\pi f_0 \Delta t_r SNR}\sqrt{\frac{T_c}{\tau}} \qquad \tau \geq T_c \geq 2\Delta t_r \tag{15}$$

where f_0 is the frequency of the local oscillator, SNR, the signal to noise ratio, Δt_r characterises the line width, T_c the measurement cycle time and τ the duration of the probe interval.

In the microwave frequencies, the best way to reach Allan variance magnitude competitive with atomic fountains clock [25] is to work with an ion cloud which at the same time increases the SNR ratio and reduce the parasitic frequency shifts. The $^{199}Hg^+$ ion species is a good candidate with the

possibility to prepare easily the ion [26]. The clock transition at 40.5 GHz corresponds to the hyperfine separation of the fundamental level. The linear ion traps (LITS) at JPL (NASA) provide successful stability (short stability of $3 \times 10^{-14}/\sqrt{\tau}$ [27], which allows to measure that temporal variation of the fine constant is smaller than 4×10^{-14} per year [28], while the Grand Unified String Theory predicts a time variation at the level of 10^{-20} per year. Improvement of the system is now undergo by replacing the four rods of the linear trap by a judicious set of 12 round rods with a $60°/30°$ symmetry and by using an extended trap in order to separate the optical state selection region from the clock resonance region [27].

The quality of an atomic clock depends also on its accuracy. The NIST (Boulder) project develops small linear ion traps working with laser- cooled ions [30] and interrogation based on a temporal Ramsey technique. Actually , using linear ionic crystals allows to expect systematic shifts to be less than a part in 10^{16}, since reducing the micro-motion heating, the electric and magnetic field allows to minimize all the systematic effects. Nevertheless, reducing the number of ion will decreases the SNR ratio and long stability is more difficult to be reached. The Ramsey technique for the detection intoduces at this level of sophistication limit of the accuracy due to the quantum projection noise , evidenced in [29]

3.2 The optical domain

Working at higher frequency with narrow forbidden lines can be very attractive since the limit of the Allan variance is expected to be smaller. For these spectral regions, the Lamb–Dicke regime [6] or the "strong binding conditions" profile [31] requires Doppler laser cooling. Due to presence of the micro-motion, the required temperature (few mK) can only be attained with single ions stored in miniature trap or with ion chains in a linear trap. The spectrum of the considered transition is then discrete showing a central carrier and sidebands which are separated by multiples of the ion's frequencies of motion [5,32]. Numerous possibilities exist. For instance, projects with Yb^+ are developed at PTB (Braunschweig), NPL (Teddington), with Sr^+ at NPL and NRC (Ottawa), with In^+, at MPQ (Munich) [33] and Washington University, Hg^+, at NIST (Boulder), finally Ca^+ potentialities are investigated in Kobe and Marseille.

In the optical region, the error signal of the local oscillator ("clock-laser") does not result from a collective response (even from few ions as in Mercury with twelve-ion chain) but from a single ion using the Dehmelt shelving amplification scheme [34]. Actually, the observation of single ion "quantum jumps" indicates the internal quantum state of a two-level ion with 100% of confidence, assuming that the power of the laser will not couple the involved levels. For this reason, the cooling laser is stopped during the interrogation timt. However influence of the Ramsey technique in the optical domain, as

proposed at NIST must be evaluated. The measured signal is the fluorescence light corresponding to the cooling transition. When the probe laser is locked on the clock transition, which is not coupled to the cooling cycle, the fluorescence disappears. Example of quantum jumps are shown in Fig.6.

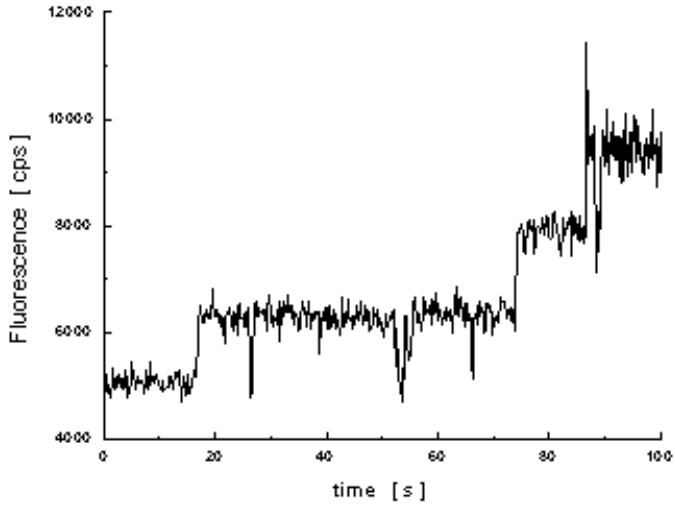

Fig. 6. Collision-induced quantum jumps with one, two and three Ca^+ laser-cooled ions in a medium-size trap at PIIM.

The efficiency of its amplification is of the order of the ratio of the strength of the lines, that is, for instance, more than 10^8. The error signal will then be built from the accumulation of the detected quantum jumps for two values of the frequencies of the clock laser corresponding to symmetric positions in the spectral line of the clock transition.

At this time, only frequency measurements (in the optical domain) with Sr^+ are published (NPL [35] and NRC [36]). The last measurements at NRC referenced to a Cs atomic clock [37] involving a diode laser present an accuracy of 200 Hz at 445 THz on the $5s^2S_{1/2}$ - $4d^2D_{5/2}$ (natural width of 46 Hz).

One major problem with the optical frequency standard is to fulfil the required condition on the local oscillator, that is here the laser linewidth which should be of the same order as the passive oscillator, which is generally about 1Hz. A noteworthy step was achieve at NIST: a two-step stabilised doubled frequency laser was able to give spectral width of 0,84 Hz (beat-note spectrum) [38]. This remarkable result is obtained from dye frequency doubled lasers, but the group is now preparing an solid-state system based upon a frequency-doubled Nd^{3+}- doped fluorapatie (Nd:FAP), the frequency

instability of this arrangment being dominated by low-frequency acoustical and mechanical noise.

Then, the suitable very high laser stability can be only reached when locking is done with very high-finesse cavities ($12000 \leq \mathcal{F} \leq 200000$) [38,35] placed in a very well controlled environment.

3.3 The Ca^+ at PIIM, Marseille

Among the ions proposed for a frequency standard in the optical domain, Ca^+ is one of the most favourable candidates due to its wavelengths accessible with all-solid-state laser systems. This allows to imagine an ultimate experimental setup which would be simple and compact [39]. The $4s^2S_{1/2} - 3d^2D_{5/2}$ electric quadrupole transition at 729 nm with a natural linewidth of 200 mHz serves as clock transition (giving rise to a Q-factor of the transition higher than 10^{15}), while laser cooling on the $4s^2S_{1/2} - 3d^2P_{1/2}$ resonance line is carried out at 397 nm (Fig. 7).

The D-doublet lifetimes were studied in detail in differents laboratory and also in Marseille where systematic investigations allowed to estimate quenching and j-mixing rate constants under collisions with different atomic and molecular neutral gases. [40]

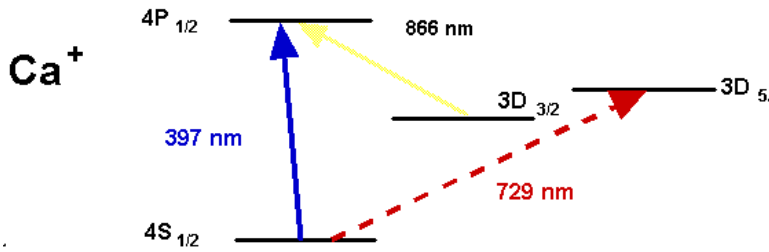

Fig. 7. Lowest energy levels of the $^{40}Ca^+$ ion

The miniature ion trap of the Paul-Straubel-type [41] consists of a cylindrical ring of 1.4 mm inner diameter and a wall thickness of 0.3mm being surrounded by four compensation electrodes (for ion positionning and electric field control), the total height of the cylinder ($2z_0$) equals 0.85mm (Fig.8). The confinement frequency is $\Omega/2\pi = 11.6$ MHz with an AC amplitude $V_{\rm ac}$ up to 1500 V_{rms}, the related frequencies of ion motion are then in the MHz range.

Ions are created by electron bombardement of a slow atomic calcium beam produced by an oven. After loading of the trap, the oven and the electron gun are turned off to avoid collisions and a high stray light level. Collimation of the ion's fluorescence is made by an aspheric lens at about 16mm away from the trap center. This very open structure provides a solid

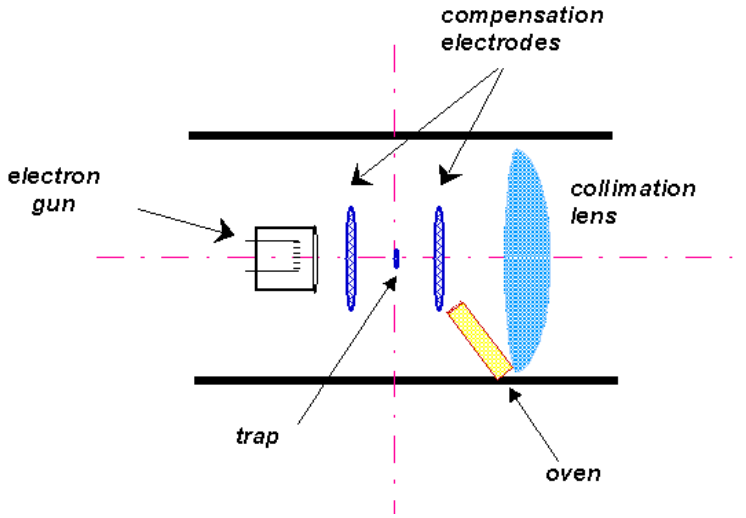

Fig. 8. Miniature trap for the Calcium project at PIIM, schematic design of the miniature trap, the compensation electrodes in x and y are not shown

angle of observation of almost 90°. The base pressure in the vessel is below 5×10^{-10} mbar as measured by a Bayard–Alpert-gauge. The composition of the partial pressures of the residual gas can be determined by a mass-spectrometer with a resolution up to 1×10^{-9} mbar.

First laser-cooled small ion clouds were obtained [42]. The cooling wavelength is obtained at this time by frequency doubling of a 794 nm Ti-Sa-laser. In Fig.9 are plotted the signals of clouds containing approximately 50 to 150 ions. These signals have been obtained with the use of a light buffer gas (p(He)=1×10^{-7} mbar) for the pre-cooling of the clouds. Different cloud sizes can easily be produced by changing the heater current of the calcium oven. The temperature of the clouds can roughly be estimated from the envelopping Doppler profile. They lie between 400 and 700K, which represents an equilibrium value between the initial ion temperature due to rf heating, the infinite heat reservoir of the buffer gas at 300K and the laser-cooling process.

Colder ions (of the order of 50K) have been trapped without buffer gas. After termination of the testing process of our trap, lower temperatures will be reached by reducing the size of the cloud in the trap.

4 Quantum Optics with Laser-Cooled Ions

Obtention of one single ion or small laser-cooled ion cloud allows one to manipulate internal states and the ion motion with external laser fields and to then perform efficient probing of the considered atomic levels. These last

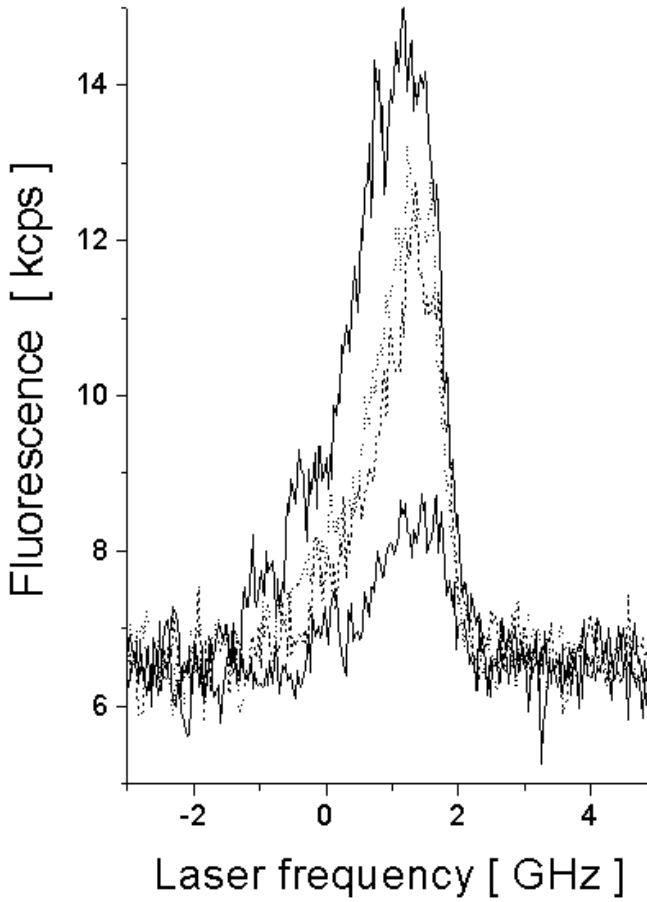

Fig. 9. Fluorescence of Ca^+ ion clouds of different sizes in the miniature trap. The different curves correspond to various values of the heater current of the calcium oven. These signals have been obtained with a moderate buffer gas pressure

years saw very rich harvest of findings thanks to such techniques. Thanks to the quantification of the motion, more performing cooling was achieved, with the resolved sideband Raman cooling [43,33]. Study of the statistics of "quantum jumps" shows photon antibunching and sub-Poissonian statistics of the spontaneous emission. Single ions are largely utilised for such illustration [44] measuring the photon correlation statistics with a Hanbury–Brown–Twiss experiment and experimentally evidencing the predicted value, zero, of the correlation function in such conditions. Interference fringes with light scattered from two ions were observed with mercury species [45], the fringe pattern can be used to investigate the degrees of ion localization the dis-

tance between the two ions. Generation of non classical states of ion motion as Fock states was successively succeeded in a standing wave , via observation of quantum jumps [48], via adiabatic passage [49], or via trapping states [50] and also in travelling waves on Raman transition [51]. Coherent states has been prepared and the related Poissonian distribution of the Fock states measured [51].

Preparation of Schrödinger cats [52] and entangled states [46] shows that cold ions provide a realistic physical system for the implementation of a quantum computer. Indeed, delicate preparations using several steps for manipulating both internal and external states allows to success for the realisation or "controlled-NOT" gate [43,47], when the decoherence is controlled.

5 Conclusion

Ion storage technique provide a large range of example of individual and coupled oscillators. The couplings between them can be done macroscopically, it gives good examples of non linear dynamics and can be applied for sophisticated high resolution mass spectrometry methods [53].

A very large review on physics and instrumentation with ion traps can be found in [54]. In the quantum domain, entangled states are formed with ion chain characterised by their common frequencies. The challenge to exploit these new quantum systems is continuously increasing, as the last propositions to use entangled states for optical frequency standards [55] or to propose atomic interferometry [56] show it.

References

1. R.C.Thompson, Spectroscopy of Trapped Ions, Adv. Atom. Molec. Opt. Phys. **31**, 63 (1993) and R. Blatt, Atomic Physics **14**, D.J. Wineland , C.E. Wieman and S.J. Smith (eds), Am. Inst. of Physics, New York (1995).
2. W. Paul, O. Osberghaus and E. Fischer, Forschungsberichte des Wirtschafts- und Verkehrs-ministeriums, Nordrhein-Westfalen, n°415, West-deutscher Verlag, Kvln und Opladen (1958).
3. M. Vedel, M. Knoop, D. Lunney, I. Rebatel and F. Vedel, Phys. Rev. **A 51**,2294 (1995).
4. J. Rocher, M. Vedel and F. Vedel, Int. J. Mass Spectrom. Ion Processes,**181**,173 (1998).
5. F.G. Major and J.L. Duchêne, J. Phys. **36**, 953 (1975).
6. R.H. Dicke, Phys.Rev. **89**, 472-473 (1953).
7. A.J. Lichtenberg and M.A. Lieberman, "Regular and Stochastic Motion", Springer, NY. (1983).
8. F. Vedel, Int. J. Mass Spectrom. Ion Processes, **106**,33 (1991).
9. H.G. Schuster, Deterministic chaos, VCH Verlag, Weinheim (1989).
10. L. D. Landau and E. M. Lifchitz, Mechanics, Pergamon Press, N.Y. (1976).
11. G. Kotowski, Z. Angew. Math. Mech. **23**, 213 (1943).

12. H. Dawson: "Quadrupole Mass Spectrometry", Elsevier, Amsterdam (1976).
13. R.Alheit, C.Hennig, R.Morgenstern, F.Vedel and G.Werth, Appl. Phys. **B 61**, 277 (1995).
14. R. Alheit, S. Kleineidam, F. Vedel, M. Vedel and G. Werth, Int. J. Mass Spectrom. Ion Processes, **154**, 155 (1996).
15. Y. Wang, J. Franzen, K. P. Wanczek, Int. J. Mass Spectrom. Ion Processes. **124**, 125 (1993).
16. F.Guidugli and P.Traldi, Rapid Comm. Mass Spectrom. **5**, 491 (1991), and F.Guidugli, P.Traldi, A.M.Franklin, M.L.Langford, J.Murell and J.F.J. Todd, Rapid Commun.Mass Spectrom. **6**, 229 (1992).
17. K.L.Morand, S.A.Lammert and R.G.Cooks, Rapid Comm. Mass Spectrom. **5**, 491 (1991).
18. F.v. Busch, W. Paul, Z. Phys. **164**, 581 (1961).
19. M. Vedel, J. Rocher, M. Knoop and F.Vedel, Appl. Phys. **B 66**, 191 (1998).
20. H. G. Dehmelt, Adv. At. Mol. Mol. Phys. **3**, 53 (1967).
21. F. Vedel and M. Vedel, Phys. Rev. **A 41**, 2348-2351 (1990).
22. J. Rocher, Thèse de l'Univerisité de Provence, Marseille (1998)
23. H. Dehmelt, B.A.P.S. . 18, 1521 (1973), I.E.E.E. Trans. Instrum. Meas. **IM-31**, 83 (1982).
24. D.W. Allan, I.E.E.E., Trans. Ultrason., Ferroelectric, Freq. Contr. **34**,647 (1987).
25. E. Simon , P. Laurent and A. Clairon, Phys. Rev. **A 57**, 436 (1998).
26. M. Jardino and M. Desaintfuscien, I.E.E.E. Trans. Instrum. Meas. **IM-29**, 163 (1980).
27. J. D. Prestage, R.L. Tjoelker and L. Maleki, CP457, Trapped charged particles and fundamentals physics, p. 357, ed. D. H.E. Dubin and D. Schneider, A. I. P. (1999).
28. J. D. Prestage, R.L. Tjoelker and L. Maleki, Phys. Rev. Lett. **74**, 3511 (1995).
29. W.M. Itano et al.,, Phys. Rev.**A 47**, 3554 (1993) **74**, 3511 (1995).
30. D.J. Berkeland et al., Phys. Rev. Lett. **80**, 2089 (1998).
31. D.J. Wineland, Phys. Rev. **A 20**,1521 (1979).
32. D.J. Wineland, W.M. Itano, J.C. Bergquist, and R.G. Hulet, Phys. Rev. **A 36**, 2220-2232 (1987).
33. E. Peik, J. Abel, Th. Becker, J. von Zanthier, and H. Walther, Phys. Rev. **A 60**, 439 (1999)
34. W. Nagourney, JH. Sandberg and H. Dehmelt, Phys. Rev. Lett. **56**, 2797(1986).
35. G. P. Barwood, P. Gill, H. A. Klein and W. R. C. Rowley, I.E.E.E. Trans. Instrum. Meas. **46**, 135 (1997).
36. J.E. Bernard, L. Marmet and A. Madej, Opt. Comm. **150**, 170 (1998).
37. J.E. Bernard, A. Madej, et al. Phys. Rev. Lett, **82**, 3228 (1999).
38. B.C. Young et al., CP457, Trapped charged particles and fundamentals physics, ed. D. H.E. Dubin and D. Schneider, The American Institute of Physics (1999), more recent performances are given in B.C. Young et al.,Phys. Rev. Lett, **82**, 3799 (1999). .
39. M. Knoop, M. Vedel and F. Vedel, J. Phys. **II**, 4, 1639 (1994).
40. M. Knoop, M.Vedel, F.Vedel, Phys.Rev.**A 52**, 3763 (1995), for a review see E. Biemont and C.J. Zeippen, Comments At. Mol. Phys. **33**, 29 (1996) and M. Knoop, M.Vedel, F.Vedel, Phys.Rev.**A 58**, 264 (1995) and ref. therein.
41. C.A. Schrama, E. Peik, W.W. Smith, H. Walther, Opt.Comm.**101**, 32 (1993)

42. M.Knoop, M.Vedel, M.Houssin, M. Herbane, T.Pawletko, F.Vedel, Proceedings of the Joint Meeting of the 13th European Forum of Time and Frequency and the 1999 IEEE International Frequency Control Symposium, Besangon, avril 1999, in press.
43. C. Monroe et al., Phys. Rev. Lett. **75**, 4714 (1995).
44. F. Diedrich et al, Phys. Rev. Lett **58**, 203 (1987).
45. U. Eichmann, et al. Phys.Rev.Lett.**70**, 2359 (1993)
46. Q.A. Turchette et al. LANL e-print archive quant-phys/9806012 (1998).
47. I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).
48. I. Cirac, R. Blatt, A.S. Parkins and P. Zoller, Phys. Rev. Lett. **70**, 762 (1993).
49. I. Cirac , R. Blatt and P. Zoller, Phys. Rev. **A 49**, R3174 (1994).
50. R. Blatt, I. Cirac and P. Zoller, Phys. Rev. Lett. **52** ,518 (1995).
51. D.M. Meekhof et al., Phys. Rev. Lett. **76**, 1796 (1996).
52. C. Monroe et al, Science, **272** (1996).
53. "Electrodynamic Ion traps" A special issue honoring the carreers and contributions of John Todd and Ray March, Int. J. Mass Spectrom. Ion processes, in press
54. P. K. Ghosh, Ion traps, Clarendon Press, Oxford (1995).
55. S.F. Huelga, P.L. Knight, C. Macchiavello, M.B. Plenio and V. Vedra. Appl. Phys. **B 67**, 723 (1998).
56. R. Huesman, Ch. Balzer, Ph. Courteille, W. Neuhauser and P.E. Toschek, Phys. Rev. **82**, 1611 (1999).

$1/f$ Fluctuations in Cosmic Ray Extensive Air Showers

E. Faleiro[1] and J.M.G. Gómez[2⋆]

[1] Departamento de Física Aplicada e I.S., E.U.I.T. Industrial, Universidad Politécnica de Madrid, E-28012 Madrid, Spain
[2] Departamento de Física Atómica y Nuclear, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain

Abstract. The fluctuations of the particle density distributions in extensive air showers have been studied at ground level. In order to achieve meaningful statistics, the interaction of cosmic rays with the earth atmosphere has been simulated by means of the CORSIKA Monte Carlo code. It is shown that the fluctuations of the particle density distributions as a function of the polar angle have features typical of a $1/f$ noise. The sample is then analysed in order to study its scaling behaviour and we find that it can be parametrized by means of a universal multifractal approach.

1 Introduction

Since the discovery of cosmic rays at the beginning of this century, their study has provided substantial contributions to physics which could not have been obtained otherwise, because it is not possible to achieve so high particle energies in laboratories. Traditional questions related to cosmic rays are, for example, Where do these particles come from? What is their age? How did they get such a high energy? ... But in this work we pose different questions: What kind of statistical fluctuations are observed in cosmic rays? Do they contain any information?

The interaction of cosmic rays with the high altitud atmosphere produces secondary particles, and successive interactions of the primary cosmic rays and their secondaries give rise to multiplicative cascades called Extensive Air Shower (EAS). For a good survey of EAS physics see for example ref.[1].

Current models of the particle density in an EAS describe average values and it is usually assumed that the fluctuations around the mean must have zero mean value and are totally independent, i.e., they should show no correlation structure. Thus for example the widely used NKG[2] formula describes the particle density as a function of the distance to the core, but it is assumed to be independent of the polar angle. In this paper we study the characteristics of fluctuations in the particle density of EAS at ground level. We shall see that these fluctuations are not a simple white noise. They contain a complex structure which can be interpreted as a white noise plus a $1/f$ noise.

⋆ gomezk@nuclear.fis.ucm.es

M. Planat (Ed.): LNP 550, pp. 125–136, 2000.
© Springer-Verlag Berlin Heidelberg 2000

Furthermore, they exhibit statistical selfsimilarity properties which can be quantified by means of the universal multifractal formalism.

We note that similar structures have been found as well in the analysis of fluctuations in the semiinclusive *rapidity* distribution in high-energy hadron-hadron collisions and scaling laws have been observed in particle density and correlations functions in high-energy multiparticle dynamics[3,4].

We shall first give a brief introduction to the physics of cosmic rays and Extensive Air Showers (EAS) in section 2, and we describe in section 3 the databasis of EAS events obtained by Monte Carlo simulation for statistical studies. In section 4 we present an analysis of the statistical fluctuations in the secondary particle density of EAS events at ground level, and in section 5 we analyse their multifractal structure. Finally, the conclusions are summarized in section 6.

2 Extensive Air Showers

The high energy particles and nuclei arriving at the Earth from space are collectively called *cosmic rays*. They consist mainly of protons and α particles, an appreciable fraction of heavier atomic nuclei, a small fraction of electrons ($\sim 0.1\%$) and an even smaller fraction of photons ($\sim 0.01\%$). The energy spectrum of cosmic rays varies from 10^9 eV to 10^{20} eV. The latter particle energies are very seldom, since the flux of cosmic rays decreases quickly with increasing energy, following a power law $E^{-\gamma}$.

Cosmic rays can be directly detected in balloons and satellites up to $\sim$ 100 TeV for charged particles, or up to $\sim$ 30 GeV for the less abundant γ rays. For higher energy particles the flux is so small that direct detection and measurements become practically impossible. The main experimental technique is then based on detection at ground level of the *secondary* particles arising from the collision of *primary* cosmic rays with the earth atmosphere.

Fig. 1 illustrates how an electromagnetic EAS arises from a primary γ ray. In the neighbourhood of an atmospheric nucleus, the photon materializes creating a positron and an electron pair. Each of these high energy leptons produces bremsstrahlung γ rays, which lead again to pair production, and the process continues quickly producing an electromagnetic shower.

The physical processes giving rise to a hadronic EAS are also shown in Fig.1. A cosmic hadron collides with an atmospheric nucleus producing mainly pions and possibly also nucleon and other mesons. The π^0, π^+ and π^- particles are produced in approximately the same quantities. The π^0 decay spontaneously (mean life 1.78×10^{-16} s), by electromagnetic interaction, into two γ rays, which again produce new electromagnetic showers. The charged pions decay by weak interaction (mean life 2.55×10^{-8} s) according to the reactions

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \ , \tag{1}$$

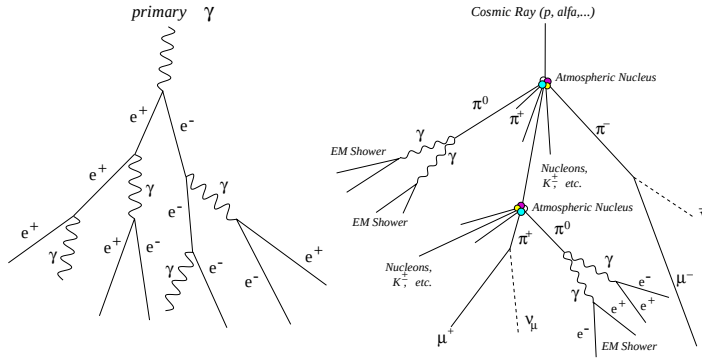

Fig. 1. Schematic diagram of the basic reactions producing a pure electromagnetic shower iniciated by a primary γ ray, and a hadronic shower iniciated by a proton or some other nucleus.

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \ . \tag{2}$$

The muons decay according to the reactions

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \ , \tag{3}$$

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \ , \tag{4}$$

with a mean life of $2.2 \cdot 10^{-6}$ s in their own reference frame. However, for an external observer the mean life is multiplied by the Lorentz factor. Thus only muons with a Lorentz factor larger than 20 survive a flight from the upper atmosphere to the ground before they decay. Therefore the highest energy muons provide information on the hadronic EAS component and on the primary cosmic ray.

Fig. 2 illustrates an extensive air shower in which a considerable number of secondary particles reach the earth surface, where they are detected. At this point the secondaries are mostly e^+, e^-, γ and μ, and the figure shows a grid of scintillator counters, Cerenkov telescopes and other detectors.

3 Simulation of EAS Events

The detailed data needed to study fluctuations of the EAS particle density at ground level were obtained by a Monte Carlo simulation using the CORSIKA[5] code. The observation level was set to 2200 meters above sea level, corresponding to the altitude of the Roque de los Muchachos Observatory in the Canary Islands, where the HEGRA[6] experiment is located. Particles were tracked down to this height and the parameters of e^+, e^- and γ rays at their impact point with the surface were stored. The rest of the EAS

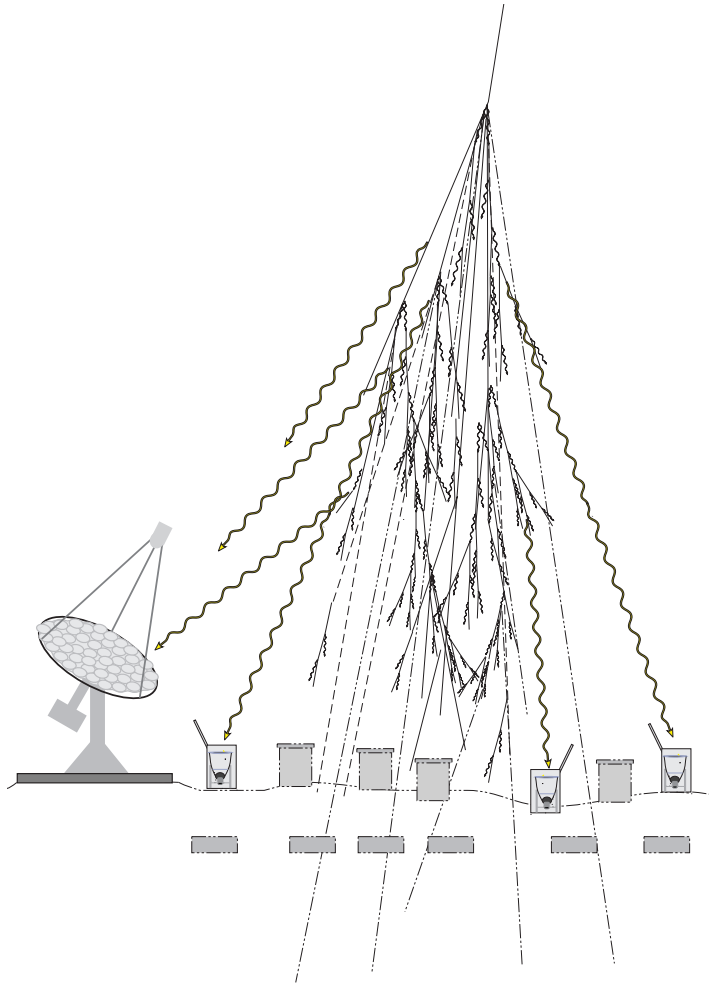

Fig. 2. Schematic extensive air shower reaching a grid of different photon and charged particle detectors placed at ground level, and underground muon detectors.

components, particularly μ and Cherenkov photons, were not used at this point. With the mentioned setup we generated over 1000 γ rays and over 2000 protons with energies ranging from 10 to 50 TeV. As a simplification, all the showers have been generated at 90^0 incidence angle with respect to the earth surface.

Fig.3 shows some examples of EAS particle distributions at the observation level. Two events generated by 10 TeV primary photons having their first collision al different altitudes are shown, and similarly two proton events are also shown. It can be seen that the hadronic EAS contain more inhomogeneities than the electromagnetic EAS particle distributions.

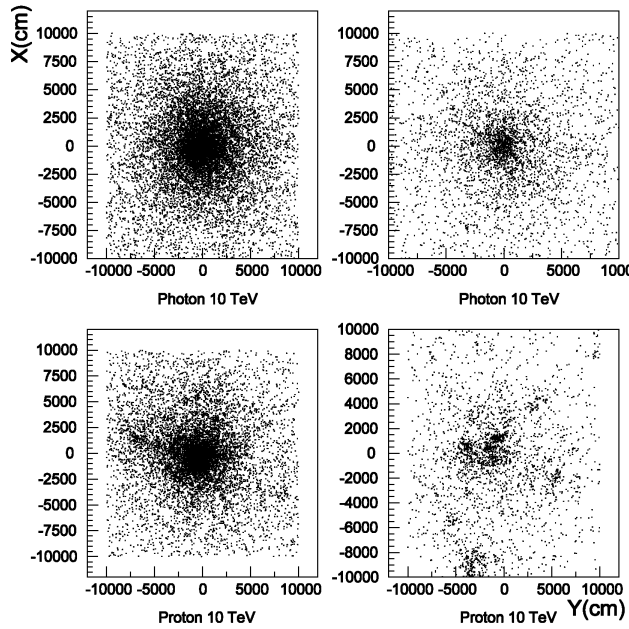

Fig. 3. Examples of secondary particle distributions at the observation level in EAS generated by 10 TeV cosmic rays. Two primary photon and two primary proton events are shown. The core of the shower is at the geometrical centre.

We have taken as a starting point for our analysis the particle densities at ground level. To gain in simplicity and get a clearer physical view of the problem, we have reduced it to a one dimensional case by considering the density of particles as a function of the azimuthal angle in a ring, comprised between radii r_{min} and $r_{\max}$, around the shower axis. The inner radius $r_{\min}$, chosen at 50 meters, is determined by the convenience to be sufficiently far away from the core of the shower, so that the uncertainties in the very forward region of the cross sections do not influence the results significantly. The outer radius $r_{\max}$ was taken as 100 meters in order to have a sufficient number of particles and thus reduce the importance of Poisson fluctuations in the number of particles. Nevertheless we have tested that changing the values of $r_{\min}$ and $r_{\max}$ in a wide range esentially does not alter our conclusions.

Table I shows the distribution of primary protons and γ rays according to their number of secondary particles arriving to the ring, classified into bins of size 2500 particles. With this databasis we achieve reasonable statistics for EAS events giving up to about 15000 secondary particles in the ring.

Table 1. Main features of the Monte Carlo simulation of primary photon and proton samples ranging from 10 to 50 TeV. Here N_p is the number of primary particles whose number of secondary particles arriving to the ring is within the bin ΔN_s; A and C are the parameters of the power spectrum density χ^2 fit, and α and C_1 are the universal multifractal parameters

Primary	N_p	ΔN_s	A	C	χ^2_r	α	C_1
proton	706	0−−2500	0.46	0.0192	1.3	1.629	1.57×10^{-2}
	572	2500−−5000	1.22	0.0534	1.2	1.999	6.04×10^{-3}
	450	5000−−7500	1.99	0.0906	1.2	2.012	3.33×10^{-3}
	284	7500−−10000	2.49	0.128	1.0	2.011	2.26×10^{-3}
	186	10000−−12500	3.3	0.161	1.3	2.011	1.81×10^{-3}
	104	12500−−15000	3.4	0.203	1.0	2.005	1.24×10^{-3}
	75	15000−−17500	4.4	0.236	1.2	2.005	1.16×10^{-3}
γ	514	0−−2500	0.054	0.0252	1.0	1.870	1.51×10^{-3}
	359	2500−−5000	0.119	0.0587	1.6	1.969	6.23×10^{-4}
	201	5000−−7500	0.21	0.1005	1.3	1.982	3.75×10^{-4}
	161	7500−−10000	0.31	0.143	1.1	2.003	2.73×10^{-4}
	110	10000−−12500	0.37	0.178	1.0	2.008	2.10×10^{-4}
	119	12500−−15000	0.58	0.218	1.3	1.995	2.18×10^{-4}
	79	15000−−17500	0.50	0.263	1.2	1.995	1.43×10^{-4}

In order to study the fluctuations of the particle density in the ring, it was divided in 256 equal sectors in azimuthal angle and the number of particles in each sector was computed. Thus an event is characterised by a primary particle and the array of its secondary particle density distribution in the ring sectors.

4 Analysis of Fluctuations

Statistical analysis of fluctuations normally considers each event as the concrete realization of a stochastic process whose statistical properties are being determined. The stochastic process could be defined by the set of all primaries of the same type and energy, or even by the totality of events generated. Each process is then characterised by its statistical parameters, which may be studied in the coordinate or in the frequency domain. In this work, although the spatial coordinate is the polar angle around the shower axis, we use instead the ring sector number t.

The autocorrelation information about a process is contained in the autocorrelation function $R(t, t+\tau) =< \epsilon(t)\epsilon(t+\tau) >$, where $\epsilon(t)$ is the content of sector t and the symbol $< . >$ indicates ensemble averaging[7]. For a special class of processes, the so-called *wide sense stationary processes* (WSS), it simplifies to $R(\tau)$. In this case it is related to the mean power spectrum density (PSD) denoted $P(k)$, by a Fourier transformation[8]

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(k) e^{ik\tau} dk \ , \tag{5}$$

where τ is a variable that here represents the separation between sectors. The PSD is usually estimated from

$$P(k) \approx \langle |g(k)|^2 \rangle \ , \tag{6}$$

where angular brackets represent again the ensemble averaging of the quantity inside them, and $g(k)$ stands for the amplitude corresponding to frequency k in the Fourier transform of the data.

In fig. 4 the natural logarithm of $P(k)$ for our events, is represented versus $\log k$. The plots show two important facts. In the first place, instead of finding a constant $P(k)$, as would be the case for a signal whose fluctuations were uncorrelated, a PSD showing a complex structure is found. As a second remark, the shape of $P(k)$ suggest a simple model of functional form

$$P(k) \sim \frac{A}{k^B} + C \ . \tag{7}$$

The constant term C represents a white noise which is clearly seen in the constant behaviour at high k, while the first term stems for the nearly linear behaviour in $1/k$ seen at low k. The appearance of a plateau in the power spectrum for high k is related to the discrete nature of collecting secondary particles in the sectors. If N_e is the average number of arriving particles per sector, for a white Gaussian noise we have $C = \sigma^2 \equiv N_e$. Thus $P(k)$ becomes constant for large k and represents the average particle density. The direct estimation of B is difficult, specially for photons showers, due to the limited sample size at low k. A direct fit gives values close to 1. To test the consistency of the data with the proposed function, we have filtered the constant component using a real non random filter with transfer function

$$h(k) = \sqrt{1 - \frac{C}{\frac{A}{k} + C}} \ . \tag{8}$$

The filter is a variant of the Wiener filter[8], which is known to be adecuate for elimination of the white noise in the signal provided that it is uncorrelated to the $1/f$ component. After the filter is applied, a power spectrum analysis of the residual signal confirms the value $B \simeq 1$. Data have therefore been

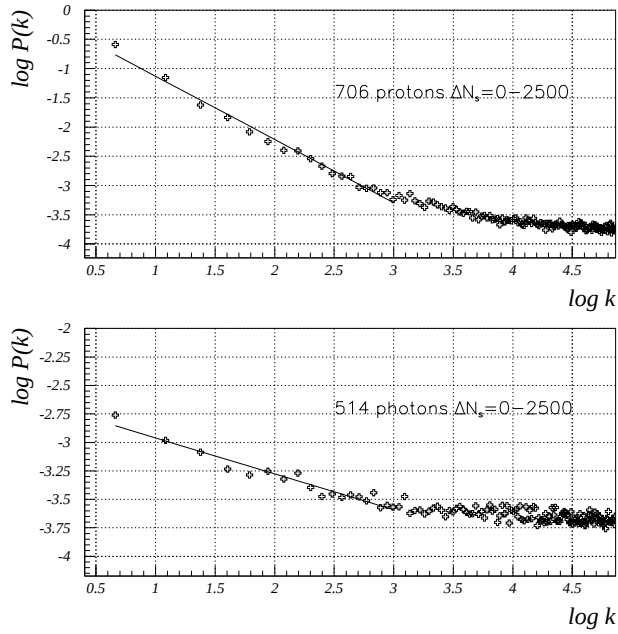

Fig. 4. Power Spectrum Density P(k) for two of the considered samples. The spatial frequency k is in fundamental frequency units.

fitted to the model PSD

$$P_{\mathrm{mod}}(k) \sim \frac{A}{k} + C \ . \tag{9}$$

Table I contains the best fit parameters A and C and the chi square per degree of freedom χ_r^2 obtained for our model power spectrum. The fit errors of A and C affect the last digit of the values given in the table. We see that the model power spectrum (9) fits well the data. Thus it can be concluded that the fluctuations of the particle density are composed of a white noise plus a noise $1/f$ which gives rise to its complex structure. This kind of noise is usually called a *flicker noise* in the literature [9]. The value found for the parameter B, approximately one, shows that the scaling component of the signal may *a priori* be assumed to be *conservative* within the multiplicative cascade model that we should now propose [10–12]. Finer corrections could later be included in further work.

5 Multifractal Analysis of the 1/f Noise

Provided that the $1/f$ noise found in data can be approximately considered conservative, we study here the scaling properties of its statistical moments $\langle \epsilon_\lambda^q \rangle$, where ϵ_λ is the content of a generic sector at resolution λ and $q > 0$ is a real mumber. The signal at resolution $\mu < \lambda$ is obtained by a simple coarse-graining procedure that is represented by [12,13]

$$\epsilon_\mu = \frac{1}{\mu_r} \sum_i \epsilon_\lambda \tag{10}$$

where μ_r is the ratio of scales and the sum is taken over the sectors at scale λ, covering the considered sector at scale μ. The scaling law for statistical moments in a multifractal scheme states that [11,12]

$$\langle \epsilon_\lambda^q \rangle \sim \lambda^{K(q)} \tag{11}$$

where the $K(q)$ function characterises the scaling behaviour and also the statistics of the data and it is called *moment scaling function*. The non linearity of $K(q)$ involves a multiscaling description corresponding to a nontrivial power law. Fig. 5 shows the empirical function $K(q)$ obtained from (11), using the average $\langle \epsilon_\lambda^q \rangle$ calculated over all the events and sectors at scale λ in a bin ΔN_s. It is clearly seen that $K(q)$ is a smooth non linear function and therefore the particle density distribution has a multifractal structure.

Let us now see whether this multifractal fits into the universal multifractal scheme [11,12]. In this scheme the theoretical value of $K(q)$ is given by

$$K(q) = \frac{C_1}{\alpha - 1}(q^\alpha - q) \ , \ \alpha \neq 1 \tag{12}$$

with $0 \leq \alpha \leq 2$, called Levy index, and $0 \leq C_1 \leq 1$ for unidimensional data, called mean fractal inhomogeneity measure. Eq.(12) arises considering the empirical signal as an outcome of a multiplicative stochastic process governed by Levy stable probability distributions of index α. To determine the indices α and C_1 we applied a direct analysis technique called DTM (Double Trace Moment technique [3,14]) to the data. The theoretical $K(q)$ given by (12) is shown in fig. 5. The agreement with the empirical $K(q)$ is very good. In all the other bins ΔN_s the agreement is also good at least up to $q = 4$.

The parameters of the DTM analysis are given in table I. We can see that both primary types show an extremely low intermittency, corresponding to a great homogeneity. On the other hand, the value of the α parameter, near $\alpha = 2$, shows a typical log-normal behaviour. The small values of C_1 obtained might suggest that we are dealing with nonconservative multifractals. However, calculations of the first order structure function exponent $\zeta(1)$, give values about 0.12, corresponding to quasi-conservative multifractals. Using this exponent, we performed a fractional derivation of the signal followed by

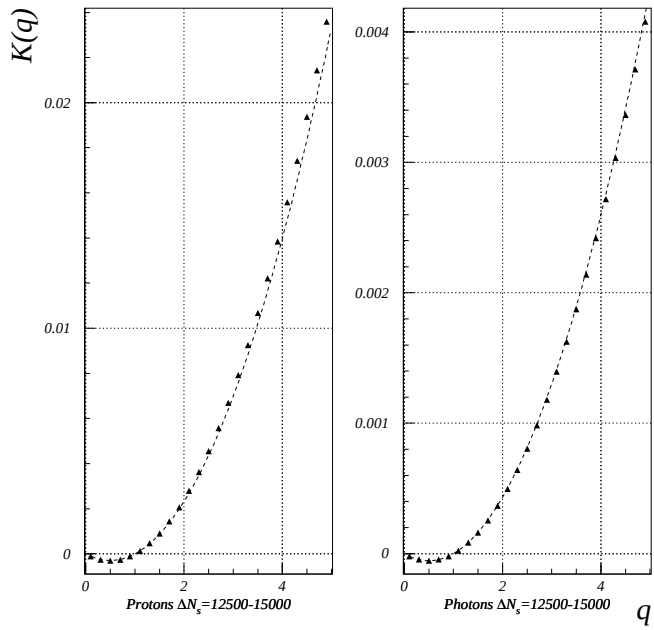

Fig. 5. Comparison of the empirical moment scaling function $K(q)$ (triangles) for two of the samples, with the theoretical value in the universal multifractal scheme (dashed line).

a DTM reanalysis, and we found that α does not substantially change and C_1 changes roughly by a factor of 2 and thus it is still very small.

In order to check the reliability of the values obtained for α and C_1, they were used in a multiplicative cascade simulation following the method of ref. [15], and finally a white noise was added. The power spectrum of the resulting signal is very similar to the $P(k)$ shape shown in fig. 4.

Finally, from the equation [11,12,16]

$$B = 1 - K(2) \ , \tag{13}$$

it is seen that the low value of $K(2) \simeq 0$ is coherent with the value found for the PSD exponent, $B \simeq 1$, corresponding to $1/f$ noise.

6 Conclusions

Using Monte-Carlo simulations, we have generated large sets of extensive air shower events arising from the interaction of high-energy (10 to 50 TeV) protons and γ rays with the earth atmosphere. The density distribution of

secondary particles at ground level has been analysed in detail as a function of azimuthal angle in a ring centered at the core of the EAS.

It is found that the power spectrum density $P(k)$ of the particle distribution fluctuations behaves as a flicker noise plus a white noise. When the white noise component is filtered, the remaining fluctuations correspond quite accurately to a $1/f$ noise.

The scaling properties of the particle density statistical moments have been studied. The moment scaling function $K(q)$ is found to be nonlinear and is thus characterised as a multifractal structure. Furthermore, the moment scaling function is shown to fit well into a universal multifractal scheme. The Levy index α and the mean fractal inhomogeneity measure C_1 have been calculated for all the samples. Their values depend on the energy of the primary particle, but this dependence is found to be very similar for protons and γ rays.

Acknowledgements

This work is supported in part by DGES and CICYT projects No. PB96-0604 and No. AEN96-1676.

References

1. Sokolsky P. (1989) *Introduction to Ultrahigh Energy Cosmic Ray Physics*, Addison Wesley Redwood City, CA.
2. Greisen K. (1956) in Prog. Cosmic Ray Phys. **3**, 3.
   Kamata K. and Nishimura J. (1958) in Prog. Theoretical Phys. Suppl. **6**, 93.
3. Ratti S. P. et al. (1994) Z. Phys. C, **61**, 229.
4. De Wolf E. A. et al. (1996) Phys. Rept. **270**, 1.
5. Capdevielle J.N. et al. (1992) in KfK Report 4998 (Kernforschungzentrum, Karlsruhe).
6. HEGRA Collaboration (1997) in Proceedings of the 25th International Cosmic Ray Conference of Durban **5**, 113.
7. Feller W. *An introduction to probability theory and its applications* Vol.II (2nd ed.), Wiley.
8. Peebles P.Z. (1987) *Probability, Random Variables and Random Signal Principles* (2nd ed.), McGraw-Hill.
9. M.B. Weissman (1988) Rev. Mod. Phys. **60**, 537.
10. Davis A. et al. (1994) J. Geophys. Res. **99**, D4, 8055.
11. Schertzer D. and Lovejoy S. (1991) in Scaling, Fractals and Nonlinear Variability in Geophysics, edited by Schertzer D. and Lovejoy S. (Kluwer, Holland), 41.
12. Schertzer D. and Lovejoy S. (1993) in Nonlinear Variability in Geophysics 3: Scaling and multifractal processes. (Lecture Notes NVAG3, EGS).
13. Schertzer D. et al. in XXII Colloque GRETSI, 1313–1325.
14. Lavallee D. (1991) in Ph.D. Thesis, University McGill (Montreal, Canada).

15. Schertzer D. and Lovejoy S. (1989) in Nonlinear Variability in Geophysics: Multifractal analysis and simulation. (Fractals: Physical Origin and Consequences), edited by L. Pietronero (Plenum, New York), 49-79.
16. Schertzer D. and Lovejoy S. (1987) J. Geophys. Res. **92**, 9693.

Stochastic Resonance and the Benefit of Noise in Nonlinear Systems

François Chapeau-Blondeau *

Laboratoire d'Ingénierie des Systèmes Automatisés (LISA), Université d'Angers, 62 avenue Notre Dame du Lac, 49000 Angers, France.
web : http://www.istia.univ-angers.fr/∼chapeau/

Abstract. Stochastic resonance is a nonlinear effect wherein the noise turns out to be beneficial to the transmission or detection of an information-carrying signal. This paradoxical effect has now been reported in a large variety of nonlinear systems, including electronic circuits, optical devices, material-physics phenomena, neuronal systems, chemical reactions. Stochastic resonance can take place under various forms, according to the types considered for the noise, for the information-carrying signal, for the nonlinear system realizing the transmission or detection, and for the quantitative measure of performance receiving improvement from the noise. These elements will be discussed here so as to provide a general overview of the effect. Various examples will be treated that illustrate typical types of signals and nonlinear systems that can give rise to stochastic resonance. Various measures to quantify stochastic resonance will also be presented, together with analytical approaches for the theoretical prediction of the effect. For instance, we shall describe systems where the output signal-to-noise ratio or the input–output information capacity increase when the noise level is raised. Also temporal signals as well as images will be considered. Perspectives on current developments on stochastic resonance will be evoked.

1 Introduction

When linear coupling takes place between signal and noise, usually the noise acts as a nuisance degrading the signal. In contrast, when certain types of nonlinear interaction take place between signal and noise, there may exist a possibility of cooperation between the signal and the noise. The presence of the noise then becomes beneficial to the signal, up to a point where an increase of the noise may improve the performance for transmitting or detecting the signal. Stochastic resonance (SR) designates this type of nonlinear effect whereby the noise can benefit to the signal [55,73].

This paradoxical effect was first introduced some twenty years ago in the domain of climate dynamics, as an explanation for the regular recurrences of ice ages [2]. Since this origin, SR has been largely developed and extended to a broad variety of domains [54,26,53]. Today, it is possible to summarize the various forms observed for SR by means of the scheme of Fig. 1.

* chapeau@univ-angers.fr

M. Planat (Ed.): LNP 550, pp. 137–155, 2000.
© Springer-Verlag Berlin Heidelberg 2000

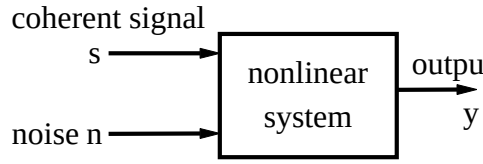

Fig. 1. A general scheme for stochastic resonance, which consists in the possibility of increasing the "similarity" between the information-carrying or coherent input signal s and the output signal y by means of an increase of the level of the noise η

Stochastic resonance, as illustrated by Fig. 1, involves four essential elements:
(i) an information-carrying or coherent signal s: it can be deterministic, periodic or non, or random;
(ii) a noise η, whose statistical properties can be of various kinds (white or colored, Gaussian or non Gaussian, ...);
(iii) a transmission system, which generally is nonlinear, receiving s and η as inputs under the influence of which it produces the ouput signal y;
(iv) a performance or efficacy measure, which quantifies some "similarity" between the output y and the coherent input s (it may be a signal-to-noise ratio, a correlation coefficient, a Shannon mutual information, ...).
SR takes place each time it is possible to increase the performance measure by means of an increase in the level of the noise η. Historically, the developments of SR have proceeded through variations and extensions over these four basic elements.

From the origin, and for a relatively long period of time, SR studies have concentrated on a *periodic* coherent signal $s(t)$, transmitted by nonlinear systems of a *dynamic* and *bistable* type [27,52]. This form of SR now appears simply as a special form of SR. Nevertheless, it bears important historical and conceptual significance, and we shall present in the next Section this form of (periodic) SR as our first detailed example of an SR phenomenon.

2 Periodic SR in Bistable Dynamic Systems

This form of SR is based on the evolution equation

$$\tau_a \dot{x}(t) = -\frac{dU(x)}{dx} + s(t) + \eta(t) \; . \tag{1}$$

Such an equation represents a dynamic system whose state $x(t)$ is forced by the input $s(t)+\eta(t)$, and whose free relaxation $\tau_a \dot{x} = -dU/dx$ is governed by a potential $U(x)$ which generally is a double-well potential. A form frequently chosen is the "quartic" potential

$$U(x) = -\frac{x^2}{2} + \frac{x^4}{4X_b^2} \tag{2}$$

with parameter $X_b > 0$, whose shape is depicted in Fig. 2.

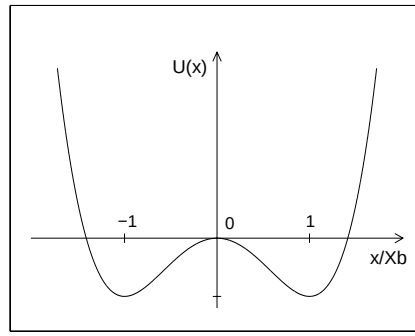

Fig. 2. Double-well bistable potential $U(x)$ of (2)

Because of its double-well potential $U(x)$, the dynamic system of (1) has two stable stationary states ($x = \pm X_b$ in the case of (2)) corresponding to the two minima of the potential ($U(x = \pm X_b) = -X_b^2/4$ in the case of (2)), and separated by a potential barrier (with height $X_b^2/4$ in the case of (2)).

A mechanical interpretation of this system allows a concrete description of the occurence of the SR phenomenon. In such an interpretation, (1) describes the motion in an overdamped regime (inertia $\ddot{x}$ is neglected relative to viscous friction $\dot{x}$), of a particle in a potential $U(x)$ subjected to the external force $s(t) + \eta(t)$. If a periodic input $s(t) = A\cos(2\pi t/T_s)$ is applied alone and with a too weak amplitude A, then the particle cannot jump over the potential barrier between the two wells; it remains confined in one of the wells around a potential minimum, with no transitions between wells. One can introduce here a binary output signal $y(t)$, with two states say $y(t) = \pm 1$, indicating which of the two wells the particle is in at time t, for instance

$$y(t) = \text{sign}[x(t)] \tag{3}$$

with the potential of (2). As no transitions take place between wells, $y(t)$ remains stuck in one of its two states. Then, if a small noise $\eta(t)$ is added, a cooperative effect between the signal $s(t)$ and the noise becomes possible, enabling occasionally the particule to jump over the potential barrier. This translates into transitions between wells which are correlated with the periodic input $s(t)$ as it plays a part in their production (in conjunction with the noise). When the noise level is raised, the probability of occurence of such coherent transitions first increases, thus reinforcing the correlation of the output $y(t)$ with the periodic input $s(t)$. For stronger noise levels, incoherent transitions induced by the noise alone will become more and more frequent, and will gradually destroy the correlation of the output with the periodic input. The noise thus has a nonmonotonic influence, first enhancing

the correlation of the output with the periodic input, up to an optimum, after which the correlation is gradually destroyed.

The output $y(t)$ is a random signal, because of the influence of the noise input $\eta(t)$, yet it bears correlation with the periodic input $s(t)$. To quantify the correlation of $y(t)$ with $s(t)$, the standard method starts with the calculation of the autocorrelation function of $y(t)$, and then through Fourier transform, to its power spectral density [55,73,26]. In the power spectral density of $y(t)$, the influence of the periodic input $s(t)$ shows up as spectral lines at integer multiples of the coherent frequency $1/T_s$. These lines emerge out of a broadband continuous noise background stemming from incoherent transitions due to $\eta(t)$. This typical constitution of the output power spectral density in periodic SR is depicted in Fig. 3A. The SR effect is identified by coherent spectral lines whose emergence out of the noise background gets more pronounced when the level of the input noise is raised. This is quantified [55,73,26] by a signal-to-noise ratio (SNR) at the output, defined as the power contained in the spectral line at $1/T_s$ divided by the power contained in the noise background in the region of $1/T_s$. When the level of the input noise is raised, the ouput SNR experiences a nonmonotonic evolution culminating at a maximum for an optimal nonzero noise level, whence the term resonance.

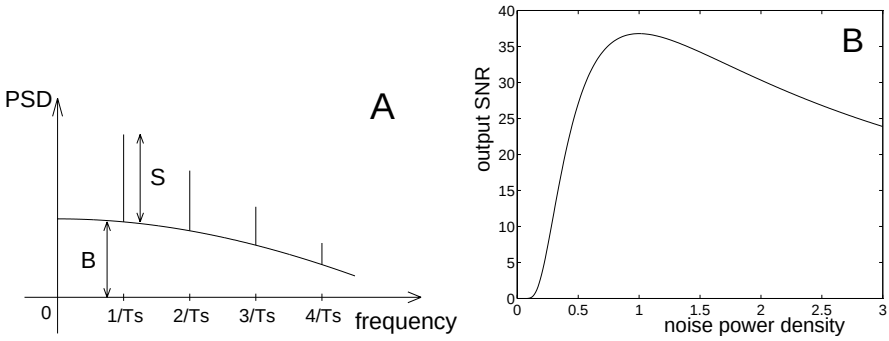

Fig. 3. (A) Power spectral density at the output serving for the definition of the signal-to-noise ratio $\mathcal{R}_{\rm out} = S/B$ offering a quantitative measure to periodic SR. (B) Output signal-to-noise ratio $\mathcal{R}_{\rm out}$ from (4) as a function of the input noise power density D, demonstrating SR in the bistable dynamic system of Eqs. (1)–(3)

For the case of the quartic potential of (2) with a Gaussian white noise $\eta(t)$ of autocorrelation function $\mathrm{E}[\eta(t)\eta(t+\tau)] = 2D\delta(\tau)$, in the regime where the coherent signal is both small and slow, the theory of [52] (based on a master equation for the output state probabilities governed by transition rates of a Kramers type) derives an explicit expression for the SNR under the form:

$$\mathcal{R}_{\rm out} = \frac{A^2 X_b^2/4}{(D/\tau_a)^2} \exp\left(-\frac{X_b^2/4}{D/\tau_a}\right) . \tag{4}$$

A typical evolution of the output SNR of (4) is depicted in Fig. 3B. The nonmonotonic evolution of the SNR in Fig. 3B with the noise power density D is the signature of SR, which is also verified by simulation of the system of Eqs. (1)–(3) [52]. This SR theory of [52] is applicable for any double-well bistable potential $U(x)$, but with a sinusoidal input $s(t)$ both slow and small and with a white Gaussian input noise. This treatment was for example used with a double-well potential differing from the quartic potential, and which describes certain regimes of operation of neurons in networks [4].

It is under this form, associated to a bistable dynamic system of the type of (1) excited by a noisy sinusoid, that the phenomenon of SR was originally introduced [2]. Under this form also, that SR has received the most sustained attention for a long period of time [26]. This form of SR has been gradually observed experimentally in various processes obeying the bistable dynamics of (1), and including electronic circuits [24], lasers [72], electron paramagnetic resonance [28], neurons [4,18], a magnetoelastic pendulum [66], chemical reactions [45,21], superconducting devices [37]. In these observations, the characterization scheme exposed above for SR, which proceeds through the computation of the output autocorrelation function then to the power spectral density to access the SNR in the frequency domain, can be numerically realized on measured signals. In addition, the theory of [52] can be used to derive an approximate expression of the SNR for a theoretical prediction of SR.

Other theories have been proposed to predict the behavior of the SNR in SR based on (1). Some approaches consider the Fokker-Planck equation [29] associated to the stochastic differential equation (1) [26,43,42,39]. In this context which is both nonlinear and nonstationary, these approaches lead to approximate expressions for the SNR.

Another method proposed for the theoretical analysis of SR is linear-response theory [22,20,23,57]. This theory is a perturbative method based on the linearization of the response of a nonlinear system for a small coherent input added to the noise. In principle, it can be applied to any nonlinear system, not necessarily bistable dynamic; an application is performed in [68] exhibiting periodic SR in a monostable nonlinear dynamic system. But linear-response theory is restricted to perturbative conditions defined by a small coherent input. However, SR can well take place outside these perturbative conditions [11], especially if one wants to obtain the important property of an output SNR larger than the input SNR in SR [7,9].

3 Periodic SR in Static Nonlinear Systems

SR in nonlinear dynamic systems as those based on (1) is usually difficult to theoretically analyze, and deriving an explicit expression for the SNR in given conditions usually requires to ressort to approximations (like those of a small or a slow coherent signal $s(t)$). There exists another class of nonlinear

dynamic systems where the theoretical analysis of SR can be performed exactly in many conditions [6,11]. This class consists of the nonlinear dynamic systems for which the nonlinear and the dynamic characters are separated. These systems are formed by the association of a static or memoryless nonlinearity followed by an arbitrary linear dynamic system, according to Fig. 4, and they are amenable to a general theory [11,7].

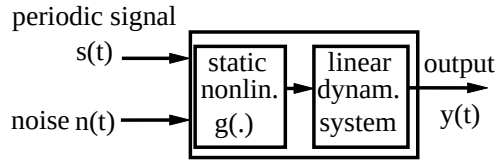

Fig. 4. A general class of nonlinear dynamic systems exhibiting SR, and describable by the theory of [11,7]

We shall now expose an example of periodic SR in a simple instance of the class of Fig. 4, with a complete theoretical description based on the approach of [11]. We shall deal with a nonlinear electronic circuit involving one of the most common nonlinearities found in electronic devices, i.e. the threshold nonlinearity of a PN junction diode [33].

We consider the circuit shown in the inset of Fig. 5A, where the input voltage consists of the sum $s(t) + \eta(t)$, with $s(t)$ a T_s-periodic signal and $\eta(t)$ a stationary white noise with the probability density function $f_\eta(u)$. The output voltage $y(t)$ results as a nonstationary, yet cyclostationary [61], random signal bearing correlation with the periodic input $s(t)$.

This correlation is quantified here by means of the output autocorrelation function [11]

$$R_{yy}(k\Delta t) = \frac{1}{N}\sum_{j=0}^{N-1} \mathrm{E}[y(j\Delta t)y(j\Delta t + k\Delta t)] \;, \tag{5}$$

where, for comparing theory and experiment, we use a discrete-time formulation in which the signals are sampled at a step $\Delta t \ll T_s = N\Delta t$.

The output power spectral density follows, from a discrete Fourier transform of R_{yy} over an integer number $2M$ of period T_s, as

$$P_{yy}(\ell\Delta\nu) = \sum_{k=-MN}^{MN-1} R_{yy}(k\Delta t)\exp\left(-i2\pi\frac{k\ell}{2MN}\right) \;, \tag{6}$$

with the frequency resolution $\Delta\nu = 1/(2MN\Delta t)$.

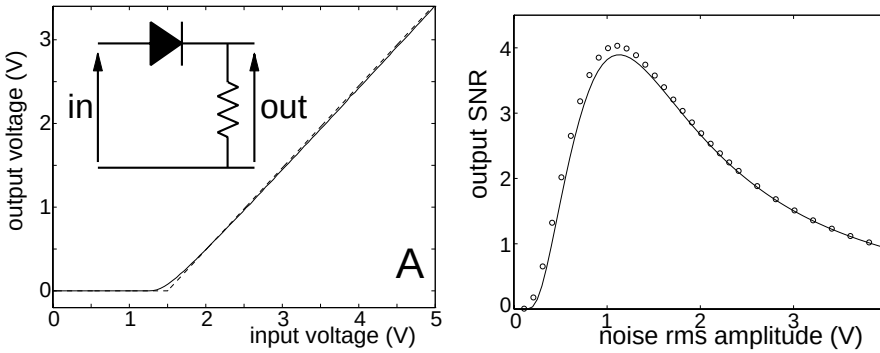

Fig. 5. Panel A: Diode circuit and its input–output static characteristics experimentally recorded (solid line) and its simple piecewise-linear theoretical model (dashed line). Panel B: Experimental (circles) and theoretical from (12) (solid line) ouput SNR $\mathcal{R}_{\rm out}(1/T_s)$ at the fundamental $1/T_s$, as a function of the rms amplitude σ_η of a zero-mean Gaussian input noise $\eta(t)$, for the diode circuit with a sinusoidal input signal $s(t) = A\cos(2\pi t/T_s)$ with $A = 1\,\mathrm{V}$

The theory of [11] leads in these conditions to an explicit expression for the output power spectral density as

$$P_{yy}(\ell\Delta\nu) = \overline{\mathrm{var}(y)} + \sum_{n=-\infty}^{+\infty} |\overline{Y}_n|^2\, \widehat{\delta}(\ell - 2Mn)\ , \tag{7}$$

with $\widehat{\delta}(0) = 2MN$ and $\widehat{\delta}(j) = 0$ for an integer $j \neq 0$. In (7),

$$\overline{Y}_n = \frac{1}{N}\sum_{j=0}^{N-1} \mathrm{E}[y(j\Delta t)]\exp\left(-i2\pi\frac{jn}{N}\right) \tag{8}$$

is the order n Fourier coefficient of the T_s-periodic output expectation $\mathrm{E}[y(j\Delta t)]$ computable as

$$\mathrm{E}[y(t)] = \int_{-\infty}^{+\infty} g(u) f_\eta[u - s(t)]du\ . \tag{9}$$

Also in (7),

$$\overline{\mathrm{var}(y)} = \frac{1}{N}\sum_{j=0}^{N-1} \mathrm{var}[y(j\Delta t)]\ , \tag{10}$$

with the T_s-periodic output variance $\mathrm{var}[y(t)] = \mathrm{E}[y^2(t)] - \mathrm{E}^2[y(t)]$ computable with

$$\mathrm{E}[y^2(t)] = \int_{-\infty}^{+\infty} g^2(u) f_\eta[u - s(t)]du\ . \tag{11}$$

In (9) and (11), the function $g(u)$ represents the input–output static characteristics of the nonlinear circuit realizing $y = g(s+\eta)$. For the circuit of the inset of Fig. 5A, the experimental characteristics has been measured and is also represented in Fig. 5A. We chose the diode (a red LED TLHR 5400 from TEMIC) so as to have an almost linear characteristics above the conduction threshold, well approximated by the simple model $g(u) = u - V_{\rm th}$ for $u > V_{\rm th}$, and $g(u) = 0$ otherwise, with $V_{\rm th} = 1.5\,{\rm V}$.

The output power spectral density P_{yy} of (7) has a shape which conforms to the general model of Fig. 3A. It is formed by spectral lines with amplitude $|\overline{Y}_n|^2$ at integer multiples of the coherent frequency n/T_s, emerging out of a continuous noise background whose magnitude is measured by $\overline{{\rm var}(y)}$ constant in the present case of a white noise. The standard ouput SNR given by the power in the line at frequency n/T_s divided by the power in the noise background in a small frequency band ΔB around n/T_s, results as

$$\mathcal{R}_{\rm out}\left(\frac{n}{T_s}\right) = \frac{|\overline{Y}_n|^2}{\overline{{\rm var}(y)}\Delta t \Delta B}\ . \tag{12}$$

In the case where $\eta(t)$ is a zero-mean Gaussian noise with rms amplitude σ_η, (9) leads to

$${\rm E}[y(t)] = \frac{\sigma_\eta}{\sqrt{2}}\left\{\frac{1}{\sqrt{\pi}}\exp[-z^2(t)] - z(t)\,{\rm erfc}[z(t)]\right\}\ , \tag{13}$$

and (11) to

$${\rm E}[y^2(t)] = \frac{\sigma_\eta^2}{2}\left\{[1+2z^2(t)]\,{\rm erfc}[z(t)] - \frac{2}{\sqrt{\pi}}z(t)\exp[-z^2(t)]\right\}\ , \tag{14}$$

with $z(t) = [V_{\rm th} - s(t)]/(\sigma_\eta\sqrt{2})$.

From (13) and (14), the output SNR of 12) has been computed. Also, this quantity has been experimentally evaluated on the circuit of Fig. 5A in the interesting regime of a subthreshold coherent input $s(t)$. Both results are shown in Fig. 5B (with $\Delta B = 1/T_s$) revealing very good agreement between theory and experiment, taking into account the simple piecewise-linear model adopted for the static characteristics. The nonmonotonic evolution of the output SNR with the input noise level, which culminates at a maximum value for an optimal nonzero noise level, demonstrates the SR effect, by which here a subthreshold coherent input $s(t)$ is aided by noise to overcome the threshold.

The present electronic circuit stands for one of the simplest conceivable stochastic resonators. In particular, it comes here with a complete theoretical analysis, which is not always feasible for more complex stochastic resonators. Other electronic circuits also shown to exhibit SR, like a Schmitt trigger [24] or a chaotic Chua circuit [1], have more complicated nonlinearities that hinder

an exact theoretical treatment of the effect, which was essentially exhibited through experiments or numerical simulations.

SR in our diode circuit relies here on the presence of a threshold and a subliminal signal whose transmission is only possible in the presence of the noise, with maximum efficacy for an optimum noise level explicitly predictable with the theory. Other situations exist where a subliminal signal or a threshold are not necessary for SR [11,3]. Also for SR, the periodic signal $s(t)$ can be replaced either by a high-frequency carrier modulated by a low-frequency message, or by a broadband aperiodic signal [1,14,8].

4 Aperiodic SR in a Nonlinear Information Channel

We shall now expose an example of aperiodic SR, i.e. SR with an aperiodic coherent signal $s(t)$ in the scheme of Fig. 1. At the same time, we shall introduce another measure of the performance receiving enhancement from the noise, under the form of an information-theoretic quantity. This will result in an information channel whose capacity is improvable via noise addition.

We consider the transmission of information over a memoryless channel. The input to the channel is a discrete random variable, that we write s to conform to the general scheme of Fig. 1, and which assumes values 1 or -1 respectively with probabilities p_1 and $p_{-1} = 1 - p_1$. Transmission over the channel involves two effects. First a noise η is added to the input s to produce $s+\eta$. Next, $s+\eta$ is compared to a fixed double threshold $\theta > 0$ to determine the discrete output y of the channel according to

$$\begin{aligned} s+\eta < -\theta &\Rightarrow y = -1, \\ -\theta \le s+\eta \le \theta &\Rightarrow y = 0, \\ s+\eta > \theta &\Rightarrow y = 1. \end{aligned} \tag{15}$$

The noise η is a random variable, continuous or discrete, with the cumulative distribution function $F_\eta(u) = \Pr\{\eta \le u\}$. The successive realizations of the noise η are independent and identically distributed, and the same for the random input s. The noise η and the input s are statistically independent.

The present channel can be considered as a binary channel with erasure [16] where the input binary information $s = \pm 1$ can be received as $y = \pm 1$ possibly with an error, or erased when $y = 0$.

For the input–output transfer probabilities of this channel we have, for example, the probability $p_{11} = \Pr\{y = 1 \,|\, s = 1\}$ which is also $\Pr\{s + \eta > \theta \,|\, s = 1\}$ amounting to $\Pr\{\eta > \theta - 1\} = 1 - F_\eta(\theta - 1)$. With similar rules we

arrive at

$$p_{11} = \Pr\{y = 1 \,|\, s = 1\} = 1 - F_\eta(\theta - 1) \,, \tag{16}$$

$$p_{-1,1} = \Pr\{y = -1 \,|\, s = 1\} = F_\eta(-\theta - 1) \,, \tag{17}$$

$$p_{1,-1} = \Pr\{y = 1 \,|\, s = -1\} = 1 - F_\eta(\theta + 1) \,, \tag{18}$$

$$p_{-1,-1} = \Pr\{y = -1 \,|\, s = -1\} = F_\eta(-\theta + 1) \,, \tag{19}$$

$$p_{01} = \Pr\{y = 0 \,|\, s = 1\} = [1 - F_\eta(-\theta - 1)]F_\eta(\theta - 1) \,, \tag{20}$$

$$p_{0,-1} = \Pr\{y = 0 \,|\, s = -1\} = [1 - F_\eta(-\theta + 1)]F_\eta(\theta + 1) \,. \tag{21}$$

Once the transfer probabilities are known, the input–output mutual information $I(s;y)$ can be evaluated from the entropies [16] as

$$I(s;y) = H(y) - H(y \,|\, s) \,. \tag{22}$$

With $h(u) = -u \log_2(u)$, the output entropy $H(y) = \sum_y h(\Pr\{y\})$ is here

$$\begin{aligned} H(y) = {} & h[p_1 p_{11} + (1 - p_1)p_{1,-1}] + h[p_1 p_{-1,1} + (1 - p_1)p_{-1,-1}] \\ & + h[p_1 p_{01} + (1 - p_1)p_{0,-1}] \,, \end{aligned} \tag{23}$$

and the input–output conditional entropy $H(y|s) = p_1 \sum_y h(\Pr\{y|s=1\}) + (1 - p_1) \sum_y h(\Pr\{y|s=-1\})$ is

$$\begin{aligned} H(y \,|\, s) = {} & p_1[h(p_{11}) + h(p_{01}) + h(p_{-1,1})] \\ & + (1 - p_1)[h(p_{1,-1}) + h(p_{0,-1}) + h(p_{-1,-1})] \,. \end{aligned} \tag{24}$$

Equations (22), (23) and (24) provide an explicit expression for the mutual information $I(s;y)$ as a function of the transfer probabilities (16)–(21) and the input probability p_1. In the case where the input noise η is symmetric, i.e. with an even probability density, the transmission process is symmetric, and the mutual information is invariant in the exchange of the values of p_1 and $p_{-1} = 1 - p_1$. As the mutual information of a discrete memoryless channel is always a concave function of the input probability distribution [16], we deduce that for η symmetric, the maximum of $I(s;y)$ which defines the information capacity C of the channel, is reached for $p_1 = 0.5 = p_{-1}$. In case of a symmetric noise, Eqs. (22), (23) and (24) with $p_1 = 0.5$, thus allow an explicit evaluation of the capacity C of the channel. We can now examine, on the capacity C, the influence of the noise η as conveyed by the function $F_\eta(u)$.

Figure 6A shows the capacity C as a function of the rms amplitude σ_η of the noise η chosen zero-mean Gaussian. When the threshold $\theta < 1$, an input $s = 1$ (respectively $s = -1$) is alone sufficent to trigger an output $y = 1$ (resp. $y = -1$). In the absence of the noise η the capacity is thus $C = 1$ bit. Addition of noise is then only felt as a nuisance, entailing a decay of C as σ_η increases above zero. In contrast, when $\theta > 1$, and input $s = 1$ (or $s = -1$) is alone insufficent to trigger an output $y = 1$ (or $y = -1$). In the absence of the noise η the capacity is thus $C = 0$ and the channel is unable to transmit

any information. It is the addition of noise which allows the transmission of information, through a cooperative effect where the noise and the input s collaborate to overcome the threshold. This translates into a nonzero capacity C of the channel in the presence of the noise, with a region where C increases as σ_η increases, up to an optimal noise level where C is maximized.

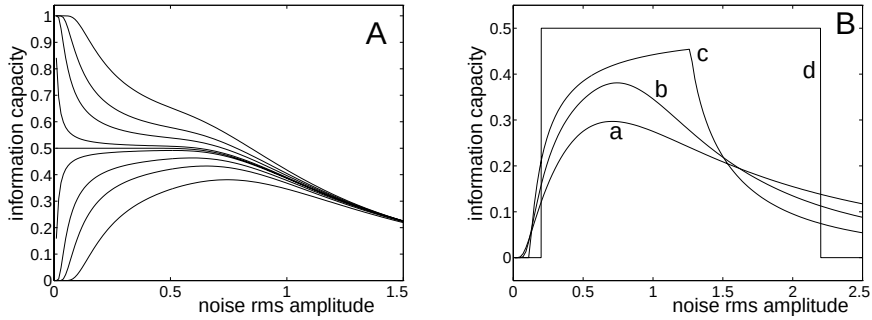

Fig. 6. Information capacity C (in bits) of the channel of (15) as a function of the rms amplitude σ_η of the noise η. Panel A: η is zero-mean Gaussian; the 9 curves are obtained for 9 values of the threshold θ, with successively from the uppest curve to the lowest: $\theta = 0.8$, 0.9, 0.95, 0.99, 1, 1.01, 1.05, 1.1, 1.2. Panel B: $\theta = 1.2$ and the symmetric zero-mean noise η being (a) two-sided exponential, (b) Gaussian, (c) uniform, (d) two-level discrete

Our model also predicts an influence of the noise distribution on the SR effect of noise-enhanced capacity. For illustration, we have tested four different symmetric noises sharing the same rms amplitude σ_η. Figure 6B shows that in each case the noise enhancement of the capacity is preserved, yet with an influence from the noise distribution.

The results of the present Section demonstrate that information-theoretic measures are capable of quantifying an effect of noise-enhanced transmission of an aperiodic random information-carrying signal. This is a form of aperiodic SR. Other examples of SR characterized by information-theoretic measures can be found in [46,58,5,36,8,32,17]. Various studies have used information-theoretic measures to establish that aperiodic SR can take place in information transmission by neurons [46,5,17].

5 Aperiodic SR in Image Transmission

Recently the phenomenon of SR has been applied to the noise-enhanced transmission of bidimensional spatial signals or images, instead of monodimensional temporal signals. Various studies have considered SR in image perception by the visual nervous system [65,62,74]. SR has also been reported in image transmission by nonlinear optical devices [71].

For illustration, we shall expose here a simple example of SR on images, which again has the advantage of allowing a complete theoretical analysis. Also, at this occasion we shall introduce a new class of quantitative measures for SR based on input–output cross-correlations [15,14].

Leaning again on the general scheme of Fig. 1, we consider this time that the coherent information-carrying signal s is a bidimensional image where the pixels are indexed by integer coordinates (ℓ, m) and have intensity $s(\ell, m)$. For a simple illustration, we take here a binary image with $s(\ell, m) \in \{0, 1\}$. A noise $\eta(\ell, m)$, statistically independent of $s(\ell, m)$, linearly corrupts each pixel of image $s(\ell, m)$. The noise values are independent from pixel to pixel, and are identically distributed with the cumulative distribution function $F_\eta(u) = \Pr\{\eta(\ell, m) \leq u\}$. A nonlinear detector, that we take as a simple hard limiter with threshold θ, receives the sum $s(\ell, m) + \eta(\ell, m)$ and produces the output image $y(\ell, m)$ according to:

$$\begin{aligned} \text{If} \quad s(\ell, m) + \eta(\ell, m) > \theta \quad &\text{then} \quad y(\ell, m) = 1 \\ &\text{else} \quad y(\ell, m) = 0 \ . \end{aligned} \tag{25}$$

When the intensity of the input image $s(\ell, m)$ is low relative to the threshold θ of the detector, i.e. when $\theta > 1$, then $s(\ell, m)$ (in the absence of noise) remains undetected as the output image $y(\ell, m)$ remains a dark image. Addition of the noise $\eta(\ell, m)$ will then allow a cooperation between the intensities of images $s(\ell, m)$ and $\eta(\ell, m)$ to overcome the detection threshold. The result of this cooperative effect can be visually appreciated on Fig. 7, where an optimal nonzero noise level maximizes the visual perception.

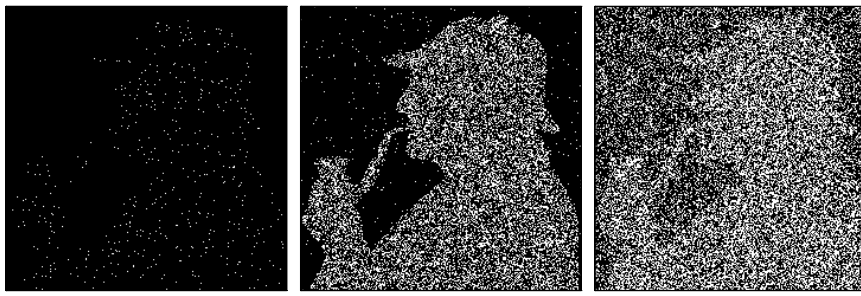

Fig. 7. The 256×256 binary image $y(\ell, m)$ at the output of the detector of (25) with threshold $\theta = 1.2$, when $\eta(\ell, m)$ is a zero-mean Gaussian noise with rms amplitude 0.1 (left), 0.5 (center) and 2 (right)

It is possible to quantitatively characterize the effect visually perceived in Fig. 7. An appropriate quantitative measure of the similarity between input image $s(\ell, m)$ and output image $y(\ell, m)$, is provided by the normalized cross-

covariance [71]

$$C_{sy} = \frac{\langle (s - \langle s \rangle)(y - \langle y \rangle) \rangle}{\sqrt{\langle (s - \langle s \rangle)^2 \rangle \langle (y - \langle y \rangle)^2 \rangle}} \ , \tag{26}$$

where $\langle . \rangle$ denotes an average over the images.

Another possibility is provided by the mutual information I_{sy} between the pixels of images $s(\ell, m)$ and $y(\ell, m)$,

$$I_{sy} = H(y) - H(y|s) \ , \tag{27}$$

with standard definitions [64] for the entropies $H(y)$ and $H(y|s)$.

Both measures C_{sy} and I_{sy} can be experimentally evaluated through pixels counting on images similar to those of Fig. 7. Also, for the simple transmission system of (25), both measures C_{sy} and I_{sy} can receive explicit theoretical expressions, as a function of $p_1 = \Pr\{s(\ell, m) = 1\}$ the probabilty of a pixel at 1 in the binary input image $s(\ell, m)$, and as a function of the properties of the noise conveyed by $F_\eta(u)$ [71].

For the scenario of Fig. 7, Fig. 8 represents these quantitative measures C_{sy} and I_{sy}, with both their experimental and theoretical evaluations which are in close agreement. Both measures identify a maximum efficacy in image transmission for an optimal nonzero noise level, another instance of SR.

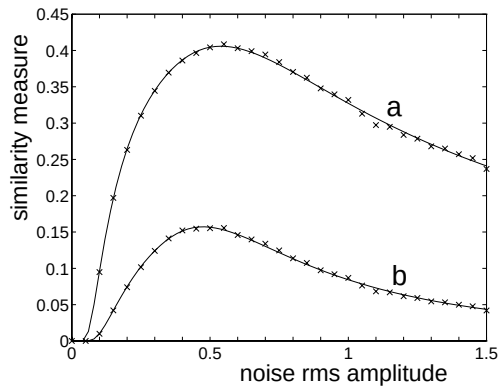

Fig. 8. Input–output similarity measures between input image $s(\ell, m)$ and output image $y(\ell, m)$, as a function of the rms amplitude of the noise $\eta(\ell, m)$ chosen zero-mean Gaussian. (a) is the cross-covariance C_{sy} of 26), (b) is the mutual information I_{sy} of (27). The crosses are experimental evaluations through pixels counting on images, the solid lines are the theoretical predictions ($p_1 = 0.6$)

6 Outlook

The above examples that we have treated are typical of the signals and nonlinear mechanisms by which SR can occur, and of how SR can be measured. Such processes have received numerous experimental embodiments materializing SR in many areas of physical sciences [54,26]. Among the systems specially studied that exhibit SR, one finds electronic circuits [24,1,31], optical devices [72,40,19,41,71], phenomena from material physics [28,44,56,35], neuronal systems [4,18,12,10], chemical reactions [45,21,38]. Most of the time SR is described at the macroscopic level by means of concepts from classical physics; but quantum forms of SR are also discussed [51,26,69].

SR, as presented here as an effect of noise-enhanced signal transmission, encompasses and generalizes other useful noise techniques that were previously known for specific purposes. This is the case with "dithering", which designates a technique of noise addition used in signal quantization [70,34] or image coding [63]. An application of dithering aims at limiting the distortion to an analog signal in its analog-to-digital conversion. The technique specifies to add to the analog signal prior to its quantization a white noise uniform over the quantization step, this to allow exact recovery of the mean of the analog signal from its digital representation. Dithering can be seen as a special form of SR, on a special nonlinear system (i.e. analog-to-digital converter) with a special performance measure (i.e. minimum input–output distortion of the mean) [25,11].

The "modern" approach to SR is thus, in the presence of given signals and nonlinear system, to select one specific performance measure endowed with a special significance in relation to the context or prospect involved, and then to examine whether conditions exist where the performance can be enhanced by addition of noise.

The examples of SR systems that we have treated here are nonlinear systems involving potential barriers or threshold nonlinearities. In such systems, SR takes place when the coherent signal alone is unsufficient to overcome the barrier or the threshold, and assistance to this aim is provided by the noise. Nevertheless, we emphasize that these nonlinear ingredients (barriers, thresholds) are not strictly necessary for the occurence of SR. The phenomenon of SR has been shown to operate in nonlinear systems without barriers [68] and without threshold [11,3]: [68] in a single-well monostable dynamic system, [11] in smooth monotonic or nonmonotonic static nonlinearities, [3] in thermal kinetics driven by the Boltzmann exponential factor. At a general level, SR can be interpreted as a displacement, by means of noise addition, of the operating zone of a nonlinear system into a region more favorable to the coherent signal. Standard techniques would use a constant bias for displacing the operating zone, meanwhile SR achieves it by noise addition.

Current developments to SR are concerned with the extension of the "signal-enhanced-by-noise" effect to broader conditions and domains. Among others one can cite the domain of array-enhanced SR where SR systems are

interconnected to interact in arrays [47,49], the domain of noise-enhanced propagation of nonlinear waves [50,75,48], extensions of SR to multidimensional signals or images [71,74], interplay between SR and chaotic dynamics [59,26]. SR studies are particularly active and relevant in the domain of neural information processing [30,67,13,60,17]. Neurons are examples of intrinsically nonlinear and noisy systems, in which SR operates, and that achieve very high performances for signal and information processing, through mechanisms that mainly remain to be understood.

In parallel to such analysis-oriented studies, another line of current development is the exploitation of SR for practical purposes and as a basis for competitive methods for signal and information processing. This direction of exploration is still in its infancy. It can have an impact if one is ready to depart, as neural systems suggest, from the realm of linear signal processing (as soon as the lower levels of processing) to come to the stage of nonlinear processing, more difficult to master but potentially rich of useful effects, as exemplified by SR.

References

1. V. S. Anishchenko, M. A. Safonova, and L. O. Chua. Stochastic resonance in Chua's circuit driven by amplitude or frequency modulated signals. *International Journal of Bifurcation and Chaos*, 4:441–446, 1994.
2. R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. Stochastic resonance in climatic changes. *Tellus*, 34:10–16, 1982.
3. S. M. Bezrukov and I. Vodyanoy. Stochastic resonance in non-dynamical systems without response thresholds. *Nature*, 385:319–321, 1997.
4. A. Bulsara, E. W. Jacobs, T. Zhou, F. Moss, and L. Kiss. Stochastic resonance in a single neuron model: Theory and analog simulation. *Journal of Theoretical Biology*, 152:531–555, 1991.
5. A. R. Bulsara and A. Zador. Threshold detection of wideband signals: A noise-controlled maximum in the mutual information. *Physical Review E*, 54:R2185–R2188, 1996.
6. F. Chapeau-Blondeau. Stochastic resonance in the Heaviside nonlinearity with white noise and arbitrary periodic signal. *Physical Review E*, 53:5469–5472, 1996.
7. F. Chapeau-Blondeau. Input–output gains for signal in noise in stochastic resonance. *Physics Letters A*, 232:41–48, 1997.
8. F. Chapeau-Blondeau. Noise-enhanced capacity via stochastic resonance in an asymmetric binary channel. *Physical Review E*, 55:2016–2019, 1997.
9. F. Chapeau-Blondeau. Periodic and aperiodic stochastic resonance with output signal-to-noise ratio exceeding that at the input. *International Journal of Bifurcation and Chaos*, 9:267–272, 1999.
10. F. Chapeau-Blondeau and X. Godivier. Stochastic resonance in nonlinear transmission of spike signals: An exact model and an application to the neuron. *International Journal of Bifurcation and Chaos*, 6:2069–2076, 1996.
11. F. Chapeau-Blondeau and X. Godivier. Theory of stochastic resonance in signal transmission by static nonlinear systems. *Physical Review E*, 55:1478–1495, 1997.

12. F. Chapeau-Blondeau, X. Godivier, and N. Chambet. Stochastic resonance in a neuron model that transmits spike trains. *Physical Review E*, 53:1273–1275, 1996.
13. D. R. Chialvo, A. Longtin, and J. Mullergerking. Stochastic resonance in models of neuronal ensembles. *Physical Review E*, 55:1798–1808, 1997.
14. J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff. Aperiodic stochastic resonance. *Physical Review E*, 54:5575–5584, 1996.
15. J. J. Collins, C. C. Chow, and T. T. Imhoff. Aperiodic stochastic resonance in excitable systems. *Physical Review E*, 52:R3321–R3324, 1995.
16. T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.
17. G. Deco and B. Schürmann. Stochastic resonance in the mutual information between input and output spike trains of noisy central neurons. *Physica D*, 117:276–282, 1998.
18. J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss. Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, 365:337–340, 1993.
19. M. I. Dykman, G. P. Golubev, I. K. Kaufman, D. G. Luchinsky, P. V. E. McClintock, and E. A. Zhukov. Noise-enhanced optical heterodyning in an all-optical bistable system. *Applied Physics Letters*, 67:308–310, 1995.
20. M. I. Dykman, H. Haken, G. Hu, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, C. Z. Ning, N. D. Stein, and N. G. Stocks. Linear response theory in stochastic resonance. *Physics Letters A*, 180:332–336, 1993.
21. M. I. Dykman, T. Horita, and J. Ross. Statistical distribution and stochastic resonance in a periodically driven chemical system. *Journal of Chemical Physics*, 103:966–972, 1995.
22. M. I. Dykman, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, N. D. Stein, and N. G. Stocks. Stochastic resonance: Linear response theory and giant nonlinearity. *Journal of Statistical Physics*, 70:463–478, 1993.
23. M. I. Dykman, D. G. Luchinsky, R. Mannella, P. V. E. McClintock, N. D. Stein, and N. G. Stocks. Stochastic resonance in perspective. *Nuovo Cimento*, 17D:661–683, 1995.
24. S. Fauve and F. Heslot. Stochastic resonance in a bistable system. *Physics Letters A*, 97:5–7, 1983.
25. L. Gammaitoni. Stochastic resonance and the dithering effect in threshold physical systems. *Physical Review E*, 52:4691–4698, 1995.
26. L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Reviews of Modern Physics*, 70:223–287, 1998.
27. L. Gammaitoni, F. Marchesoni, E. Menichella-Saetta, and S. Santucci. Stochastic resonance in bistable systems. *Physical Review Letters*, 62:349–352, 1989.
28. L. Gammaitoni, M. Martinelli, L. Pardi, and S. Santucci. Observation of stochastic resonance in bistable electron-paramagnetic-resonance systems. *Physical Review Letters*, 67:1799–1802, 1991.
29. C. W. Gardiner. *Handbook of Stochastic Methods.* Springer, Berlin, 1985.
30. X. Godivier and F. Chapeau-Blondeau. Noise-enhanced transmission of spike trains in the neuron. *Europhysics Letters*, 35:473–477, 1996.
31. X. Godivier and F. Chapeau-Blondeau. Noise-assisted signal transmission in a nonlinear electronic comparator: Experiment and theory. *Signal Processing*, 56:293–303, 1997.

32. X. Godivier and F. Chapeau-Blondeau. Stochastic resonance in the information capacity of a nonlinear dynamic system. *International Journal of Bifurcation and Chaos*, 8:581–590, 1998.
33. X. Godivier, J. Rojas-Varela, and F. Chapeau-Blondeau. Noise-assisted signal transmission via stochastic resonance in a diode nonlinearity. *Electronics Letters*, 33:1666–1668, 1997.
34. R. M. Gray and T. G. Stockham. Dithered quantizers. *IEEE Transactions on Information Theory*, IT-39:805–812, 1993.
35. A. N. Grigorenko and P. I. Nikitin. Magnetostochastic resonance as a new method for investigations of surface and thin film magnetism. *Applied Surface Science*, 92:466–470, 1996.
36. C. Heneghan, C. C. Chow, J. J. Collins, T. T. Imhoff, S. B. Lowen, and M. C. Teich. Information measures quantifying aperiodic stochastic resonance. *Physical Review E*, 54:R2228–R2231, 1996.
37. A. D. Hibbs, A. L. Singsaas, E. W. Jacobs, A. R. Bulsara, J. J. Bekkedahl, and F. Moss. Stochastic resonance in a superconducting loop with a Josephson junction. *Journal of Applied Physics*, 77:2582–2590, 1995.
38. W. Hohmann, J. Muller, and F. W. Schneider. Stochastic resonance in chemistry: The minimal bromate reaction. *Journal of Physical Chemistry*, 100:5388–5392, 1996.
39. G. Hu, G. Nicolis, and C. Nicolis. Periodically forced Fokker-Planck equation and stochastic resonance. *Physical Review A*, 42:2030–2041, 1990.
40. J. M. Iannelli, A. Yariv, T. R. Chen, and Y. H. Zhuang. Stochastic resonance in a semiconductor distributed feedback laser. *Applied Physics Letters*, 65:1983–1985, 1994.
41. B. M. Jost and B. E. A. Saleh. Signal-to-noise ratio improvement by stochastic resonance in a unidirectional photorefractive ring resonator. *Optics Letters*, 21:287–289, 1996.
42. P. Jung and P. Hänggi. Stochastic nonlinear dynamics modulated by external periodic forces. *Europhysics Letters*, 8:505–510, 1989.
43. P. Jung and P. Hänggi. Amplification of small signal via stochastic resonance. *Physical Review A*, 44:8032–8042, 1991.
44. L. B. Kiss, Z. Gingl, Z. Márton, J. Kertész, F. Moss, G. Schmera, and A. Bulsara. $1/f$ noise in systems showing stochastic resonance. *Journal of Statistical Physics*, 70:451–462, 1993.
45. D. S. Leonard and L. E. Reichl. Stochastic resonance in a chemical reaction. *Physical Review E*, 49:1734–1739, 1994.
46. J. E. Levin and J. P. Miller. Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature*, 380:165–168, 1996.
47. J. Lindner, B. Meadows, W. Ditto, M. Inchiosa, and A. Bulsara. Array enhanced stochastic resonance and spatiotemporal synchronization. *Physical Review Letters*, 75:3–6, 1995.
48. J. F. Lindner, S. Chandramouli, A. R. Bulsara, M. Löcher, and W. L. Ditto. Noise enhanced propagation. *Physical Review Letters*, 81:5048–5051, 1998.
49. J. F. Lindner, B. K. Meadows, W. L. Ditto, M. E. Inchiosa, and A. R. Bulsara. Scaling laws for spatiotemporal synchronization and array enhanced stochastic resonance. *Physical Review E*, 53:2081–2086, 1996.
50. M. Löcher, D. Cigna, and E. R. Hunt. Noise sustained propagation of a signal in coupled bistable electronic elements. *Physical Review Letters*, 80:5212–5215, 1998.

51. R. Löfstedt and S. N. Coppersmith. Quantum stochastic resonance. *Physical Review Letters*, 72:1947–1950, 1994.
52. B. McNamara and K. Wiesenfeld. Theory of stochastic resonance. *Physical Review A*, 39:4854–4869, 1989.
53. S. Mitaim and B. Kosko. Adaptive stochastic resonance. *Proceedings of the IEEE*, 86:2152–2183, 1998.
54. F. Moss, A. Bulsara, and M. F. Shlesinger, eds. Proceedings NATO Advanced Research Workshop on Stochastic Resonance in Physics and Biology. *Journal of Statistical Physics*, 70:1–512, 1993.
55. F. Moss, D. Pierson, and D. O'Gorman. Stochastic resonance: Tutorial and update. *International Journal of Bifurcation and Chaos*, 4:1383–1398, 1994.
56. Z. Neda. Stochastic resonance in 3D Ising ferromagnets. *Physics Letters A*, 210:125–128, 1996.
57. A. Neiman, L. Schimansky-Geier, and F. Moss. Linear response theory applied to stochastic resonance in models of ensembles of oscillators. *Physical Review E*, 56:R9–R12, 1997.
58. A. Neiman, B. Shulgin, V. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund. Dynamical entropies applied to stochastic resonance. *Physical Review Letters*, 76:4299–4302, 1996.
59. G. Nicolis, C. Nicolis, and D. McKernan. Stochastic resonance in chaotic dynamics. *Journal of Statistical Physics*, 70:125–139, 1993.
60. V. V. Osipov and E. V. Ponizovskaya. The nature of bursting noises, stochastic resonance and deterministic chaos in excitable neurons. *Physics Letters A*, 238:369–374, 1998.
61. A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.
62. M. Riani and E. Simonotto. Stochastic resonance in the perceptual interpretation of ambiguous figures: A neural network model. *Physical Review Letters*, 72:3120–3123, 1994.
63. L. G. Roberts. Picture coding using pseudo-random noise. *IRE Transactions on Information Theory*, IT-8:145–154, 1962.
64. J. C. Russ. *The Image Processing Handbook*. CRC Press, Boca Raton, 1995.
65. E. Simonotto, M. Riani, C. Seife, M. Roberts, J. Twitty, and F. Moss. Visual perception of stochastic resonance. *Physical Review Letters*, 78:1186–1189, 1997.
66. M. Spano, M. Wun-Fogle, and W. L. Ditto. Experimental observation of stochastic resonance in a magnetoelastic ribbon. *Physical Review A*, 46:5253–5256, 1992.
67. M. Stemmler. A single spike suffices: The simplest form of stochastic resonance in model neurons. *Network: Computation in Neural Systems*, 7:687–716, 1996.
68. N. G. Stocks, N. D. Stein, and P. V. E. McClintock. Stochastic resonance in monostable systems. *Journal of Physics A*, 26:L385–L390, 1993.
69. J. J. L. Ting. Stochastic resonance for quantum channels. *Physical Review E*, 59:2801–2803, 1999.
70. J. Vanderkooy and S. P. Lipshitz. Resolution below the least significant bit in digital systems with dither. *Journal of the Audio Engineering Society*, 32:106–113 (correction p. 889), 1984.

71. F. Vaudelle, J. Gazengel, G. Rivoire, X. Godivier, and F. Chapeau-Blondeau. Stochastic resonance and noise-enhanced transmission of spatial signals in optics: The case of scattering. *Journal of the Optical Society of America B*, 15:2674–2680, 1998.
72. G. Vemuri and R. Roy. Stochastic resonance in a bistable ring laser. *Physical Review A*, 39:4668–4674, 1989.
73. K. Wiesenfeld and F. Moss. Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDs. *Nature*, 373:33–36, 1995.
74. T. Yang. Adaptively optimizing stochastic resonance in visual system. *Physics Letters A*, 245:79–86, 1998.
75. Y. Zhang, G. Hu, and L. Gammaitoni. Signal transmission in one-way coupled bistable systems: Noise effect. *Physical Review E*, 58:2952–2956, 1998.

Time is Money

Marcel Ausloos[1⋆], Nicolas Vandewalle[1], and Kristinka Ivanova[2,3]

[1] SUPRAS and GRASP, Institut de Physique B5, Université de Liège, B-4000 Liège, Belgium
[2] Department of Meteorology, Pennsylvania State University, University Park, PA 16802, USA
[3] *permanent address* : Institute of Electronics, Bulgarian Academy of Sciences, Boul. Tzarigradsko chaussée 72, Sofia 1784, Bulgaria

Abstract. Specialized topics on financial data analysis from a numerical and physical point of view are discussed when pertaining to the analysis of coherent and random sequences in financial fluctuations within (i) the extended detrended fluctuation analysis method, (ii) multi-affine analysis technique, (iii) mobile average intersection rules and distributions, (iv) sandpile avalanches models for crash prediction, (v) the (m, k)-Zipf method and (vi) the i-variability diagram technique for sorting out short range correlations. The most baffling result that needs further thought from mathematicians and physicists is recalled: the crossing of two mobile averages is an original method for measuring the "signal" roughness exponent, but why it is so is not understood up to now.

1 Introduction

It is fortunate to recall from the start that Louis Bachelier (1870-1946) was a mathematician at the University of Franche-Comté in Besançon. He was the first to develop a theory of Brownian motion, – in his Ph. D. thesis [1], in fact for the pricing of options in speculative markets. Later on he wrote down what is known as the Chapman–Kolmogorov equation. Alas, he motivated his approach on what is known nowadays as the *efficient market theory*, basically a Gaussian distribution for the price changes. This is known to be incorrect, at least for economic indices [2,3].

Recently the statistical physics community has been reattracted into investigating economic and financial problems. Two modern reasons can certainly be given: (i) economic systems, like stock markets produce quite complex signals due to a high number of parameters involved, and (ii) models developed so far in actual econometry do seem irrelevant for mimicking available signals, – at least on the level expected by usual physical models for natural phenomena. A list of recent progress is too long to be cited or discusssed here. Several books are already of interest. One aim should be first to review a few technical details in a global context. Even for a general audience, with mathematical orientation, it is in fact hopefully possible to

⋆ ausloos@gw.unipc.ulg.ac.be

M. Planat (Ed.): LNP 550, pp. 156–171, 2000.
© Springer-Verlag Berlin Heidelberg 2000

give nonrigourous informations on how physicists pretend that an increase in revenue can be obtained if general rules are found from non linear dynamic-like analysis of financial time series. Some general views have been already presented in "Money Games Physicists Play" [4]. More specialized topics are discussed here as were already sketched in ref. [5].

There are six methods or so that we have been discussing and using in the Liège GRASP[1], when performing investigations in the context of sorting deterministic features from apparently stochastic noise contained in economic and financial data. The investigations pertain to considerations on different time correlation ranges. First it has been observed a long time ago that stock market fluctuations were not Brownian motion-like, but some long range correlation existed [3]. We have tested that idea on foreign exchange currency (FXC) rates [6]. Using the detrended fluctuation analysis method (DFA), it was shown that profit making in the FXC market can be made by bankers. This leads to the introduction of a turbulence-like picture in order to discuss the spareness and roughness of FXC rates. It can be shown that not all FXC rates belong to the same so-called universality class, but nobody knows at this time why, nor what universality classes exist.

Next, some medium range correlation can be discussed. First, a technique due to stock analysts, known as the mobile (or moving) average technique which allows for predicting gold or death crosses, whence suggesting buying or selling conditions will be discussed. It can be shown to be a rather delicate (a euphemism !) way of predicting what to do on a market. This will lead to a very interesting, and apparently unsolved problem for mathematicians and physicists. Moreover, the behavior of major stock market average indices will be recalled, and it will be observed that so-called crashes have well defined precursors. The crash of October 1987 could be seen as a phase transition [7]. The amplitude and the universality class can be discussed as well, thereby indicating how the financial crash of October 1997 could have been (and was) predicted. This will lead to indicate a microscopic model for such a set of crashes, model based on sandpile avalanches on fractal lattices. This will lead to emphasize a very interesting problem for the dynamics of numbers.

Moreover, a claim will be substantiated that the (m, k)-Zipf analysis and low order variability diagrams can be used for demonstrating short range correlation evidence in financial data.

2 Detrended Fluctuation Analysis Techniques

The Detrended Fluctuation Analysis technique consists in dividing a time series or random one-variable sequence $y(n)$ of length N into N/T nonoverlapping boxes, each containing T points [8]. Then, the local trend in each box

[1] GRASP = Group for Research in Applied Statistical Physics
http://www.supras.phys.ulg.ac.be/statphys/statphys.html

is *a priori* defined. A linear trend $z(n)$ like

$$z(n) = an + b, \tag{1}$$

or a cubic trend like

$$z(n) = cn^3 + dn^2 + en + f, \tag{2}$$

can be assumed [6,9]. In a box, the linear trend might be way-off from the overall intuitive trend, henceforth shorter scale fluctuations might be missed if the box size becomes quite large with respect to the apparent short time fluctuation scale of the signal. Thus the interest of using non linear trends. The parameters a to f are usually estimated through a linear least-square fit of the data points in that box. The process is repeated for all boxes. The detrended fluctuation function $F(T)$ is then calculated following

$$F^2(T) = \frac{1}{T} \sum_{n=kT+1}^{(k+1)T} |y(n) - z(n)|^2, \qquad k = 0, 1, 2, \cdots, (\frac{N}{T} - 1). \tag{3}$$

Usually only one type of trend is taken for the whole analysis, but mixed situations could be envisaged. Averaging $F(T)$ over all N/T box sizes centered on time T gives the fluctuations $\langle F(T) \rangle$ as a function of T. If the $y(n)$ data are random uncorrelated variables or short range correlated variables, the behavior is expected to be a power law

$$\langle F^2(T) \rangle^{1/2} \sim T^{\alpha} \tag{4}$$

with an exponent α =1/2 [8]. An exponent $\alpha \neq 1/2$ in a certain range of T values implies the existence of long-range correlations in that time interval as e.g. in the fractional Brownian motion [10]. Correlations and anticorrelations correspond to $\alpha > 1/2$ and $\alpha < 1/2$ respectively.[2]

If a signal has a fractal dimension D, its power spectrum is supposed to behave like

$$S(f) \sim f^{-\beta} \tag{5}$$

where

$$D = E + \frac{(3 - \beta)}{2} \tag{6}$$

[2] Notice that the procedure to estimate α in [11] includes an *a priori* integration of the tested signal, and these authors measure in fact an $\alpha' = \alpha + 1$.

or

$$\beta = 2H + 1, \tag{7}$$

in terms of the Hurst exponent H such that $D = E + 1 - H$ [10,12-14]; e.g. $\beta = 2$ for Brownian motion. Therefore, since $\alpha = H$

$$\beta = 2\alpha + 1. \tag{8}$$

In so doing one defines pink (or black) noise depending whether H is less (or greater) than 1/2. Black noise is related to long-memory effects, and pink noise to anti-persistence. processes are dominant over the external influences and perturbations [10].

Such power laws are the signature of a propagation of *information* across a hierachical system over very long times. Two cases are shown in Fig. 1. For time scales above 2 years, a crossover is however observed on studied financial data. This crossover suggests that correlated sequences have a characteristic duration of ca. 2 years along the whole financial evolution at least for the 16 years cases studied in ref. [6]. In order to probe the existence of *correlated and decorrelated sequences*, a so-called observation box of "length" 2 year was constructed and placed at the beginning of the data. The exponent α for the data contained in that box was calculated at each step. The box was then moved along the historical time axis by 20 points (4 weeks) toward the right along the financial sequence. Iterating this procedure for the sequence, a "local measurement" of the degree of "long-range correlations" is obtained, i.e. a local measure of the Hurst or α exponent. The results indicate that the α exponent value varies with the date. This is similar to what is also observed along DNA sequences where the α exponent drops below 1/2 in so-called non-coding regions.

The opposite has been observed for the breaking apart and disappearance of stratus clouds (over Oklahoma) [15]. The exponent α jumps from much below 1/2 to about 1/2 and drops back to a low value when the clouds scattered all over the area. By analogy with DNA and cloud behaviors, our findings suggest that financial markets loose the controlled structure (*following some loss of "information"*) at such a time. It should be noticed that in ref. [6] both sequences observed around 1983 and 1987 were not immediately seen from the rough data nor the value of α, and were missed by R/S and Fourier analysis.

Therefore, two of the main advantages of the DFA over other techniques like Fourier transform, or R/S methods [3] are that (i) local and large scale trends are avoided, and (ii) local correlations can be easily probed [6,9]. In economic data like stock exchange and currency fluctuations, long or short scale trends are *a posteriori* obvious and are of common evidence. The DFA method allows one to avoid such trend effects which can be considered as the envelope of the signal and could mask interesting details. Thus, we expect

that DFA allows a better understanding of apparently complex financial signals.

In so doing, correlations can be sorted out and a strategy for profit making can be developed in terms of persistence and antipersistence signals [6].

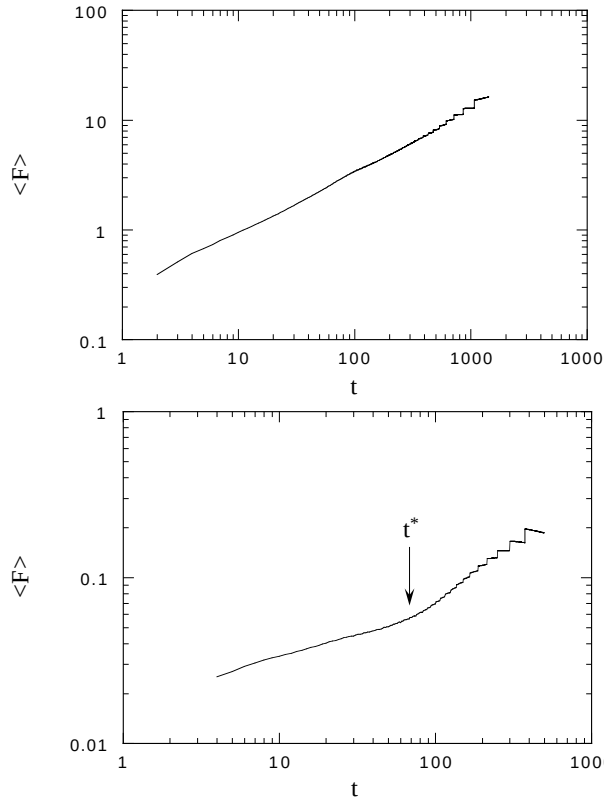

Fig. 1. Linearly Detrended Fluctuation Analysis function for two typical foreign currency exchange rates, i.e. JPY/USD and NLG/BEF between Jan. 01, 1980 and Dec. 31, 1995. The Brownian motion behavior corrresponds to a slope 1/2 on this log-log plot as indicated. The notation $\langle F \rangle$ is used for $\langle F^2(T) \rangle^{1/2}$ for conciseness in labeling the y-axis.

3 Multiaffine Analysis Techniques

A locally varying value of α suggests a multifractal process. A multi-affine analysis of several currency exchange rates has been performed in ref.[16–18], and also for Gold price, and Dow Jones Industrial Average (DJIA) in ref.[17]. In order to do so the roughness (Hurst) exponent H_1 and the intermittency

exponent C_1 are calculated for the correlation function $c(\tau)$ supposed to behave like

$$c_1(\tau) = \langle |y(t) - y(t')| \rangle_\tau \sim \tau^{H_1}. \tag{9}$$

The technique consists in calculating the so-called 'qth order height-height correlation function" [19] $c_q(\tau)$ of the time-dependent signal $y(t)$

$$c_q(\tau) = \langle |y(t) - y(t')|^q \rangle_\tau \tag{10}$$

where only non-zero terms are considered in the average $\langle . \rangle_\tau$ taken over all couples (t, t') such that $\tau = |t - t'|$. The roughness exponent H_1 describes the excursion of the signal. For the random walk (Brownian motion), $H_1 = 1/2 = H$. In the case of white noise $H_1 = 0$ [10]. Notice also that $H_1 \sim H_2 = H$.

The generalized Hurst exponent H_q is defined through the relation

$$c_q(\tau) \sim \tau^{qH_q}. \tag{11}$$

The C_1 exponent [19–21] is a measure of the intermittency in the signal $y(t)$

$$C_1 = - \left. \frac{dH_q}{dq} \right|_{q=1}. \tag{12}$$

which can be numerically estimated by measuring H_q around $q = 1$.

It appears that in a (H_1, C_1) diagram (Fig. 2) the currency exchange rates are dispersed over a wide region around the Brownian motion case ($H_1 = 0.5, C_1 = 0$) and have a significantly non-zero thus intermittent component, i.e. ($C_1 \neq 0$) – the value of which might depend on the nature of the trading market, thereby indicating that economic policy seems to be probed through the analysis and its role should be taken into account in further microscopic modeling [17].

4 Moving Average Techniques

A stock market index has often been considered as cyclic, but so-called unpredictable events, like crashes are fascinating. It should be noted that they take place at the end of a period of euphory, when some anxiety builds in. Surely it is not obvious from the general trend nor from the apparently stochastic noise when a crash is forthcoming. Can we find some deterministic content beside the official trend from a basic noise characteristic, e.g. the fractal dimension has beentaken as a fundamental question.

Roughness or Hurst exponents are commonly measured in surface science [22] and also in time series analysis [23]. From a usual technique by analysts, known as the mobile (or moving) average technique, an interesting way can

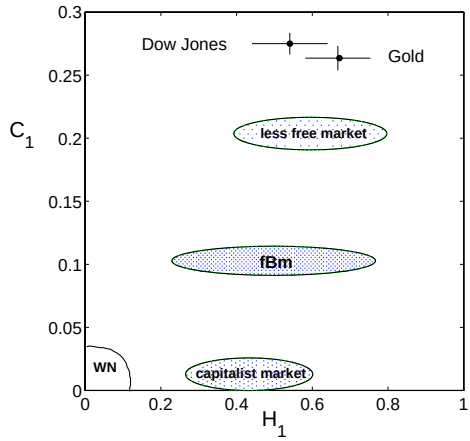

Fig. 2. Roughness(H_1), intermittency (C_1) parameter phase diagram of a few typical financial data and mathematical, i.e.fractional Brownian motion (fBm) and white noise (WN) signals.

be proposed for determining H and how to apply it right away to many cases with persistent or antipersistent correlations.

Consider a time series $y(n)$ given at discrete times n. At time n, the mobile average $\bar{y}$ is defined as

$$\bar{y}(n) = \frac{1}{N} \sum_{i=0}^{N-1} y(n-i), \tag{13}$$

i.e. the average of y for the last N data points. One can easily show that if y increases (resp. decreases) with time, $\bar{y} < y$ (resp. $\bar{y} > y$). Thus, the mobile average captures the trend of the signal over a time interval N. Such a procedure can be used in fact on any time series like in atmospheric or meteorological data, DNA, electronic noise, fracture, internet, traffic, and fractional Brownian motions.

Let two different mobile averages $\bar{y_1}$ and $\bar{y_2}$ be calculated respectively over e.g. T_1 and T_2 intervals such that $T_2 > T_1$. Since $T_1 \neq T_2$, the crossings of $\bar{y_1}$ and $\bar{y_2}$ coincide with drastic changes of the trend of $y(n)$. If $y(n)$ increases for a long period before decreasing rapidly, $\bar{y_1}$ will cross $\bar{y_2}$ from above. This event is called a "death cross" in empirical finance [24,25]. On the contrary, if $\bar{y_1}$ crosses $\bar{y_2}$ from below, the crossing point coincide with an upsurge of the signal $y(n)$. This event is called a "gold cross". Financial analysts usually try to '"extrapolate" the evolution of y_1 and y_2 expecting "gold" or "death" crosses. Most computers on trading places are equiped for performing this kind of analysis and forecasting [25]. Even though mobile averages seem to be "arbitrary" measures, they present some very practical interest for physicists and raise new questions. It is paradoxical to have such a type of nalysis

performed while on the other hand the "efficient market hypothesis" is the basis of most econometry theories.

It is well known [23] that the set of crossing points between a signal $y(n)$ and the $y = 0$-level is a Cantor set with a fractal dimension $1 - H$. The related physics pertain to so-called studies about first return time problems [26]. However, we have checked that the density ρ of crossing points between $\bar{y_1}$ and $\bar{y_2}$ curves is homogeneous along a signal and is thus *not* a Cantor set.

In so doing, the fractal dimension of the set of crossing points is one, i.e. the points are homogeneously distributed in time along $\bar{y_1}$ and $\bar{y_2}$. Due to the homogeneous distribution of crossing points, the forecasting of 'gold" and "death" crosses even for self-affine signals $y(n)$ seems unfounded.

However, it is of interest to observe how ρ behaves with respect to the choice in the relative difference $T_1 - T_2$. More precisely, consider the relative difference $0 < \Delta T < 1$ defined as $\Delta T = (T_2 - T_1)/T_2$. It has been found [27] that the density of crossing points $\rho(\Delta T)$ curve is fully symmetric, has a minimum and diverges for $\Delta T = 0$ and for $\Delta T = 1$, *with an exponent which is the Hurst exponent*. This remarkable and puzzling result does not seem to have been mentioned previously to ref. [27] due to the fact that some theoretical framework for the mobile average method is missing. The behavior of ρ is analogous to the age distribution of domains after coarsening in spin-like models [28] and to the density of electronic states on a fractal lattice in a tight binding approximation. This method of mobile averages can in fact serve to measure the Hurst exponent in a very fast and continuous way.

5 Sandpile Model for Rupture and Crashes

Another investigation of the relationship between the trend and local structure of a signal, like that of stock market measures like the DJIA and the Standard & Poor 500 (S&P500) has led us into examining regions where huge variations were taking place. These are usually associated to rupture phenomena'and 'crashes".

It has been proposed [29] that an economic index $y(t)$ increases as a complex power law, i.e.

$$y(t) = A + B\left(\frac{t_c - t}{t_c}\right)^{-m}\left[1 + C\sin\left(\omega \ln\left(\frac{t_c - t}{t_c}\right) + \phi\right)\right] \quad \text{for } t < t_c \tag{14}$$

where t_c is the crash-time or rupture point, A, B, m, C, ω and ϕ are free parameters. The law for $y(t)$ diverges (converges) at $t = t_c$ if the exponent m is positive (negative) while the period of the oscillations converges to the rupture point at $t = t_c$. The real part of the law is similar to that of critical points at so-called second order phase transitions [30] but generalizes the scaleless situation to cases in which a discrete scale invariance [31] is

presupposed when a complex exponent $m + i\omega$ exists. This relationship was already proposed in order to fit experimental measurements of sound wave rate emissions prior to the rupture of heterogeneous composite stressed up to failure [32]. Such log-periodic corrections have been recently reported in biased diffusion on random lattices [33], and in our sandpile studies is found when the underlying base is quasi-fractal [34]. Thus, an avalanche sand pile model can be imagined for financial indices [35]

A logarithmic divergence, corresponding to the $m = 0$ limit, can be also proposed, [36] i.e. the divergence of the index y for t close to t_c should be such that

$$y(t) = A + B \ln\left(\frac{t_c - t}{t_c}\right)\left[1 + C \sin\left(\omega \ln\left(\frac{t_c - t}{t_c}\right) + \phi\right)\right] \quad \text{for } t < t_c \tag{15}$$

In August 1997, a series of investigations was performed in order to test the existence of crash precursors. Daily data of the DJIA and the S&P500 was used. A strong indication of a crash event or a rupture point in between the end of october 1997 to mid-november 1997 was numerically discovered [37], and later predicted to occur during the week of Oct. 27, 97, and it was observed to occur on Oct. 27, 97 [38]. This resulted from an analysis of the similarities between two long periods: 1980-87 and 1990-97. The number of open days per year on Wall Street is about 261 days, - the exact value depending on the number of holidays falling on week ends. For the first period, the analysis was performed on data ending two months before the so-called Black Monday, i.e. October 19, 1987. For the second period, the data was considered till August 20th, 1997. In fact, we have separated the search of the crash day into two problems, that of the divergence itself and that of the oscillation convergences on the other hand, i.e. (i) t_c^{div} for the power (or logarithmic) divergence and (ii) t_c^{osc} for the oscillation convergence.

Sometimes it might be natural to be contempting, and/or displeased by, the eye balling technique we are supposedly using [39–42]. We should totally disagree concerning this gross misunderstanding of our technique. Our statistical analysis takes into account the approximate location of the maxima, and in a recent paper it has been precisely shown one good way of taking the maximum location into account. It is true that in ref.[37] the arrows pointing at maxima and minima look rather thick, but this is for a display purpose. In fact the statistical data analysis takes into account the number of data points in the best possible interval, as it is standard in critical point (exponent) analysis [43,44]. In so doing the origin of the time interval is obtained indeed. This time origin the definition of the phase) should lead to some interesting questions in fact. It might be of interest to recall that the closing value of the DJIA was used. This is not necessarily the intraday maximum value in fact, nor the intraday average value. One might wonder if the former or the latter would give a better estimate of the upper bound of

the predicted crash day. One might search whether rather than the closing value or the maximum the average DJIA, or average of range, or something else over a one day interval should be better used for better predictability, etc. It is known that there are larger fluctuations at the begining and at the end of a day. These are left for further investigations. The stability of the fit parameters can also be checked on these closing values with respect to random noise and through a Monte-Carlo data in order to take into account some sort of uncertainty in a bivariate data analysis (with error bars on the x and y axes). However due to the apparent precision of the technique at this time no robustness test has been performed as of now.

It should be pointed out that we do not expect any real divergence in the stock market indices; this is total non sense of course. However a divergence is predicted by us to occur at some upper bound of t_c. This is exactly the same as in phase transitions, where there is never any infinite divergence at the critical temperature. The divergence of the correlation length, specific heat, etc. is a virtual (mathematical) image of physical reality. There is no infinity (nor zero in fact) in physics due to finite size effects, inhomogeneities, noise, etc. Therefore to argue on the true existence of zeroes and infinity [33,40–42] is rather meaningless. We consider that to give an upper bound is certainly an as good predictive technique in data analysis and for modeling, as good as to give a determinis tic finite value at t_c.

Moreover a true drop certainly exists at a crash and is the signature of the crash, and the formula of ref. [29] would seem therefore appropriate. According to ref. [29,40], the drop goes to a finite value. Notice that there is some sense indeed to examine the size of jumps at crashes though. Such an attempt has been made in ref. [7].

index - (period)	$t_c^{div}(m=0)$	$t_c^{div}(m\neq 0)$	t_c^{osc}
DJIA (80-87)	87.85 ± 0.02	88.46 ± 0.04	87.91 ± 0.10
DJIA (90-97)	97.92 ± 0.02	98.68 ± 0.04	97.89 ± 0.06
S&P500 (80-87)	87.89 ± 0.03	88.78 ± 0.05	87.88 ± 0.07
S&P500 (90-97)	97.90 ± 0.02	98.67 ± 0.04	97.85 ± 0.08

Table 1: Fundamental parameters found for the DJIA and S&P500 indices during 1980-87 and 1990-97 periods. Time is expressed in years. The notations for t_c are such that e.g. 97.90 means the calendar date corresponding to the 90-th day as if there are 100-days in 1997. Two values of t_c^{div} correspond to a fit using a logarithmic divergence ($m = 0$) and a fit using a power law divergence ($m \neq 0$) respectively. The true date of the October 1987 crash in the above units gives $t_c = 87.79$ and for the October 1997 crash is $t_c = 97.81$, i.e. *quasi* the predicted dates.

6 (m, k)-Zipf Techniques

For testing and emphasizing short range correlations, the (m, k)-Zipf and the i-Variability Diagram (i-VD) techniques have been used. The Zipf analysis consists

in counting the number of words of a certain type appearing in a text, calculating the frequency of occurence f_o of each word in a given text, and sorting out the words according to their frequency, i.e. a rank R is assigned to each word, with $R = 1$ for the most frequent one, and rank R_M for the word appearing the less. Moreover call f_M the frequency (occurrence) of the most often observed word.

For natural languages, one observes a power law

$$f_o \sim R^{-\zeta} \tag{16}$$

with an exponent ζ close to one for any language. This has been applied to various complex signals or "texts" [45–47], economy (size of sales and firms) data [48], financial data [5], meteorological [17,49], sociological [50] or even random walk [51] and percolation [52] after translating whatever signal into a text based on an alphabet of k characters. The appearance of this power law is due to the presence of a so-called hierarchical structure of long range correlations in words, sentences, paragraphs, and so on for the given set of characters in an alphabet used for writing a text [46]. A simple extension of the Zipf analysis is to consider m-words only, i.e. the words strictly made of m characters without considering the white spaces.

Let for the sake of argument, only a binary alphabet with u and d characters, and the translation of a signal into a text (Fig.3). Let the probability to find a u in the text be p. The deviation from $p = \frac{1}{2}$, i.e. $p = \frac{1}{2} + \epsilon$ where $0 \leq \epsilon \leq \frac{1}{2}$ is called the bias. The bias is in fact a local measure of the trend in a stock price or index value.

We have chosen to examine $(m = 6, k = 2)$ cases. It may be remarked that this is useful for attempting to observe short range (weekly) fluctuations in (weather or financial) data for example. The aim of the study is to find the exponent ζ. By the way, it has been conjectured [53,54] that ζ is related to the Hurst exponent H, thus to the fractal dimension D [10,12,14] of the signal as

$$\zeta = |2H - 1|. \tag{17}$$

Therefore, for H different from 0.5, and thus ζ different from zero, the signal is not Brownian-like, whence some predictability can be expected because non trivial correlations exist between successive daily fluctuations.

One case can serve as an illustration herebelow. As *experimental data* among the many indices and stocks available on Internet, let us choose an insurance company Oxford Health Plan ($OXHP$), treated on the NASDAQ. From Aug. 8, 1991 till March 15, 1999, this consists in about 1900 data points [15]. The daily closing price signal is shown in Fig. 4. The fractal dimension D, or power spectrum characterized by β, the DFA exponent α and ζ can be examined as well. The latter from a Zipf analysis for the $OXHP$ closing price is given in Fig. 5, and values of exponents summarized in Table 2. The corresponding results for the Brownian motion are also given and serve as an estimate of the validity of the analysis.

Name of the signal	ζ	α	β
Brownian motion	0.08 ± 0.0007	0.50 ± 0.01	1.79 ± 0.20
OXHP: Closing price	0.27 ± 0.02	0.56 ± 0.03	1.75 ± 0.25

Table 2: The ζ, α and β values for OXHP closing price for the time spans from August 8, 1991 till March 15, 1999.

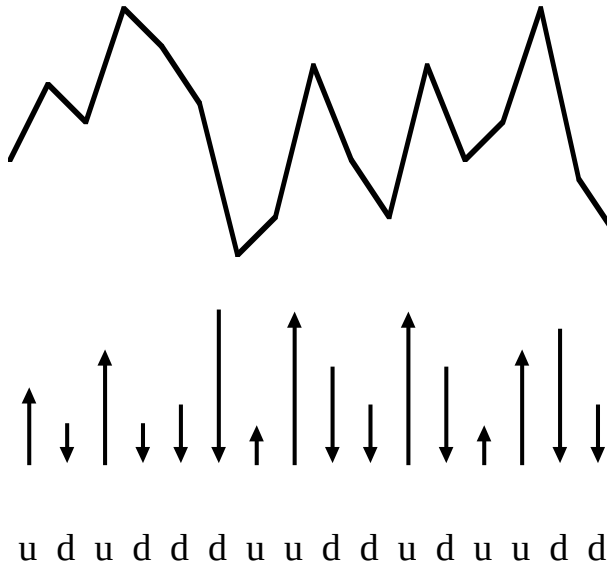

Fig. 3. Translation of part of a random walk sequence ("fluctuations") into a binary sequence made of two characters u and d.

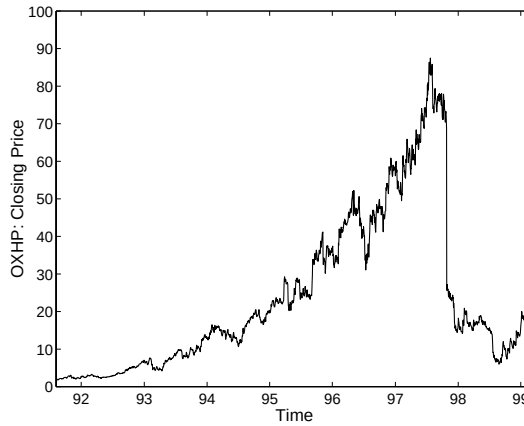

Fig. 4. Daily closing price of Oxford Health Plan ($OXHP$) stock, treated on the NASDAQ, from mid-91 till Jan. 99, i.e. about 1900 data points.

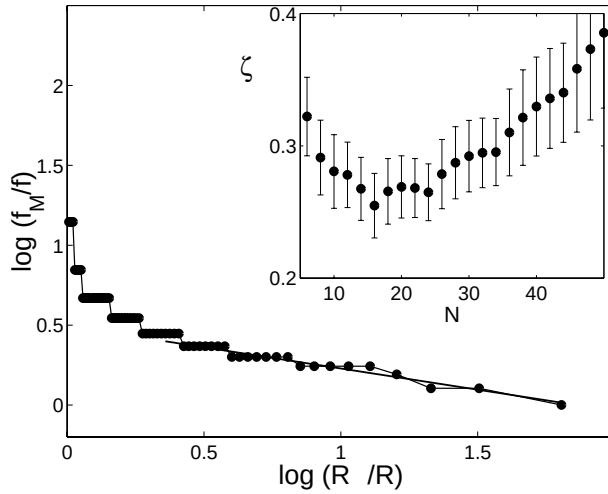

Fig. 5. (6,2)-Zipf analysis for the $OXHP$ stock closing price; the estimate of the ζ exponent is shown in the inset for R small and as a function of the number N of data points used to calculate the best slope from the main graph.

7 Basics of *i*-Variability Diagram Techniques

One disadvantage of the Zipf-method is that it is not possible to distinguish between persistent and antipersistent sequences. Only the departure from randomness is easily observed. Another way to sort out short range correlations is the i-Variability Diagram technique, used for example in heart beat [55] and meteorological [49] studies. Recall that the first return map (r_i, r_{i-1}) or the τ-return map $(r_i, r_{i-\tau})$ of a signal are often used for revealing a possible dominant correlation between the events of the data set. This leads to studies of *strange attractors* and the embedding dimension of a signal.

The return map of the *first derivative* of the signal, i.e. the so-called first order variability diagram (1-VD) [55] correlates every three consecutive points of the series,

$$\begin{aligned} s_{i+1} &= r_{i+1} - r_i \\ s_i &= r_i - r_{i-1} \end{aligned} \tag{18}$$

The curvature of the signal, thus relating every four consecutive events as

$$\begin{aligned} u_{i+1} &= r_{i+1} - 2r_i + r_{i-1} \\ u_i &= r_i - 2r_{i-1} + r_{i-2}. \end{aligned} \tag{19}$$

is called the second order variability diagram (2-VD).

The links between the 4-points e.g. defining the above relationship are seen through a phase space diagram for the curvature. A non-trivial shape of

the point cloud and point distribution itself on such a diagram indicate an asymmetry between the different consecutive curvatures, and can be used for predictability.

It has been found [56] that for free market financial series (DJIA and Gold price) the local trend behaves like a $\sqrt{t}$. However the BGL/USD exchange rate variability seems to be different: the set of events leads to a line structure with slope $I = -1.21$ in the curvature return map. The differences can be conjectured to depend on economic policy grounds.

A combination of Zipf and i-VD has been recently attempted for the local curvature correlations in financial signals [56]. This method leads to suggest tests based on microscopic models.

Acknowledgments

KI thanks the hospitality of T. Ackerman and the Department of Meteorology at Penn State University and NRC for a research grant during which part of this work was completed. Thanks to the F-532 grant of the Bulgarian NFSI as well. MA and NV thank the ARC 94-99/174 for financial support. MA thanks the organizers of the Ecole thématique de Chapelle des Bois and in particular Michel Planat for inviting him to present the above results.

References

1. L. Bachelier, *Théorie de la spéculation*, Ph. D. thesis, Acad. Paris (1900)
2. R.N. Mantegna and H.E. Stanley *Nature*, **376**, 46-49 (1995)
3. E.E. Peters, *Fractal Market Analysis : Applying Chaos Theory to Investment and Economics*, (Wiley Finance Edition, New York, 1994)
4. M. Ausloos, *Europhys. News* **29**, 70-72 (1998)
5. N. Vandewalle and M. Ausloos, in *Fractals and Beyond. Complexity in the Sciences*, M.M. Novak, Ed. (World Scient., Singapore, 1999) p.355-356.
6. N. Vandewalle and M. Ausloos, *Physica A*, **246**, 454-459 (1997)
7. N. Vandewalle, Ph. Boveroux, A. Minguet and M. Ausloos, *Physica A* **255**, 201-210 (1998)
8. H. E. Stanley, S. V. Buldyrev, A. L. Goldberger, S. Havlin, C.-K. Peng and M. Simmons, *Physica A* **200**, 4-24 (1996)
9. N. Vandewalle and M. Ausloos, *Int. J. Comput. Anticipat. Syst.* **1**, 342-349 (1998)
10. B. J. West and B. Deering, *The Lure of Modern Science: Fractal Thinking*, (World Scient., Singapore, 1995)
11. C.-K. Peng, J.M. Hausdorff, S. Havlin, J.E. Mietus, H.E. Stanley, and A.L. Goldberger, *Physica A* **249**, 491-500 (1998)
12. M. Schroeder, *Fractals, Chaos and Power Laws*, (W.H. Freeman and Co., New York, 1991)
13. K. J. Falconer, *The Geometry of Fractal Sets*, (Cambridge Univ. Press, Cambridge, 1985)

14. P. S. Addison, *Fractals and Chaos*, (Inst. of Phys., Bristol, 1997)
15. K. Ivanova, *unpublished*
16. N. Vandewalle and M. Ausloos, *Int. J. Phys. C* **9**, 711-720 (1998)
17. K. Ivanova and M. Ausloos, *Eur. J. Phys. B* **8** 665-669 (1999)
18. N. Vandewalle and M. Ausloos, *Eur. J. Phys. B* **4**, 257-261 (1998)
19. A. L. Barabási and T. Vicsek, *Phys. Rev. A* **44**, 2730-2733 (1991)
20. A. Davis, A. Marshak, and W. Wiscombe, in *Wavelets in Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds. (Academic Press, New York, 1994) pp. 249-298
21. A. Marshak, A. Davis, R. Cahalan, and W. Wiscombe *Phys. Rev. E* **49**, 55-69 (1994)
22. A.-L. Barabási and H.E. Stanley, *Fractal Concepts in Surface Growth*, (Cambridge Univ. Press, Cambridge, 1995)
23. J. Feder, *Fractals*, (Plenum, New-York, 1988)
24. E. Labie, *private communication*
25. A.G. Ellinger, *The Art of Investment*, (Bowers & Bowers, London, 1971)
26. J.-P. Bouchaud and A. Georges, *Phys. Rep.* **195**, 127-294 (1990)
27. N. Vandewalle and M. Ausloos, *Phys. Rev. E* **58**, 6832-6834 (1998)
28. L. Frachebourg, P.L. Krapivsky, and S. Redner, *Phys. Rev. E* **55**, 6684-6689 (1997)
29. D. Sornette, A. Johansen, and J.-Ph. Bouchaud, *J. Phys. I (France)* **6**, 167-175 (1996)
30. H.E. Stanley, *Phase Transitions and Critical Phenomena*, (Oxford Univ. Press, Oxford, 1971)
31. D. Sornette, *Physics Reports* **297**, 239-270 (1998)
32. J.C. Anifrani, C. Le Floc'h, D. Sornette and B. Souillard, *J. Phys. I (France)* **5**, 631-63 (1995) 277
33. D. Stauffer and D. Sornette, *Physica A* **252**, 271- (1998)
34. N. Vandewalle, R. D'hulst, and M. Ausloos, *Phys. Rev. E* **59**, 631-635 (1999)
35. M. Ausloos and N. Vandewalle, *unpublished*
36. N. Vandewalle and M. Ausloos, *Eur. J. Phys. B* **4**, 139-141 (1998)
37. H. Dupuis, *Trends Tendances* **22**(38), (september 18, 1997) p.26-27
38. H. Dupuis, *Trends Tendances* **22**(44), (october 30, 1997) p.11
39. D. Daoût, *Le Vif L'Express* (october 30, 1997) p.124-125
40. D. Sornette and D. Stauffer, *private communicaion*
41. J.-Ph. Bouchaud, P. Cizeau, L. Laloux, and M. Potters, *Phys. World* **12**, 25-29 (1999)
42. L. Laloux, M. Potters, R. Cont, J.-P. Aguilar, and J.-P. Bouchaud, *Europhys. Lett.* **45**, 1-5 (1999)
43. J.R. Macdonald and M. Ausloos, *Physica A* **242**, 150-160 (1997)
44. M. Ausloos, *J. Phys. A* **22**, 593-609 (1989)
45. G.K. Zipf, *Human Behavior and the Principle of Least Effort*, (Addisson-Wesley, Cambridge, MA, 1949)
46. W. Ebeling and A. Neiman, *Physica A* **215**, 233-241 (1995)
47. B. Vilensky *Physica A* **231**, 705-711 (1996)
48. M.H.R. Stanley, S.V. Buldyrev, S. Havlin, R.N. Mantegna, M.A. Salinger, and H.E. Stanley, *Ecomonics Letters*, **49**, 453-457 (1995)
49. K. Ivanova, M. Ausloos, A. Davis, and T. Ackerman, *XXIV General Assembly of the EGS*, the Hague, Netherlands, April 19-23 (1999)

50. M. Marsili and Y.-C. Zhang, *Phys. Rev. Lett.* **80**, 2741-2744 (1998)
51. N. Vandewalle and M. Ausloos, *unpublished*
52. M. S. Watanabe *Phys. Rev. E* **53**, 4187-4190 (1996)
53. A. Czirok, R.N. Mantegna, S. Havlin, and H.E. Stanley, *Phys. Rev. E* **52**, 446-452 (1995)
54. G. Troll and P.B. Graben, *Phys. Rev. E* **57**, 1347-1355 (1998)
55. A. Babloyantz and P. Maurer, *Phys. Lett. A* **221**, 43-55 (1996)
56. K. Ivanova and M. Ausloos, *Physica A* **265**, 279-286 (1999).

Oscillators and the Characterization of Frequency Stability: an Introduction

Vincent Giordano[1⋆] and Enrico Rubiola[1,2⋆⋆]

[1] LPMO – CNRS UPR3203 associée à l'Université de Franche-Comté, 32 Rue de l'Observatoire, F-25044 Besançon, France
[2] Politecnico di Torino, Dipartimento di Elettronica, and INFM UDR Politecnico c.so Duca degli Abruzzi no. 24, I-10129 Torino, Italy

Abstract. This paper provides an introduction to basic concepts commonly used in time and frequency metrology, and is addressed to other scientific communities. Thus no attempt is made to provide an exhaustive review. Instead, attention is focused on the most important tools used by physicists and engineers involved in time and frequency metrology. We first explain the principles of the oscillator through an example. Then we introduce the concepts of frequency reference, oscillating loop, frequency stability and accuracy. Finally, we define the power spectrum density of frequency fluctuations and the Allan variance as means to characterize the stability of frequency standards.

1 Introduction

The ideas presented in this paper originated in the early 1960s, and have been widely exploited not only for research purposes, but also for the technological development of time and frequency industry. These ideas have led to a rapid progress in the knowledge of noise in electronic circuits and oscillators, and to an impressive stability improvement of frequency standards and clocks. Original works and further information dealing with the characterization of frequency stability can be found in [1–4]. Up-to-date information on various topics – both fundamental and applied – on time and frequency metrology can be found in the proceedings of the conferences on that subject [5–7]. Finally, we advise consulting some books dealing with oscillators and atomic frequency standards [8–10].

2 Basic Definitions and Principles

The ideal oscillator can be regarded as a black box that transforms the supplied energy into a sinusoidal signal $A(t)$ of amplitude A_0 and frequency ν_0

$$A(t) = A_0 \cos(2\pi\nu_0 t) \; . \tag{1}$$

⋆ giordano@lpmo.univ-fcomte.fr
⋆⋆ rubiola@polito.it

M. Planat (Ed.): LNP 550, pp. 175–188, 2000.
© Springer-Verlag Berlin Heidelberg 2000

Actual oscillators show frequency offset, drift and fluctuations that are not accounted in (1). We restrict our attention to frequency stability, discarding all the problems connected with amplitude stability. The main reason for this is that frequency is integrated over a long time, while amplitude has just a local effect. Characterization methods are therefore needed to describe oscillator non-idealities and to provide efficient tools suitable to compare the performances of frequency standards. The concepts related to frequency stability are being introduced taking as an example the traditional grandfather's clock known in the region of Chapelle des Bois as the *horloge comtoise*.[1]

2.1 The pendulum as a frequency reference

As the frequency of an oscillator is determined by a resonator, the overall performance of the oscillator relies upon the performance of that resonator. In the case of the horloge comtoise, this oscillator is a gravitational pendulum of length L_0 and mass m (Fig. 1).

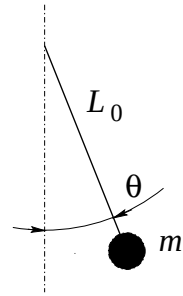

Fig. 1. Freely swinging pendulum

If the pendulum is set off from its equilibrium state, it gets potential energy. The mass let free starts swinging, continuously interchanging potential and kinetic energy. In the small oscillation approximation, which consists of replacing $\sin\theta = \theta$ in the derivation of the motion equations, and neglecting losses, the total energy conservation yields

$$\ddot{\theta}(t) + \frac{g}{L_0}\,\theta(t) = 0\,, \tag{2}$$

where $g \simeq 9.81$ m/s^2 (at the sea level) is the local gravity acceleration. The general solution of the above equation is

$$\theta(t) = \theta_0\ \cos\left(\frac{2\pi}{T_0}t\right)\ . \tag{3}$$

[1] "horloge comtoise" is the French term for the traditional clock made in the French region of Franche-Comté.

Thus, the pendulum swings regularly with oscillation period

$$T_0 = 2\pi\sqrt{\frac{L_0}{g}} \tag{4}$$

and peak amplitude θ_0. The number of cycles per unit of time is the oscillation frequency $\nu_0 = 1/T_0$.

As we are confident of the *horloge comtoise* – the Franche Comté artisans are quite competent watchmakers – we can base the unit of time on it. A possible definition could be: "the 'Franche Comté second' is the period of a pendulum 24.9 mm long freely swinging in vacuum at Chapelle des Bois (France) on August 11th 1999 at noon, during the millennium eclipse". This definition includes date and place, which is needed to define the local value of g; thus the definition can be extrapolated to all other cases. A time scale, consisting of a counter incremented at each period of the pendulum oscillation, is the coordinate frame on which we can date coffee break, lunch, and any other relevant event. Our pendulum can also be used as a frequency standard. A simple way to compare an oscillator to our frequency standard is the measurement of the beat-note. If two frequencies are not sufficiently close, a gearbox – i.e. a frequency synthesizer – may be inserted.

2.2 Damped and stationary oscillations

Due to dissipation forces acting on the rod and mass, actual resonators are damped. Thus the angular position results from

$$\ddot{\theta}(t) + 2\xi\dot{\theta}(t) + \frac{g}{L_0}\,\theta(t) = 0\,. \tag{5}$$

If losses are negligible ($\xi \ll 2\pi\nu_0$), the solution is

$$\theta(t) = \theta_0 \exp\left(-\frac{t}{\tau}\right)\cos\left(\frac{2\pi}{T_0}\,t\right). \tag{6}$$

Oscillation is exponentially damped with a time constant $\tau_a = 1/\xi$, and therefore the total energy decreases with a time constant $\tau_e = 1/2\xi$. The resonator loss is often reported in terms of the *merit factor* Q, also called *quality factor*

$$Q = \pi\frac{\tau_a}{T_0}\,, \tag{7}$$

that is the time constant of the energy decay normalized over the duration of one radian.

Continuous operation requires stationary amplitude, thus power must be spent to compensate for losses. In the grandfather's clock this is accomplished by an escapement that also ensures the proper phase relation of the stimulus

with respect to the rod position. Obviously, the frequency of the steady state oscillation and of the external stimulus must be the same. The pendulum can be regarded as a bandpass filter whose −3 dB bandwidth is $\Delta\nu = Q_0\nu_0$. The higher the resonator Q factor is, the better is the spectral definition of the resonance. Hence, the value of Q has a direct impact on the frequency stability of oscillators.

2.3 Frequency stability

As a result of experimental observations, frequency stability turns out not to be a single and constant parameter. Even in the presence of stationary processes only, it depends on the observation time. Thus we have to distinguish short term and long term stability, and in case of need, medium term stability.

The short term frequency stability of our pendulum may be limited by the brownian motion of the residual gas in the vacuum chamber. Due to the high collision rate and to the small mass of the molecules, the collision pulses are averaged by the mass of the pendulum, and therefore frequency stability increases proportionally to the observation time. As the sustaining force is controlled by a position detector, the noise of that detector originates additional frequency fluctuations. The Q factor turns out to be equal to the ratio P_r/P_d between the reactive power P_r and the dissipated power P_d. High Q means low dissipated energy and consequently low coupling to the environment, which warrants high short term stability.

Some other physical effects act directly on the resonator frequency, turning the environment fluctuations into frequency instability. In our pendulum, this may occur with the thermal expansion of the rod and with the fluctuation of the local gravity acceleration g. Because there is no averaging mechanism, these phenomena acts in the long term. In addition, aging due to deterministic or quasi-deterministic phenomena may be a relevant factor.

2.4 Accuracy

The ideal conditions assumed in the definition of the unit of time are never realized in practice [11]. In our example, the actual rod length differs from L_0. We can measure that length and correct the clock frequency, but a residual uncertainty remains. If, for example, $\delta L_0/L_0 = 10^{-3}$ is the rod length uncertainty, the frequency uncertainty derived from (4) is $\delta\nu/\nu_0 = \frac{1}{2}\delta L/L_0 = 5{\times}10^{-4}$. The uncertainty of other parameters acting on ν_0 can be accounted in similar ways. The overall accuracy of the clock results from the sum of the individual uncertainties, added with the appropriate statistical rules.

Finally, actual accuracy limitations may be divided in two types, namely the uncertainty of the realization of the SI unit of time, which applies to fundamental metrology, and the residual uncertainty after calibrating against a primary standard.

3 Electronic Oscillators

The electronic oscillator is described by the same differential equations as the pendulum, provided that mass and force are replaced by voltage and current. Accordingly, the RLC resonator (Fig. 2) consists of an inductance L and a capacitance C, plus a resistance R_0 that accounts for losses.

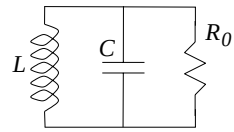

Fig. 2. RLC resonator

The resonator acts as a band-pass filter characterized by the center frequency $\nu_0 = 1/2\pi\sqrt{LC}$ and by the -3 dB bandwidth $\Delta\nu = \nu_0/Q$. The frequency response of the parallel RLC filter (Fig. 3) shows its maximum at $\nu = \nu_0$, where the phase lag is 0.

Let us now consider the circuit of Fig. 4. Due the positive feedback, the transfer function is

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{G}{1 - GA} \,. \tag{8}$$

If $GA = 1$, the denominator of (8) is zero. In this condition, the switch-on transient circulates permanently in the loop, and consequently no input is necessary for a periodic signal to be present at the output (Fig. 5). The con-

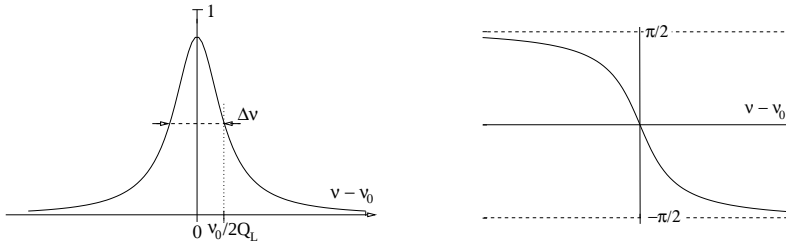

Fig. 3. Modulus (left) and phase (right) of the parallel RLC filter response to a current excitation

dition $GA = 1$, often referred as the *Barkhausen condition*, involves complex quantities, hence $|GA| = 1$ and $\arg GA = 0$ must be ensured separately. Usually the amplitude condition is guarented by saturation of the amplifier or by a separate gain control, while the phase condition determines the actual oscillation frequency. As a consequence, the effect of a phase perturbation

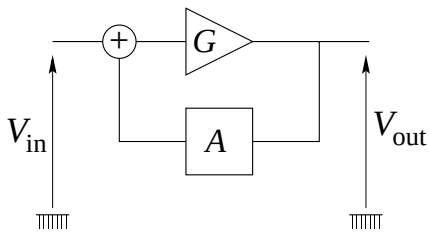

Fig. 4. Amplifier with positive feedback loop

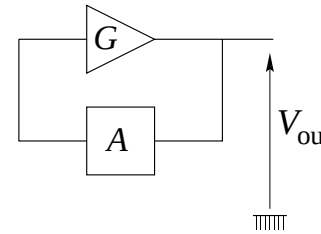

Fig. 5. Oscillator

$\Delta\varphi$ in the loop is a phase variation $-\Delta\varphi$ in the resonator, which causes a frequency change

$$\frac{\Delta\nu}{\nu_0} = \frac{\Delta\varphi}{2Q} \ . \tag{9}$$

Obviously, this phase-to-frequency conversion also takes place with random phase fluctuations of any part of the loop, whereat higher values of Q yield better frequency stability. Table 1 shows the merit factor of some resonator types.

Table 1. Merit factor of some resonators

resonator	frequency	Q
pendulum swinging in vacuum	1 Hz	10^3
watch quartz crystal	32768 Hz	5×10^4
Bulk Acoustic Wave (BAW) quartz crystal	10 MHz	10^6
Surface Acoustic Wave (SAW) quartz crystal [12]	400 MHz	1.6×10^4
dielectric resonator [13]	10 GHz	5×10^3
cryogenic sapphire (4 K) whispering gallery [14]	10 GHz	10^9

4 Frequency-Domain Characterization of Frequency Stability

Due to internal noise sources, the amplitude and the frequency of the oscillator output signal fluctuate. In general, amplitude is sufficiently stable for applications, while phase is not. This occurrs because frequency is integrated over a long time. Hence a term $\varphi(t)$ must be introduced in (1), that becomes

$$A(t) = A_0 \cos\left[2\pi\nu_0 t + \varphi(t)\right] \ . \tag{10}$$

The above model does not contain information on the nature of $\varphi(t)$. The latter can result from either deterministic (drift or modulation) or random processes. Nevertheless, as we are interested in random fluctuations, we assume that all deterministic phenomena are removed. In addition, we assume that $\varphi(t)$ is stationary and derivable. These hypotheses, despite some intrinsic difficulties [2], enable the description of experimental results.

The instantaneous frequency of $A(t)$ is

$$\nu(t) = \frac{1}{2\pi} \frac{d}{dt} \left[2\pi\nu_0 + \varphi(t)\right] , \tag{11}$$

from which we define the fractional frequency offset

$$y(t) \;=\; \frac{\Delta\nu(t)}{\nu_0} \;=\; \frac{1}{2\pi\nu_0} \frac{d\varphi(t)}{dt} . \tag{12}$$

As $\varphi(t)$ is stationary, $y(t)$ also is stationary. Therefore the autocorrelation function

$$R_y(\tau) = \langle y(t)\, y(t-\tau) \rangle \tag{13}$$

is a function of a single variable τ. The Fourier transform of $R_y(\tau)$ is the double-sided power spectrum density (PSD) of y

$$S_y^{DS}(f) = \int_{-\infty}^{\infty} R_y(\tau) \exp\left(-2i\pi f\tau\right)\, d\tau . \tag{14}$$

As instruments generally show positive frequencies only, the one-sided power spectrum density

$$S_y(f) \;=\; \begin{cases} 2S_y^{DS}(f) & \text{for} \quad f \geq 0 \\ 0 & \text{for} \quad f < 0 \end{cases} \tag{15}$$

is generally preferred to $S_y^{DS}(f)$.

Finally, the PSD of other quantities related to phase fluctuations are of interest. Table 2 reports a summary.

4.1 The power law model

As a result of experimental observations, the instability of actual oscillators is well approximated by the so called power law

$$S_y(f) = \sum_{\alpha=-2}^{2} h_\alpha f^\alpha , \tag{16}$$

which corresponds to a superposition of the five independent noise processes reported in Table 3. As an example, Fig. 6 shows the typical spectrum density $S_y(f)$ of some oscillators.

Table 2. Useful quantities and their spectrum densities

quantity	definition	PSD	PSD unit
fractional frequency fluctuations	$y(t) = \frac{1}{2\pi\nu_0}\frac{d\varphi(t)}{dt}$	$S_y(f)$	Hz^{-1}
phase fluctuations	$\varphi(t)$	$S_\varphi(f) = \frac{\nu_0^2}{f^2} S_y(f)$	$\mathrm{rad}^2\mathrm{Hz}^{-1}$
angular frequency fluctuations	$\dot{\varphi}(t) = \frac{d\varphi(t)}{dt}$	$S_{\dot{\varphi}}(f) = (2\pi\nu_0)^2 S_y(f)$	$\mathrm{rad}^2\mathrm{Hz}$
frequency fluctuations	$\Delta\nu(t) = \frac{1}{2\pi}\frac{d\varphi(t)}{dt}$	$S_{\Delta\nu}(f) = \nu_0^2 S_y(f)$	Hz^{-1}

Table 3. Summary of the oscillator noise types

noise type	$S_y(f)$	$S_\varphi(f)$
frequency random walk	$h_{-2}f^{-2}$	$\nu_0^2\, h_{-2}f^{-4}$
frequency flicker noise	$h_{-1}f^{-1}$	$\nu_0^2\, h_{-1}f^{-3}$
white frequency noise	h_0	$\nu_0^2\, h_0 f^{-2}$
flicker phase noise	$h_1 f$	$\nu_0^2\, h_1 f^{-1}$
white phase noise	$h_2 f^2$	$\nu_0^2\, h_2$

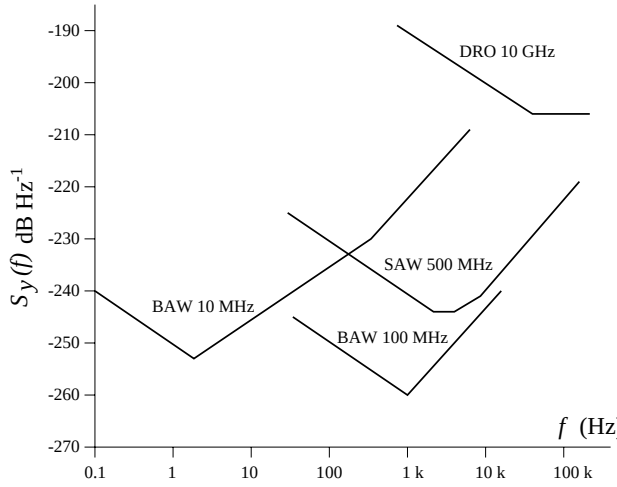

Fig. 6. $S_y(f)$ for some oscillators. DRO stands for dielectric resonator oscillator.

The phenomenological model f^α suffers from a basic difficulty due to the fact that the total power

$$P = \int_0^\infty S_y(f)\, df \tag{17}$$

is infinite. In fact, $\int f^\alpha\, df$ diverges at high frequencies for $\alpha \leq -1$, and at low frequencies for $\alpha \geq -1$; in addition, a noise process that diverges for $f \to 0$, i.e. in the long term, can be regarded as a non-stationary one. But actual measurements make sense, in spite of these theoretical difficulties. In fact, the measurement bandwidth is bounded by the maximum frequency of the instrument input stage and by low frequency that results from the finite duration of the measurement. While the upper bound can be easily accounted for, the lower bound requires a different formalism. That is why variance σ^2 must be replaced with the Allan variance $\sigma_y^2(\tau)$, as described in the next section.

5 Time-Domain Characterization of Frequency Stability

For technical reasons, commercially available counters are not suitable to measure slow frequency or phase variations, which would require an extremely high resolution. Therefore, the *beat note* method (Fig. 7) is commonly used as a replacement for direct frequency measurement schemes.

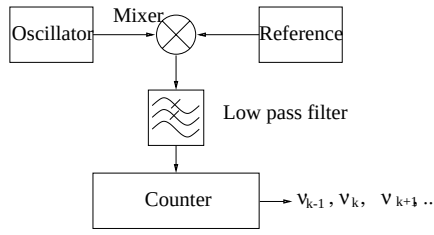

Fig. 7. The beat note method

A reference frequency ν_0 must be chosen close to the oscillator frequency ν, so that the beat frequency $\nu_b = |\nu - \nu_0|$ is of some 1–100 Hz. As the low pass filter removes the $\nu + \nu_0$ component, the counter measures ν_b only. Because the measurement resolution is a critical point, it is necessary to take ν_b from the measurement of the beat period $T_b = 1/\nu_b$ averaged over a suitable interval τ; obviously, τ is an integer multiple of T_b. In a well designed system the quantization noise $\delta\nu_b = \nu_b/(\tau\nu_c)$ is made lower than the analog circuit noise; ν_c is the counter clock frequency.

The output of the counter is a stream of values

$$\overline{\nu_{b\,k}} = \frac{1}{\tau} \int_{t_k-\tau}^{t_k} \nu_b(\theta)\, d\theta \tag{18}$$

of the beat frequency averaged over τ, that is converted into a stream of fractional frequency fluctuation values

$$\overline{y_k} = \frac{1}{\tau} \int_{t_k-\tau}^{t_k} y(\theta)\, d\theta\ . \tag{19}$$

The measurement intervals of duration τ are assumed to be contiguous, which means that there is no dead time. With modern equipment this is a generally good approximation of the reality. If the dead time is not negligible, results must be corrected using a suitable model, such as [15]. Alternatively two counters can be swapped, so that one counter can measure when the other one is busy with data transfer or with internal housekeeping [16].

Yet, the above details should not distract from the main problem, that is the suitability of statistical tools to represent frequency fluctuations.

True variance. The true variance, indicated as σ^2, is widely used in statistics as a means to characterize random variables. In frequency metrology, however, the true variance is often meaningless if the measurement time is not clearly indicated. The variance of y is

$$\begin{aligned} I^2(\tau) &= \left\langle\, \overline{y}^2 \,\right\rangle \\ &= \frac{1}{\tau^2} \left\langle \left[\int_{t_k-\tau}^{t_k} y(t)\, dt \right]^2 \right\rangle\ . \end{aligned} \tag{20}$$

The process of averaging over τ is equivalent to a filter whose response is defined in Fig. 8. Making this filter appear explicitely in (20) we get

$$I^2(\tau) = \left\langle \left[\int_{-\infty}^{\infty} y(t') h_1(t-t')\, dt' \right]^2 \right\rangle\ . \tag{21}$$

The variance turns out to be easier to evaluate in the frequency domain as

$$I^2(\tau) = \int_0^\infty S_y(f)\, |H_1(f)|^2\, df\ , \tag{22}$$

where

$$|H_1(f)|^2 = \left(\frac{\sin \pi\tau f}{\pi\tau f} \right)^2 \tag{23}$$

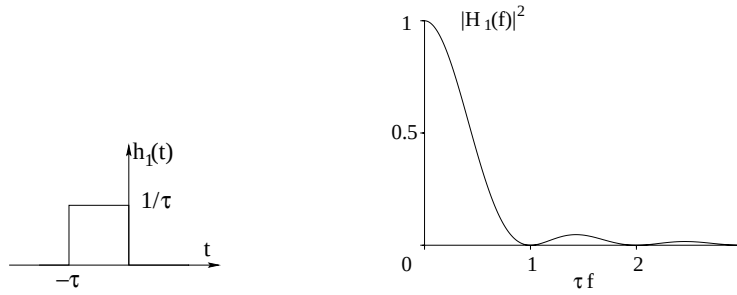

Fig. 8. Impulse response $h_1(t)$ and frequency response $|H_1(f)|^2$ of the equivalent filter inherent in the measurement of the true variance

is the filter frequency response (Fig. 8).

The key point is that the integral (22) is undefined for $\alpha \geq -1$, that corresponds to the flicker and random walk frequency noise types. Yet, the lower integration bound can not approach zero because it is limited by the duration T of the measurement. There results a high pass filter whose cutoff frequency is of the order of $1/T$. As a consequence, the estimate of $I^2(\tau)$ depends on the number of samples involved in the averaging process. That is why the true variance is scarcely useful for oscillator characterization.

Two-sample variance. The two-sample variance, also known as the Allan variance, is defined as

$$\sigma_y^2(\tau) = \frac{1}{2}\left\langle \left(\overline{y}_{k+1} - \overline{y}_k\right)^2 \right\rangle \ . \tag{24}$$

It is worth noting that the Allan Variance is a simple case of wavelet analysis [17].

Expanding (24), $\sigma_y^2(\tau)$ can be rewritten as

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2}\left\langle \left[\int_{t_k}^{t_k+\tau} y(t')dt' - \int_{t_k-\tau}^{t_k} y(t')dt'\right]^2 \right\rangle \ . \tag{25}$$

This measurement mechanism, that involves the difference of two contiguous integrals, is equivalent to the band bandpass filter whose response is shown in Fig. 9. Therefore

$$\sigma_y^2(\tau) = \frac{1}{2}\left\langle \left[\int_{-\infty}^{\infty} y(t')h_2(t-t')dt'\right]^2 \right\rangle \tag{26}$$

Consequently, $\sigma_y^2(\tau)$ is related to $S_y(f)$ by

$$\sigma_y^2(\tau) = \int_0^{\infty} S_y(f)\,|H_2(f)|^2\,df \tag{27}$$

where

$$|H_2(f)|^2 = 2\,\frac{\sin^4 \pi\tau f}{(\pi\tau f)^2} \tag{28}$$

is the filter frequency response (Fig. 9 right).

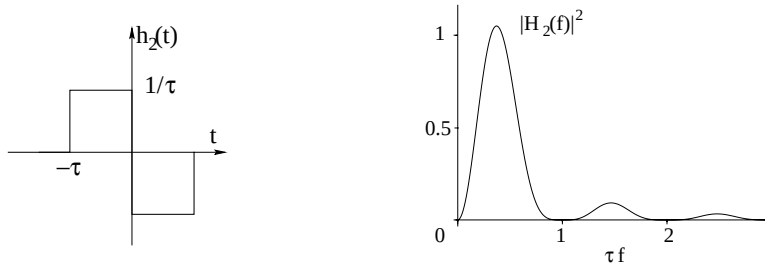

Fig. 9. Impulse response $h_1(t)$ and frequency response $|H_2(f)|^2$ of the filter introduced Allan variance measurement process

For low Fourier frequencies $|H_2(f)|^2$ is proportional to f^2, hence the integral (27) exists for both flicker frequency and random walk frequency noise types. For this reason, $\sigma_y^2(\tau)$ is suitable to characterize frequency stability in the time domain.

Actual instruments always show a high-frequency cut-off that limits the contribution of rapid frequency fluctuations. Then, the estimation of the two-sample variance is affected by the cut-off frequency f_h of this filter, as well as the roll-off This is relevant for flicker and white phase noise only.

Table 4 gives the relations between $\sigma_y^2(\tau)$ and the spectrum for the different noise types. These relations holds for $\tau \gg 1/2\pi f_h$, under the assumption of a sharp low-pass filter at the cut-off frequency f_h.

To conclude, Fig 10 shows the typical frequency stability of some clocks and oscillators. Data are taken from [19–22].

References

1. Proc. *IEEE-NASA Symposium on Short Term Frequency Stability*, Greenbelt (MD, USA), 23–24 November 1964.
2. J. A. Barnes, *Models for the interpretation of frequency stability measurements*, NBS Technical Note no. 683, August 1976.
3. J. Rutman, "Characterization of phase and frequency instabilities in precision frequency sources: fifteen years of progress", *Proc. of the IEEE*, Vol. 66, no. 9 pp. 1048–1075, September 1978.
4. E. S. Ferre-Pikal, J. R. Vig, J. C. Camparo, L. S. Cutler, L. Maleki, W. J. Riley, S. R. Stein, C. Thomas, F. L. Walls, J. D. White, "Draft revision of

Table 4. Relations between $S_y(f)$ and $\sigma_y^2(\tau)$

Noise type	$S_y(f)$	$\sigma_y^2(\tau)$
random walk frequency	$h_{-2}f^{-2}$	$\frac{2\pi^2}{3}\ h_{-2}\tau$
flicker frequency	$h_{-1}f^{-1}$	$2h_{-1}\ \ln 2$
white frequency	h_0	$\frac{h_0}{2\tau}$
flicker phase	$h_1 f$	$\frac{3h_1 \ln(2\pi f_h \tau)}{4\pi^2\tau^2}$
white phase	$h_2 f^2$	$\frac{3h_2 f_h}{4\pi^2\tau^2}$

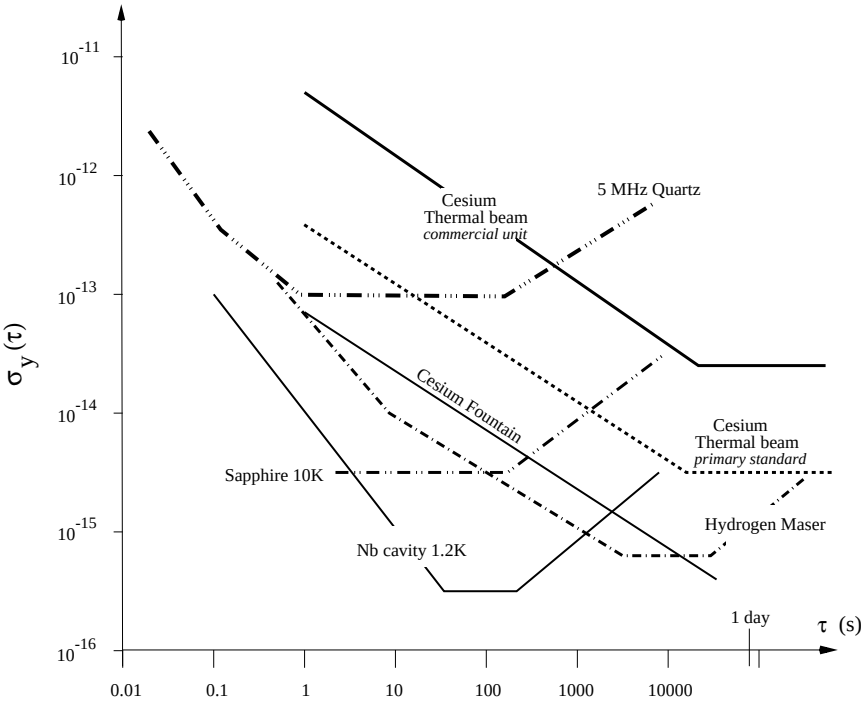

Fig. 10. Frequency stability of some frequency standards

IEEE STD 1139-1988 standard definitions of physical quantities for fundamental frequency and time metrology – random instabilities", *Proc. 51st Frequency Control Symposium* pp. 338–357, Orlando (FL, USA), 28–30 May 1997.
5. *European Frequency and Time Forum*. International conference held in Europe every year since 1987. URL http://www.fsrm.ch/eftf.
6. *Frequency Control Symposium*. International IEEE conference of the UFFC Society, hold every year in the USA. URL http://www.ieee.org.
7. *Precise Time and Time Interval*. International conference hold every year in USA. URL http://192.5.41.239/ptti.html.
8. E. A. Gerber, A. Ballato, *Precision Frequency Control*, Academiic Press, 1985. ISBN 0-12-280601-8 (vol. 1), and 0-12-280602-6 (vol. 2).
9. J. Vanier, C. Audoin, *The Quantum Physics of Atomic Frequency Standards*, Adam Hilger 1989. ISBN 0-85274-434-2.
10. Chronos Group, *Frequency Measurement and Control*, Chapman and Hall 1994. ISBN 0-412-48270-3.
11. C. Audoin, N. Dimarcq, V. Giordano, J. Viennet, "Physical origin of the frequency shifts in cesium beam frequency standards – related environmental sensitivity", *IEEE Trans. on UFFC* vol. 39 no. pp. 412–421, May 1993.
12. A. El Habti, F. Bastien, E. Bigler, T. Thorvaldsson, "High-frequency surface acoustic wave devices at very low temperature: Application to loss mechanisms evaluation", *J. Acoust. Soc. Am.* vol. 100 no. 1 pp. 272–277, July 1996.
13. D. Kajfez, Guillon, *Dielectric Resonators*, Artech House 1986.
14. A. N. Luiten, A. G. Mann, D.G. Blair, "Ultrahigh Q-factor cryogenic sapphire resonator.", *Electronics Letters* vol. 29 no. 10 pp. 879–881, May 1993.
15. J. A. Barnes, *Tables of bias functions B_1 and B_2 for variances based on finite samples of processes with power law spectral densities*, NBS Tech. Note no. 375, Washington D.C., January 1969.
16. P. Lesage, C. Audoin,"Effect of dead-time on the estimation of the two-sample variance", *IEEE Trans. on Instrumentation and Measurement* vol. IM-28 no. 1 pp. 6–10, March 1979.
17. P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion", *IEEE Trans. on Information Theory* vol. 38 no. 2 pp. 910–917, March 1992.
18. P. Lesage, C. Audoin,"Charecterization of frequency stability: uncertainty due to the finite number of measurements", *IEEE Trans. on Instrumentation and Measurement* vol. IM-22 no. 4 pp. 157–161, June 1973.
19. A. G. Mann, G. Santarelli, S. Chang, A. N. Luiten, P. Laurent, C. Salomon, D. G. Blair, A. Clairon, "A high stability atomic fountain clock using a cryogenic sapphire interrogation oscillator", *Proc. 52nd Frequency Control Symposium* pp. 13–17, Pasadena (CA, USA), 27–29 May 1998.
20. G. J. Dick, R. T. Wang, R. L. Tjoelker,"Cryo-cooled sapphire oscillator with ultra-high sability", *Proc. 52nd Frequency Control Symposium* pp. 528–533, Pasadena (CA, USA), 27–29 May 1998.
21. V. Candelier, J. Chauvin, C. Gellé, G. Marotel, M. Brunet, R. Petit, "Ultrastable oscillators", *Proc. 12th European Frequency and Time Forum* pp. 345–351, Warsaw (Poland), 10–12 March 1998.
22. V. Giordano, "Noise sources in frequency standards", *Ann. Télécommun.* vol. 51 no. 7-8 pp. 335–343, July 1996.

Phase Noise Metrology

Enrico Rubiola[1,2⋆] and Vincent Giordano[2⋆⋆]

[1] Politecnico di Torino, Dipartimento di Elettronica, and INFM UDR Politecnico c.so Duca degli Abruzzi no. 24, I-10129 Torino, Italy
[2] LPMO – UPR-3203 du CNRS associée à l'Université de Franche-Comté 32 Rue de l'Observatoire, F25044 Besançon, France

Abstract. As a result of a major technological trend towards high speed digital communications and circuits, phase noise turns out to be a relevant concern for scientists and engineers. This paper describes methods and instruments to measure the phase noise of oscillators, components and more complex devices in the radiofrequency and microwave bands, from approximately 100 kHz to 30–40 GHz, and even beyond. After a brief introduction, two sections deal with basic definitions and traditional methods, and one section presents a set of schemes that cover most actual needs. Then a new approach – known as the *interferometric method* – is discussed in detail, providing design strategies and examples; this method exhibits the highest sensitivity in real time, which can alse be exploited to dynamically correct the phase noise of amplifiers and oscillators. The last section deals with an improved version of the interferometric method, in which correlation is used to remove the instrument noise of two equal interferometers that simultaneously measure the same device. This scheme enables the measurement of low noise processes, even below the thermal floor, and therefore it represents the state of the art in the high sensitivity phase noise metrology.

1 Introduction

This paper deals with the measurement of the phase noise of radiofrequency and microwave signals. As one can expect, we are mainly interested in the measurement of *low noise* signals.

We first define a quasi-sinusoidal signal of the form

$$s(t) = \sqrt{2R_0P_c}\,[1+\alpha(t)]\cos[2\pi\nu_c t + \varphi(t)]\,. \tag{1}$$

where R_0 is the characteristic impedance and P_c is the carrier power. $\varphi(t)$ and $\alpha(t)$ are realizations of random processes that we call *phase noise* and (relative) *amplitude noise*, respectively. By definition $\alpha(t)$ and $\varphi(t)$ have zero mean, which results from an appropriate choice of P_c and of the time axis origin. Phase noise is our main concern, while amplitude noise can also be of interest in a smaller set of problems and applications. The phase noise is commonly described in terms of $S_\varphi(f)$, i.e. the power spectrum density (PSD) of $\varphi(t)$.

⋆ rubiola@polito.it
⋆⋆ giordano@lpmo.univ-fcomte.fr

M. Planat (Ed.): LNP 550, pp. 189–215, 2000.
© Springer-Verlag Berlin Heidelberg 2000

Why scientists are so worried by phase noise? As this question involves many fields of science and technology, we try to answer it through some examples.

First of all, the replacement of analog circuits with digital electronics at higher and higher clock frequencies is a major technological trend. In a digital circuit the *bit*, i.e. the quantum of information, is represented as a voltage saturated to the high or low level, 1 or 0, which makes small amplitude fluctuations not so relevant. On the other hand, for proper operation digital circuits need precise timing, within a fraction of the clock period.

Secondly, we consider a synthesizer that – by definition – transforms the driving frequency ν_i into an output frequency $\nu_o = \frac{n}{d}\,\nu_i$, where n and d are integers. Restricting our attention to an ideal synthesizer consisting of zero-delay noise-free circuits, if the driving signal is affected by a time jitter δt_i, the same time jitter is present at the output. Therefore, the synthesizer transforms the input phase fluctuations $\varphi_i(t)$ into fluctuations $\varphi_o(t) = \frac{n}{d}\,\varphi_i(t)$ of the output signal. So, any attempt to increase the frequency also increases phase noise. Even worse, if the rms value of $\varphi_o(t)$ exceeds some 1 rad, the carrier vanishes due to cycle loss and extra cycle insertion at random time.

Thirdly, we consider a feedback oscillator whose nominal frequency ν_0 is set by a resonator of merit factor Q. The oscillator loop gain must be equal to 1 for the oscillation to be stable, which means unity modulus and 0° phase. In practice the oscillation frequency is determined by the 0° phase condition only, while the unity gain condition results from saturation or from amplitude control. Thus, a *phase* perturbation φ_a present along the loop path is compensated by the resonator, whose phase changes by $\varphi_r = -\varphi_a$, which produces a *frequency* change $\delta\nu = \frac{\nu_0}{2Q}\varphi_a$. The output phase error, which is the integral of $\delta\nu$, may diverge. This description of the oscillator behaviour is known as the Leeson model [1].

Finally, get a look at the long range radar. The main lobe illuminates both the target and the ground. Obviously, a pulsed radar can not discriminate between the target and the ground clutter at the same range. But fortunately the echo from a moving target is frequency shifted by $\Delta\nu/\nu_c = 2v/c$, due to the Doppler effect; v is the range rate of the target and $c = 3{\times}10^8$ m/s is the speed of light. As a consequence, a Doppler radar can divide a moving target from the clutter, but this is possible only if the source linewidth is sufficiently narrow to allow dividing the received frequency from the transmitted one; in addition, the radar oscillator must keep its frequency constant from the emission to the reception instant. Both these features rely upon the low phase noise of the oscillator and its components [2,3].

2 Background

Phase noise is a random process, and consequently its power spectrum density $S_\varphi(f)$ can only be defined as the Fourier transform of the autocorrelation

function $\mathcal{R}_\varphi(\tau)$. The one-side spectrum is preferred because this is what spectrum analyzers display. Complying to the usual terminology, we use the symbol ν for the frequency and f for the Fourier frequency, i.e. the frequency of the detected signal when the sidebands around ν are down converted to baseband.

The power-law model is most frequently used for describing phase noise. It assumes that $S_\varphi(f)$ is equal to the sum of terms, each of which varies as an integer power of frequency. Thus each term, that corresponds to a noise process, is completely specified by two parameters, namely the exponent and the value at $f = 1$ Hz. Five power-law processes, listed below, are common in electronics.

noise type	$S_\varphi(f)$
white phase	$b_0 f^0$
flicker phase	$b_{-1} f^{-1}$
white frequency	$b_{-2} f^{-2}$
flicker frequency	$b_{-3} f^{-3}$
random-walk frequency	$b_{-4} f^{-4}$

All these noise types are generally present at the output of oscillators, while two-port devices show white phase and flicker phase noise only. For reference, Fig. 1 reports the typical phase noise of some oscillators and devices.

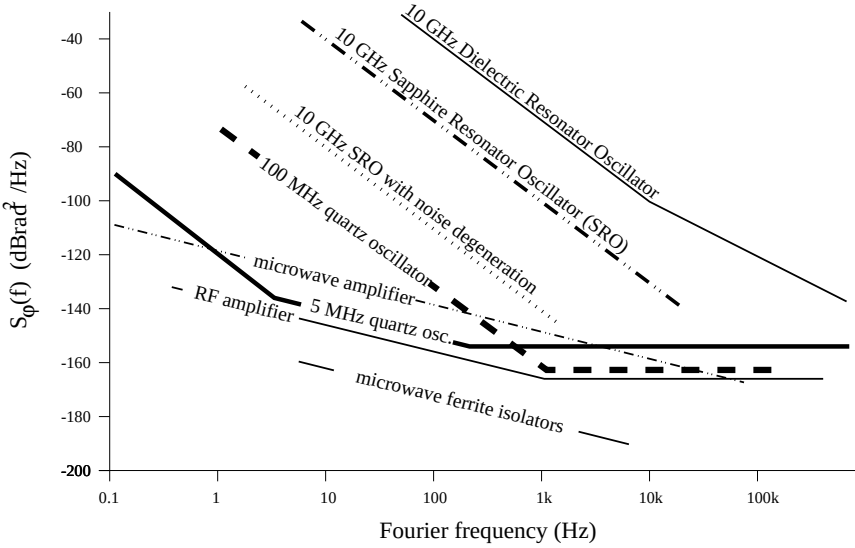

Fig. 1. Typical phase noise of some oscillators and devices

Phase noise can be measured by means of a phase-to-voltage converter in conjunction with a spectrum analyzer, which can be of the low frequency

type. On the other hand, it could be measured directly by inspecting around the carrier with a spectrum analyzer. The first method is related to the double sideband (DSB) representation of noise, while the second refers to the single sideband (SSB) representation.

2.1 Double sideband (DSB) representation of noise

Let us consider a noise process of spectrum density $N(\nu)$ symmetrical around the carrier frequency ν_c, which means that $N(\nu_c+f) = N(\nu_c-f)$. That noise process is regarded as a pair of sidebands responsible for phase and amplitude noise. In order to derive $S_\varphi(f)$ we consider two symmetrical noise slots of bandwidth B at $\pm f$ apart from ν_c, as shown in Fig. 2. The rms voltage of the carrier is $\sqrt{R_0 P_c}$. The two noise sidebands must be *in quadrature* to the carrier for only the phase to be perturbed. Assuming that the noise contributes equally to amplitude and phase, the rms voltage of the *quadrature* sidebands is $\sqrt{R_0 NB/2}$. These sidebands cause a phase modulation whose peak angle is $\varphi_p = \arctan\sqrt{2NB/P_c}$, as it results from the phasor representation of Fig. 2; the corresponding phase fluctuation is $\varphi_{\rm rms} = \arctan\sqrt{NB/P_c}$. Under the assumption of low noise-to-carrier ratio, that modulation angle becomes $\varphi_{\rm rms} = \sqrt{NB/P_c}$. Hence the spectrum density is

$$S_\varphi(f) \;=\; \frac{N(\nu_c + f)}{P_c}\;. \tag{2}$$

The physical dimension of $S_\varphi(f)$ is $\mathrm{rad}^2/\mathrm{Hz}$. In addition, the technical unit $\mathrm{dBrad}^2/\mathrm{Hz}$ is frequently used.

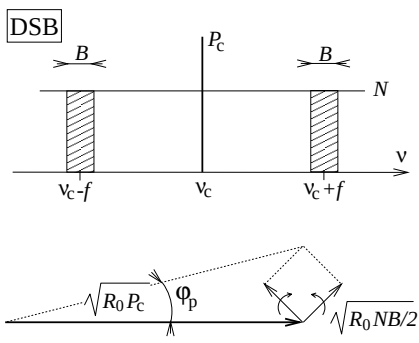

Fig. 2. Double sideband (DSB) representation of noise

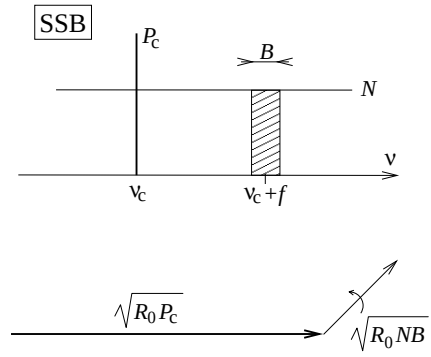

Fig. 3. Single sideband (SSB) representation of noise

In some relevant situations, the noise does *not* equally contribute to phase and amplitude. This occurs for instance with all digital circuits, in which the amplitude noise is nearly suppressed by saturation. In these cases, amplitude

and phase noise must divided and dealt with separately representing the spectrum density as $N(\nu) = N_\alpha(\nu) + N_\varphi(\nu)$. Hence, the rms voltage of the quadrature noise is $\sqrt{R_0 N_\varphi B}$ for each sideband, and consequently the phase fluctuation is $\varphi_{\rm rms} = \sqrt{2N_\varphi B/P_c}$. The phase noise spectrum density thereby obtained is

$$S_\varphi(f) \;=\; \frac{2N_\varphi(\nu_c + f)}{P_c}\;. \tag{3}$$

2.2 Single sideband (SSB) representation of noise

Let us now consider one sideband of the noise process $N(\nu)$ around the carrier, as shown in Fig. 3. Taking one slot of bandwidth B at the frequency f apart from the carrier, the corresponding rms voltage is $\sqrt{R_0 N B}$. The latter causes a phase modulation $\varphi_{\rm rms} = \sqrt{NB/2P_c}$, plus an amplitude modulation. The quantity used to describe the spectrum density thereby obtained is

$$\mathcal{L}(f) \;=\; \frac{N(\nu_c + f)}{2P_c}\;. \tag{4}$$

The physical dimension of $\mathcal{L}(f)$ is Hz^{-1}; the unit of angle (rad) should be omitted. $\mathcal{L}(f)$ is usually expressed in dBc/Hz, where "c" is intended to remind one that the $\mathcal{L}(f)$ results from the noise referred to the carrier power.

Assuming that noise contributes equally to phase and amplitude modulation, for small modulation angles it holds $\mathcal{L}(f) = \frac{1}{2}\, S_\varphi(f)$.

It should be noticed that the above derivation of (4) does not contain any explicit reference to phase, while phase noise comes from the equipartition of noise between the two degrees of freedom, i.e. phase and amplitude. Obviously, whenever the equipartition does not apply – as it occurrs with frequency multiplication – definition (4) yields a misleading result. In addition, nowadays $\mathcal{L}(f)$ is almost always measured by means of a phase to voltage converter, which is insensitive of amplitude. For this reason, the definition (4) is now being changed [4] into

$$\mathcal{L}(f) \;=\; \frac{1}{2}\, S_\varphi(f)\;. \tag{5}$$

Finally, $\mathcal{L}(f)$ is preferred to $S_\varphi(f)$ by most manufacturers; nevertheless, we use $S_\varphi(f)$ because we find it more clear.

3 Traditional Methods

The double balanced mixer (DBM), used as a phase-to-voltage converter as shown in Fig. 4, is the main tool for phase noise measurements. The mixer is driven by two signals in quadrature ($\gamma = 90°$) with nearly equal power

$$r(t) = \sqrt{2R_0 P_c}\cos\left[2\pi\nu_c t + \gamma\right] \tag{6a}$$

$$s(t) = \sqrt{2R_0 P_c}\;\left[1 + \alpha(t)\right]\cos\left[2\pi\nu_c t + \varphi(t)\right]\;. \tag{6b}$$

Basically, the mixer is a multiplier. Thus the signal at the output of the lowpass filter is $[r(t)s(t)] * h_{\rm lp}(t) = R_0 P_c[1 + \alpha(t)]\sin[\varphi(t)]$; the symbol '$*$' stands for the convolution operator and $h_{\rm lp}(t)$ is the low pass function that eliminates the $2\nu_c$ component of the IF signal. For proper operation, the mixer is saturated at both inputs. Hence the output signal is independent of the input power, and the amplitude noise $\alpha(t)$ vanishes. Linearizing $\sin(\varphi)$ for small φ, the output signal is

$$v(t) = \sqrt{K_\varphi}\,\varphi(t) \tag{7a}$$

$$S_v(f) = K_\varphi\, S_\varphi(f)\,, \tag{7b}$$

which also defines the power gain K_φ. For commodity we tend to use either K_φ or $k_\varphi = \sqrt{K_\varphi}$, depending on the measurement method; anyway, the numerical value in dB is the same.

Obviously, the measurement of $S_v(f)$ gives $S_\varphi(f)$. Often, the fast Fourier transform (FFT) analyzer is the most suitable instrument because of the wide dynamic range, typically of some 80–90 dB.

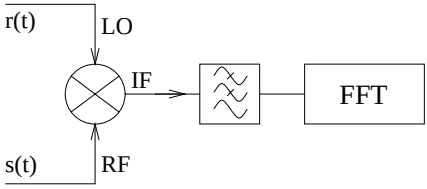

Fig. 4. Basic scheme of the traditional phase noise measurement system

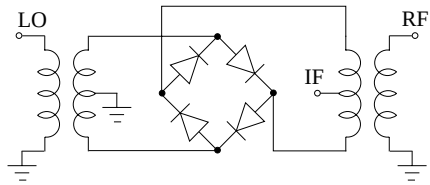

Fig. 5. Scheme of a typical radiofrequency double balanced mixer

The double balaced mixer can take various forms [5–7], depending on frequency and technology. Figure 5 shows the scheme most widely used in the HF to UHF bands (bands 7 to 9).

3.1 Instrument sensitivity

The white noise limit comes from the input noise of the amplifier inserted between the mixer and te spectrum analyzer (not shown in Fig. 4), In fact, commercial FFT spectrum analyzers show a typical input noise of the order of 20 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$, limited by the relatively high input impedance. On the other hand, the mixer output impedance is low, typically 50 Ω. Consequently, the sensitivity can be significantly improved by inserting an amplifier with lower input impedance between the mixer and the analyzer. The input noise $S_{v0}(f)$ of that amplifier can be of 1 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ (-180 dBV) or lower [8]. In this conditions, assuming that $K_\varphi = -10$ $\mathrm{dBV}^2/\mathrm{rad}^2$, a noise floor of -170 $\mathrm{dBrad}^2/\mathrm{Hz}$ can be attained. This high sensitivity refers to good or best conditions only.

It should be remarked that the best noise impedance $R_b = \sqrt{S_v(f)/S_i(f)}$ of the operational amplifiers is in the kΩ range; $S_v(f)$ and $S_i(f)$ refer to the spectral density of the voltage and current noise of the amplifier. For this reason, the operational amplifiers are misused when connected to a mixer that shows an output impedance as low as 50 Ω.

The flicker noise of the instrument comes from the mixer diodes, and it is insufficiently documented in the literature. According to our experience, the actual limit turns out to be of some -140 dBrad^2/Hz, depending on the mixer type and the driving power. The flicker of a low noise amplifier can be lower than 3 nV/$\sqrt{\text{Hz}}$ at $f = 1$ Hz, i.e. -170 dBV, which is negligible compared to the mixer noise.

Figure 6 shows the typical limit of the phase noise measurement system based on the double balanced mixer.

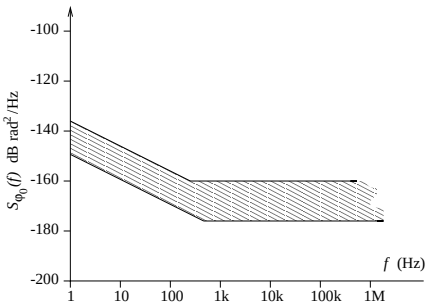

Fig. 6. Typical instrument noise of a mixer based phase noise measurement system

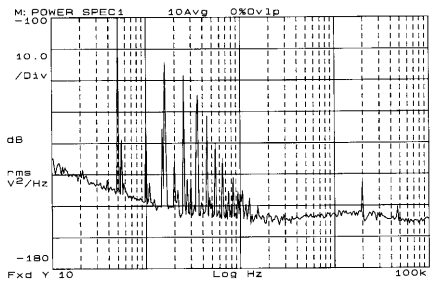

Fig. 7. Example of instrument noise $S_{\varphi 0}(f)$, measured in the absence of the device to be tested. Vertical scale is dBrad^2/Hz

3.2 Additional instrument limitations

Due to disturbances from the mains, the measured spectrum includes many lines at 50 Hz (or the appropriate frequency, out of the Europe) and multiples. These disturbances are picked up at the input of the operational amplifier, where the signal level is the lowest. Signal processing is only partially useful to clean up the measurement results because the spectrum analyzer shows a finite bandwidth, and consequently two contiguous harmonics of the mains may hide the useful information in between, thus limiting the sensitivity. Figure 7 provides a typical example of actual results.

The mixers designed for radiofrequency bands are based on ferrite tore transformers and Schottky diodes. They show a bandwidth of up to 3 decades in the region from 20 kHz to 2 GHz approximately. At higher frequencies, the transformer is replaced with a microstrip network, whereat the bandwidth

is limited to 1–3 octaves. The maximum operating frequency is of the order of 40 GHz, although some special devices may be usable up to 100 GHz, or even beyond.

The adoption of a saturated mixer as the phase-to-voltage converter turns into a severe limitation of the instrument power range. The conversion gain is roughly proportional to driving power. Accounting for the diode saturation level and maximum power, the useful dynamic range tends to be of some 10 dB. Mixers are hardly usable below approximately 5 dBm, while flicker noise increases as the power approaches the maximum value.

4 Useful Schemes

Figures 8 to 15 show some experimental configurations that cover most practical needs.

General two port devices. A two port device (DUT, i.e. the device under test) can be measured with the scheme of Fig. 8. This scheme is a direct application of the principles shown in Section 3. Both $r(t)$ and $s(t)$ originate from a single source, but only $s(t)$ is affected by the DUT noise. The quadrature condition is ensured by the variable phase shifter γ, that must be adjusted to compensate for the phase lag of the DUT and cables. The quadrature condition can be first checked observing the dc voltage at the mixer output, that must be 0 V. For highest accuracy, the dc offset that results from the diode asymmetry must be taken into account; a true phase measurement may be necessary instead of just trimming γ for 0 V_{dc} at the IF output.

The scheme of Fig. 8 works well if the DUT shows relatively low insertion loss because the mixer must be driven with the same level at the two inputs. In addition, the DUT group delay must be relatively small. This is necessary to ensure the rejection of the phase noise of the driving oscillator. Anyway, this problem will be explained underneath dealing with the discriminator.

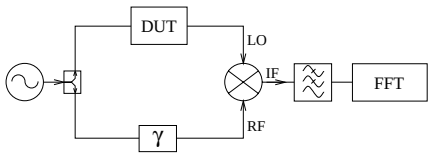

Fig. 8. Phase noise measurement of a 2 port device

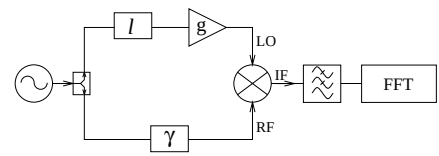

Fig. 9. Measuring an amplifier as the DUT, the amplifier must be preceeded by an attenuuator of loss ℓ equal to the gain g

Amplifiers. The phase noise of an amplifier can be measured with the scheme of Fig. 9. The amplifier gain g is compensated by an attenuator of loss $\ell \simeq g$. Obviously the attenuator and the amplifier can not be interchanged, unless the amplifier dynamics is sufficiently large to avoid saturation. Anyway, it should be noticed that the close-to-the-carrier flicker noise of the amplifier comes from the near-dc flicker of the bias current, up converted by nonliearity [9,10], and hence it strongly depends upon the output power. In addition, the presence of the attenuator increases by a factor ℓ the overall noise figure of the attenuator-amplifier compound, whereat the white noise floor of the measurement is also increased by ℓ.

High insertion-loss two port devices. The mixer requires nearly equal driving power at the two inputs. Accordingly, if the DUT shows a significant loss ℓ, an attenuator of equal loss must be inserted in the other arm, and the oscillator power must be set to an appropriately higher value. This solution requires that the increased power is tolerated by the DUT without noise increase or damnage. Alternatively, the DUT loss can be compensated inserting an amplifier of gain $g \simeq \ell$, as shown in Fig. 10. The amplifier is also needed to measure the DUTs that must work in low power conditions. Unfortunately the power required to drive the mixer is relatively high and sufficient to make the amplifier flicker, which impairs the instrument sensitivity. For this reason some low noise devices, like the high stability quartz resonators [11], can not be measured with this scheme.

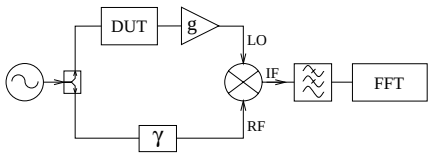

Fig. 10. In some cases an amplifier is needed to compensate for the DUT loss

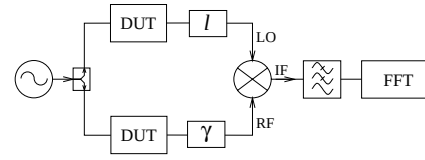

Fig. 11. Whenever possible, two equal DUTs should be measured simultaneously

Equal DUT pair. When two equal DUTs are available, they can be inserted each in one arm of the circuit, as shown in Fig. 11. To reduce the circuit complexity, the attenuation of the variable phase shifter γ is compensated by the variable attenuator ℓ, while the phase lag of the latter is compensated by γ. If the two devices are equally noisy there results a 3 dB improvement of sensitivity.

The presence of two equal DUTs is sometimes useful to improve the rejection of the source noise by compensating for the discriminator effect of each DUT.

The scheme of Fig. 11 turns out to be useful for the measurement of frequency multipliers, dividers and synthesizers in general. In fact, if two independent oscillators were used to measure a *single* sinthesizer – one oscillator drives the synthesizer and the other one serves as the reference of the mixer – it would be necessary to use two noise-free oscillators because no rejection of the fluctuations of these oscillators would take place.

Discriminator and delay line. A single oscillator can be measured with the scheme of Fig. 12, in which the resonator is used as a reference frequency discriminator. The resonator responds to a frequency change $y = \frac{\nu - \nu_c}{\nu_c}$ with a phase $\varphi_m = 2Qy$, where Q is the merit factor of the resonator. Accordingly, the spectrum density $S_{\varphi m}(f)$ of the measured phase is related to the frequency fluctuation $S_{yo}(f)$ of the oscillator by

$$\begin{aligned} S_{\varphi m}(f) &= 4Q^2 S_{yo}(f) \\ &= 4Q^2 \frac{f^2}{\nu_0^2} S_{\varphi o}(f) \; . \end{aligned} \tag{8}$$

The above hold at low Fourier frequencies ($f \ll \frac{\nu_c}{2Q}$), where the resonator phase lag can be derived from a quasistatic model. For $f > \frac{\nu_c}{2Q}$, the resonator filters out the frequency fluctuations of the oscillator. These fluctuations are still present at the other input of the mixer, and therefore the instrument measures the phase noise of the oscillator.

In most cases the discriminator shows poor sensitivity, due to the insufficient Q factor of the resonator. On the other hand, the poor sensitivity turns into a wide dynamic range.

A delay line does the same work as the resonator. The equivalent merit factor is $Q_e = \pi\tau\nu_c$, where τ is the delay of the line. Unfortunately, delay lines are scarcely useful in the frequency domain of our interest because of the insufficient value of the $\tau\nu_c$ product. Nevertheless, the delay line is of great interest for optics.

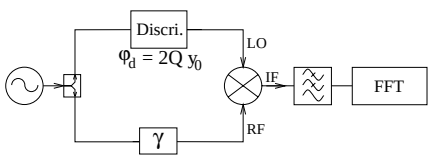

Fig. 12. A discriminator converts the frequency fluctuation of the driving oscillator into phase fluctuations that are measured by means of a mixer

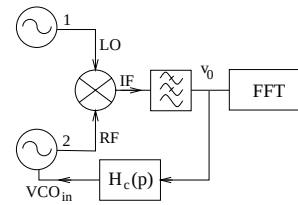

Fig. 13. Phase noise measurement of two equal oscillators

Oscillator pair The measurement of oscillators requires a phase locked loop (PLL) as in Fig. 13 – or some other locking mechanism – otherwise it would be impossible to keep the mixer inputs in quadrature. The PLL is usually regarded as a low pass filter, in which the voltage contolled oscillator (VCO) tracks the input. The corresponding transfer function is

$$\frac{S_{\varphi 2}(f)}{S_{\varphi 1}(f)} = \frac{|k_o k_\varphi H_c(f)|^2}{4\pi^2 f^2 + |k_o k_\varphi H_c(f)|^2} \,, \tag{9}$$

where k_o is the VCO gain, given in rad/sV. In order to measure phase noise, the error signal v is used as the PLL output signal, and therefore the transfer function is

$$\frac{S_v(f)}{S_{\varphi 1}(f)} = \frac{4\pi^2 f^2 k_\varphi^2}{4\pi^2 f^2 + |k_o k_\varphi H_c(f)|^2} \,, \tag{10}$$

which is a high pass function. Obviously, the loop response must be sufficiently slow for the oscillator no. 2 *not* to track the other one.

Alternatively, the scheme of Fig. 13 can be exploited to increase the dynamic range of the system. In this case, an amplifier is inserted as the H_c block, setting the loop gain to a suitable value that pushes the high pass cutoff frequecy just below the corner of the oscillator frequency flicker.

The frequency flicker is of the f^{-3} type, while the PLL response is proportional to f^2. The combined effect yields a measured spectrum of the f^{-1} type, from which the flicker coefficient can be calculated.

The PLL scheme of Fig. 13 can be used to compare an oscillator to a reference one, considered noise free, or to compare two equal oscillators. In this case, a 3 dB factor must taken into account.

Frequency multiplier. In some cases the frequency multiplier turns out to be a useful tool to enhance the instrument sensitivity (Fig. 14). In fact, if the frequency is multiplied by n, the phase also is multiplied by n. Hence the output power spectrum $S_{\varphi o}(f)$ is related to the input spectrum $S_{\varphi i}(f)$ by $S_{\varphi o}(f) = n^2 S_{\varphi i}(f)$. The obvious extension to the general case of the synthesizer is $\nu_o = \frac{n}{d}\,\nu_i$, which means $\varphi_o(t) = \frac{n}{d}\,\varphi_i(t)$ and conssequently $S_{\varphi_o}(f) = (\frac{n}{d})^2 S_{\varphi i}(f)$.

The multiplier contributes with its own noise, which must be lower than that of the mixer for the multiplier to be useful. In pratice this constraint turns into a serious difficulty for the white noise, but it is relatively easy to meet with the frequency flicker noise of the oscillator, whose slope is f^{-3}.

Narrow tuning range oscillators. With some high quality oscillators, the tuning range is significantly narrower than the initial frequency accuracy. This occurrs when the Q factor of the resonator is extremely high, 10^6 to 10^9, and for technical reasons the resonance frequency can not be changed

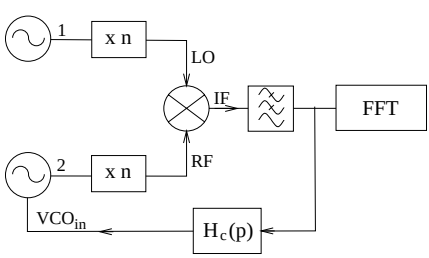

Fig. 14. Frequency multiplication enhances the instrument gain

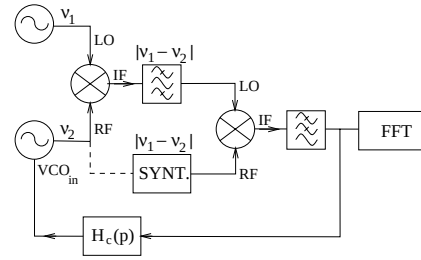

Fig. 15. Phase noise measurement of two oscillators *not* at the same frequency

with a resolution better than $\frac{\nu_c}{2Q}$. This is typical of the oscillators based on whispering gallery sapphire resonators or on cryogenic resonators. The noise of these oscillators is so low that no synthesizer would be adequate, and there is no chance to get two oscillators at the same frequency. Yet, the phase noise measurement is still possible by means of the scheme of Fig. 15. Choosing two oscillators whose frequencies are as close as possible to one another, the phase noise measurement is performed at the beat frequency $\nu_d = |\nu_1 - \nu_2|$, by comparison with an auxiliary synthesizer.

The configuration of Fig. 15 shows an additional advantage as it prevents the measurement error due to injection locking. In fact, when high Q resonators are used, the oscillators tend to lock to one another if the frequencies are sufficiently close.

5 Interferometric Noise Measurement Method

It has been shown in Sect. 3 that the saturated mixer used as a phase-to-voltage converter suffers from three basic problems, namely the narrow power range, the flicker noise of the mixer, and the relatively high white noise floor due to the poor phase-to-voltage conversion gain. An improved solution consists of the interferometric scheme, shown in Fig. 16. This scheme, inspired to [12], has been subsequently ameliorated and extended to the HF and VHF bands [13].

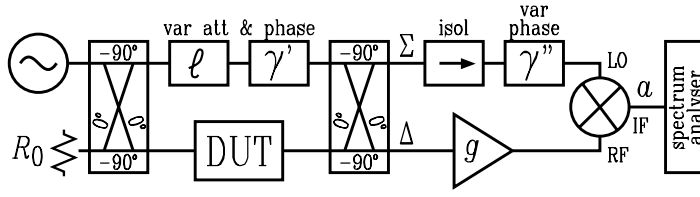

Fig. 16. Basic scheme of the radiofrequency interferometer

The left hybrid is used as a power splitter. It should be remarked that actual two way power splitters are 4 port hybrids with one port terminated to an internal resistor, otherways they could not be impedance matched; for details on these devices consult [14]. The right hybrid coupler makes the vector addition, i.e. the *interference* of its input signals. Thus, setting ℓ' and γ' so that the two paths from the oscillator to the Δ output of that hybrid are equal in amplitude and opposite in phase, all the oscillator power goes to the Σ output, while the carrier is suppressed at the Δ output. The DUT noise sidebands, which are not suppressed by the interference mechanism, are amplified by the low noise amplifier and downconverted to baseband by the mixer. After filtering out the $2\nu_c$ component, the instant voltage at the mixer output is

$$v(t) = k\left[\varphi(t)\sin\gamma'' + \alpha(t)\cos\gamma''\right] . \tag{11}$$

Thus the mixer detects PM noise, AM noise or a combination of them, depending on the detection phase γ''. Because of the particular type of conversion, the instrument gain k is written without the index φ.

Let us now introduce the following symbols

$N_\varphi(\nu)$	the power spectrum density (PSD) of the quadrature noise at the DUT output, around the carrier frequency ν_c,
$N_\alpha(\nu)$	the PSD of the in-phase noise at the DUT output, around the carrier frequency ν_c,
ℓ_h	the power loss of the hybrid, not including the 3 dB intrinsic loss. Therefore, driving one input with a power P, a power $P/2\ell_h$ is expected at each output,
ℓ_m	the power loss of the mixer, icluding the 3 dB intrinsic loss. Hence, driving the RF input with a power P, a power P/ℓ_m is expected in each output band,
R_0	the mixer output impedance,
g	the power gain of the amplifier.

A few words are to be spent about the contrast between the definition of ℓ_m, that includes the 3 dB intrinsic loss of the mixer, and the definition of ℓ_h, that does *not* include the 3 dB intrinsic loss of the hybrid. Apart from the fact that the physical origin of the intrinsic losses is not the same, the above definitions of ℓ_m and ℓ_h are consistent with the technical documentation of most commercial components.

From the scheme of Fig. 16, the noise PSD is

$$S_\Delta(\nu) = \frac{N_\varphi(\nu) + N_\alpha(\nu)}{2\,\ell_h} \tag{12}$$

at the Δ output of the hybrid, and

$$S_{\mathrm{RF}}(\nu) = g\,\frac{N_\varphi(\nu) + N_\alpha(\nu)}{2\,\ell_h} \tag{13}$$

at the RF input of the mixer.

The noise behaviour of actual DUTs is such that the sidebands are symmetrical with respect to the carrier, which means that $N_\varphi(\nu_c - f) = N_\varphi(\nu_c + f)$ and $N_\alpha(\nu_c - f) = N_\alpha(\nu_c + f)$. Hence, the DUT noise down converted to baseband is

$$S_V(f) = \frac{2R_0 g}{\ell_h \ell_m} \left[N_\varphi(\nu_c + f) \sin^2 \gamma'' + N_\alpha(\nu_c + f) \cos^2 \gamma'' \right] . \tag{14}$$

Still assuming the sideband symmetry, the PM noise is related to N_φ by

$$S_\varphi(f) = 2 \frac{N_\varphi(\nu_c + f)}{P_c} . \tag{15}$$

If γ'' is set to 90°, combining the two above equations with the instrument gain $K_\varphi = S_V(f)/S_\varphi(f)$, we obtain

$$K_\varphi = \frac{R_0 g P_c}{\ell_h \ell_m} . \tag{16}$$

The expected noise floor of the instrument can be easily derived as follows. We first remove the DUT, replacing it with a cable; obviously, ℓ' and γ' must be set to 0 dB and 0° respectively. The carrier suppression is still effective, and the whole circuit is still impedance matched. As a consequence, the thermal noise present at the input of the amplifier comes from the termination R_0 of left side hybrid.

Accounting for the amplifier noise figure F, the equivalent noise density at input of the amplifier is $S_{\Delta 0} = F_a k_B T_0$, where $k_B = 1.38 \times 10^{-23}$ W/Hz is the Boltzmann constant, and $T_0 = 290$ K is the absolute reference temperature. It is assumed that the temperature of the interferometer is close to T_0. Thus, the voltage noise at the mixer output is

$$S_{V0}(f) = 2 \frac{R_0 g}{\ell_m} F k_B T_0 . \tag{17}$$

Assuming that the carrier is perfectly suppressed at the input of the amplifier, the amplifier noise $F k_B T_0$ can *not* be related with the phase of the carrier. Accordingly, $F k_B T_0$ gives equal contributions to phase and amplitude noise. Hence, combining (16) with (17) we obtain the phase noise floor

$$S_{\varphi 0}(f) = 2\ell_h \frac{F k_B T_0}{P_c} . \tag{18}$$

Quite a similar development yields the AM noise floor

$$S_{\alpha 0}(f) = 2\ell_h \frac{F k_B T_0}{P_c} . \tag{19}$$

The practical consequences of the above theory can be better understood through the following example.

Design example 1. We assume that $\ell_h = 0.5$ dB and $\ell_m = 6$ dB, which are typical values for hybrids and mixers. We choose a $g = 37$ dB amplifier and we set the driving power for $P_c = 15$ dBm at the DUT output. From equation (16) we get $K_\varphi = 32$ dBV^2/rad^2. Assuming that the noise figure of the amplifier is $F = 2$ dB, equation (18) yields $S_{\varphi 0} = -185$ dB rad^2/Hz. These results compare favorably to the performance of a traditional system based on a saturated mixer that operates in similar conditions. The interferometer shows a gain 42 dB higher and a noise floor 15 dB lower.

5.1 Design strategies

Whereas the potential benefit of high carrier suppression is clear, a suppression specification can hardly be drawn. This occurs because the interferometer takes benefit of some features of the electronic components that are not adequately documented. Therefore, we can only give some hints derived from experience.

1. The amplifier gain g should be in the 20–40 dB range. Higher values make a sufficient carrier suppression difficult to achieve, while lower values cause the mixer noise to be taken in, impairing the sensitivity.
2. The residual carrier power at the amplifier output must be much lower than the maximum amplifier power P_m; the latter is usually specified as the "1 dB compression level". A margin $P_m/(gP_r)$ of 35–40 dB or more is needed, depending on the amplifier; P_r is the residual carrier power at the amplifier input.
3. The close-to-the-carrier flicker of the amplifier results from the combined effect of near-dc flicker and nonlinearity. Although we have no information on the former, we can infer the latter from the 3rd harmonic intercept power, which is always specified for commercially available amplifiers. Obviously, the devices showing the highest intermodulation intercept power tend to be the best ones.
4. For a given configuration, flicker noise of the amplifier is entierly determined by the power available at the output of the amplifier, that is mostly due to the residual carrier. As a consequence, the suppression ratio P_c/P_r should be regarded just as a way to specify P_r, rather than a parameter relevant by itself.
5. The presence of a low noise amplifier precedig the mixer relaxes the noise specifications for the latter.

Let us consider two examples.

Design example 2. The amplifier shows $g = 40$ dB, $P_m = 15$ dBm and needs a power margin $P_m/(gP_r) = 35$ dB for full linearity. In this condition, P_r must be less than –60 dBm. Consequently, if the DUT output power is $P_o = 15$ dBm than a carrier suppression of 75 dB must be ensured.

Design example 3. With the same instrument of the example 2, we now measure the phase noise of two piezoelectric quartz resonators, each dissipating $P_d = 10$ μW. For this purpose we insert two equal DUTs, one in each arm of the interferometer. Each DUT is a text fixture consisting of a resistive matching network that incorporates one quartz. The DUT output power is of the same order of P_d, depending on the test fixture network. Hence, assuming $P_c = -20$ dBm, a carrier suppression of 40 dB would be sufficient.

For referece, a carrier suppression of 80 dB results from an error $\delta\gamma' = 100$ μrad of the phase shifter, *or* from an error $\delta\ell = 8.7{\times}10^{-4}$ dB of the variable attenuator. Accounting for both, the accuracy specification is even more stringent.

Microwave design. Phase matching is the greatest technical difficulty at microwave frequencies. In fact, because the wavelength inside cables is about 22 mm at 10 GHz, a phase matching within 100 μrad – that is necessary for a carrier suppression of for 80 dB – is equivalent to an electrical length matching within 0.4 μm. Obviously, phase matching must be stable at that level for the duration of the experiment, say half an hour. Some commercially available phase shifters are adequate to do so, after a really patient adjustment.

A low instrument noise requires particular care with mechanical vibrations. In fact, at 10 GHz a noise floor of -180 dBrad^2/Hz corresponds to an electrical length fluctuation of $4{\times}10^{-12}$ m/$\sqrt{\text{Hz}}$. According to our experience, a sufficient stability can be obtained by fixing all the parts onto an antivibrating table – of the same type of those commonly used for optics – and securing to the table all the cables connecting the system to the external world.

Microwave hybrids and mixers show poor isolation, typically of the order of 20 dB. The obvious consequence is an unwanted feedback of the amplified signal through the mixer and the hybrid. In order to prevent oscillation or measurement alteration, isolation must be increased by inserting some ferrite isolators; the best configuration must be determined experimentally.

Microwave amplifiers show a wide bandwidth, in some cases more than 10 GHz. Noise integrated over such a wide band can push the amplifier out of linearity. If, for example, the amplifier shows a noise figure $F = 2$ dB, a gain $g = 40$ dB, and a bandwidth $B = 10$ GHz, the total integrated noise is $P_n = Fk_BT_0gB = -32$ dBm. Unfortunately, saturation is due to the peak power, which is some 20 dB higher than P_n. A bandpass filter can be needed.

VHF and HF design For technical reasons, the 5–10 MHz quartz oscillator exhibits the lowest frequency flicker, compared to similar devices at other frequencies. Besides, the quartz oscillator exhibits the lowest white noise floor

when crystal resonates around 100 MHz. Therefore, a great effort is wortly to be spent to characterize the electronic devices designed for these frequencies.

Phase matching, phase stability, and a sufficient damping of mechanical vibrations are much easier to achieve than in the microwave bands because the wavelength is 10^2–10^3 times longer. The adoption of semirigid cables, SMA type connectors and antivibrating table ensures sufficient stability. Working at $\nu_c = 10$ MHz, a 120 mm thick sand layer proved to be sufficient to damp the vibrations of the floor if some care were spent to do the measurements at certain hours, when only a few people were present in the laboratory.

In spite of the apparent simplicity, for a series of reasons the design for the HF and VHF bands turns out to be more difficult than that for the X band.

The most difficult problem arises from the variable phase shifters. Some microwave devices, consisting of a transmission line whose length can be varied by means of a micrometer, proved to be a bad choice. Apart from the small delay range (0.1–1 ns), that can be extended with a set of calibrated cables, these phase shifters turned out to be scarcely useful because of their high flicker noise; the same devices work successfully in the microwave range. We guess that this anomalous behaviour could be due to the parasitic capacitance in parallel with non-perfect contacts, which behaves as short circuit at 10 GHz and takes in acoustic noise when used at 100 MHz. Presently, a type of phase shifter specific for this application, based on a LC network with a variable capacitance, is the best known solution.

Variable attenuators suitable to the VHF band are generally based on potentiometers and for this reason they tend to flicker more than the microwave ones, based on movable absorbing surfaces. We are still searching for a more satisfactory solution, consisting either of better potentiometers or a different physical principle.

It should be remarked that the flicker performance of variable attenuators and phase shifters is usually not documented in the device specifications, and consequently the possibility to find low noise devices relies upon experience and a pinch of good luck.

Eliminating the harmonics at frequencies multiple of ν_c is a critical point because the carrier suppression mechanism has no effect on them. As almost all the components show a bandwidth of 2–3 decades, these unwanted signals would be present in the entire circuit, pushing the amplifier out of linearity and making it flicker. The only known solution consists of inserting low Q bandpass filters in certain points of the circuit.

Ferrite isolators are not available for the HF and VHF bands and must be replaced with active isolators. Although noise is not critical at the Σ output of the hybrid, where the isolators are to be placed, it is really important to drive both the active isolator and the mixer at the appropriate power level.

The presence of electromagnetic pollution can be a relevant problem at some frequencies. In fact, in highly populated areas of Europe and the U.S.A.

– well covered by FM broadcastings – the electromagnetic field is often of the order of +100 dBμV/m in the 88–108 MHz band. Besides, the Ethernet – that is probably the most popular standard for local area network – operates either at 10 Mb/s or 100 Mb/s and for this reason it turns into a source of pollution particularly difficult to eliminate from some experiments. Solutions are strongly dependent on the local situation, as well as the design.

Finally, the design for the measurement of quartz crystal resonators, that shows high Q factor and operate at low power, is subject to specific design rules [15,16].

5.2 Further remarks

The scheme of Fig. 16 can be improved by deriving the mixer LO signal from the oscillator instead of taking it at the Σ output of the hybrid. This makes the mixer pump level independent of the DUT power.

Working in the microwave bands, the 90° hybrids turn out to be the best choice. They are cheaper, smaller, and show better isolation and insertion loss than the 180° ones. The bandwidth of a 90° hybrid can be of 1–3 octaves, depending on the design and on the device size. The HF and VHF hybrids are based on lumped parameter networks. Hence, in these bands the 180° hybrids are superior to the 90° ones with respect to loss, isolation, size and cost. In addition, the 180° hybrids show a typical bandwidth of 2 decades, while the bandwidth of the 90° devices is of the order of half an octave.

In our experience the low frequency magnetic fields originated from the mains turn into a serious design problem because copper shields are not effective at these frequencies. The traditional systems, of the type described in Sections 3 and 4, are prone to that kind of disturbances because the signal, i.e. the DUT noise, is first down converted and then amplified; therefore the smallest signal, that is present at the mixer output, is a baseband one. Magnetic shielding can be used, but this solution makes the whole instrument more complicated. By contrast, the interferometer can be more effectively shielded. This occurs because the smallest signal, that consists of noise sidebands around the carrier, is amplified *before* being down converted. Moreover, low frequency magnetic fields have no effect on the high frequency noise sidebands. Figure 17, taken from [13], shows an example of instrument noise of an interferometer operating at $\nu_c = 9.1$ GHz with a carrier power $P_c = 15$ dBm. The residual of the mains can hardly be distinguished from the instrument noise.

Finally, it should be remarked that the interferometer provides the instant value of $\varphi(t)$ in real time, which makes it suitable to the dynamical removal of noise by means of a voltage controlled phase shifter in closed loop [17].

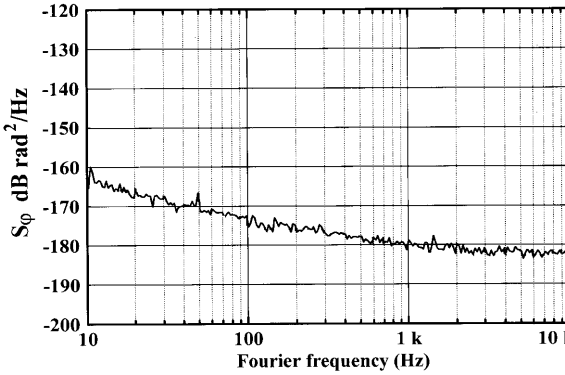

Fig. 17. Noise floor of a microwave interferometer prototype operating at $\nu_c = 9.1$ GHz. The DUT output power is 15 dBm

6 Correlation Techniques

The noise limitation of the traditional saturated mixer can be partially overcome by exploiting a correlation technique, in which two equal mixers measure the same noise process at the same time, as shown in Fig. 18. This type of measurement extracts the shared-path noise and rejects the single-arm noise processes, provided that they are independent. Obvously, the scheme of Fig. 18 can be easyly modified to measure oscillators, amplifiers, synthesizers etc., taking example from the single mixer schemes described in Section 4 (Figures 8 to 15).

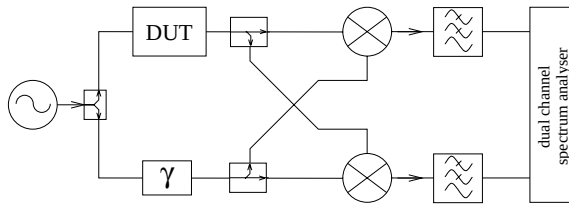

Fig. 18. Basic dual mixer correlation scheme

Searching back through the bibliography, the correlation scheme was probably used for the first time in the early '60s to measure the phase noise of hydrogen masers [18]. The dual channel correlation spectrum analyzers were not available at that time, and the correlation was evaluated through a calorimetric method. Since there, similar schemes were reproposed with updated technologies; see, for example, [19,20].

The (cross) correlation function $\mathcal{R}_{ab}(\tau)$ of two random voltages $a(t)$ and $b(t)$ is defined as

$$\mathcal{R}_{ab}(\tau) = \lim_{\theta\to\infty} \frac{1}{\theta} \int_{\theta} a(t)\, b^*(t-\tau)\, \mathrm{d}t \;, \tag{20}$$

where the symbol "$*$" stands for complex conjugate and can be omitted because we deal with real signals. The Fourier transform of $\mathcal{R}_{ab}(\tau)$ is the cross spectrum density

$$S_{ab}(f) = \int_{\infty} \mathcal{R}_{ab}(\tau) \exp(-2\pi f\tau)\, \mathrm{d}\tau \;. \tag{21}$$

Dynamic signal analyzers usually evaluate the cross spectrum density through the property

$$S_{ab}(f) = \mathcal{F}\{a(t)\}\, \mathcal{F}\{b(t)\} \;, \tag{22}$$

that holds for real signals; $\mathcal{F}\{.\}$ is the Fourier transform operator. The mathematics used in this section is clearly preseted in [21].

The signals $a(t)$ and $b(t)$ are proportional to the instant phase of the DUT, plus a random component due to the single-arm noise. Averaging on m measures, a rejection of the single-arm noise spectrum density by a factor $2\sqrt{m}$ is expected. The ultimate noise limit of the dual mixer method is not be discussed here. The noise theory of the double interferometer, that is a more sensitive instrument based on the correlation, is be given instead.

6.1 Double interferometer

An improved version of the correlation scheme, first proposed in [22], makes use of two equal interferometers that simultaneously measure the phase noise of a shared device, as shown in Fig. 19. The spectrum analyzer rejects the noise of the individual interferometers, under the assumption that the corresponding processes are independent.

The two low noise amplifiers of Fig. 19 are impedance matched, which implies that thermal noise is present at their input. In the absence of the DUT – the latter is replaced with a short cable – all the thermal noise comes from the terminations R_1, R_2 and R_3 of the hybrids used as power splitters. Because the noise coming from R_1 R_2 and R_3 is shared by the two amplifiers, at first sight one could believe that the thermal noise limits the instrument sensitivity. The full explanation, derived from [23] and explained underneath, is much more complex.

6.2 Noise theory of the double interferometer

In order to derive the noise theory of the double interferometer we analyze the case in which an attenuator of loss ℓ is inserted as the DUT. Then we define six

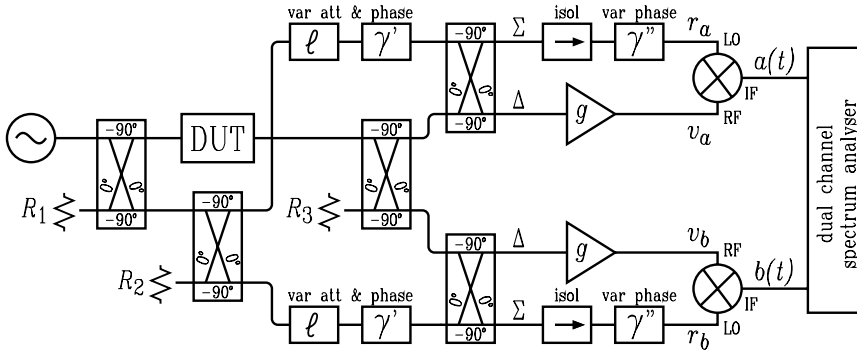

Fig. 19. Scheme of the double interferometer

thermal noise processes indicated with $n_1(t)$, $n_2(t)$, ... , $n_6(t)$, each of which coming from a resistor of value R_0, i.e. the characteristic impedance of the whole circuit. n_1, n_2 and n_3 are the available noise voltages across the resistors R_1, R_2 and R_3, respectively. n_4 and n_5 are the equivalent thermal noise of the resistive network of the variable attenuators, while n_6 is the equivalent thermal noise due to the resistive loss of the DUT. The power spectrum density $N_i(f)$ of each of these processes is equal to $R_0 k_B T_0$. The temperature of the whole instrument is assumed to be close to T_0, uniform and constant. As the DUT is an attenuator, it attenuates by ℓ any signal present at its input, including the thermal noise. As a consequence, only a fraction $\sqrt{(\ell-1)/\ell}\, n_6$. of the DUT noise n_6 is present at the output. This occurs because the total output noise at the attenuator output must be $R_0 k_B T_0$ when the attenuator input is terminated to a resistor. Similarly, the noise contributions of the two variable attenuators are $\sqrt{(\ell-1)/\ell}\, n_4$ and $\sqrt{(\ell-1)/\ell}\, n_5$. In addition, we assume that the DUT adds *extra noise* $\check{n}_6$ that can be of any type, including flicker. We use the word "extra" deliberately avoiding "excess" because the latter tend to be used as a synonym of flicker. Under the above hypotheses, the DUT output noise is

$$n_{\mathrm{dut}}(t) = \sqrt{\frac{\ell-1}{\ell}}\, n_6(t) + \check{n}_6(t)\ . \tag{23}$$

In the vicinity of the carrier frequency, a noise process can be divided in in-phase and quadrature components as

$$n(t) = n_x(t)\cos(\omega t) - n_y(t)\sin(\omega t)\ . \tag{24}$$

As we deal with thermal noise, the PSDs of the *baseband* noise processes $n_x(t)$ and $n_y(t)$ are $N_x(f) = R_0 k_B T_0$ and $N_y(f) = R_0 k_B T_0$, so that the PSD of the radiofrequency process $n(t)$ is $N(f) = R_0 k_B T_0$. Taking the oscillator signal $V\cos(\omega t)$ as the phase reference, the DUT output signal is $\sqrt{2R_0 P_c}\sin(\omega t)$. This means that the x noise component is responsible for phase noise; accordingly, $S_\varphi(f) = N_x(f)/R_0 P_c$.

For the sake of simplicity, we neglect the loss ℓ_h of the hybrids and power splitters, and the amplifier noise. The former can be introduced subsequently, and the latter vanishes in the correlation function because the two amplifiers are independent.

The reference signals at the mixer LO ports are of the form

$$r_a(t) = -V_p \cos(\omega t) \tag{25a}$$

$$r_b(t) = V_p \sin(\omega t)\ . \tag{25b}$$

Consequently, arm a detects the $\cos(\omega t)$ signal, while arm b detects the $\sin(\omega t)$ component. Due to the circuit phase relationships, the signals at the mixer RF inputs are

$$v_a(t) = \sqrt{g}\ \left[-\frac{1}{\sqrt{2\ell}}\, n_{1x} + \frac{1}{2\sqrt{\ell}}\, n_{2x} + \frac{1}{2}\, n_{3y} + \sqrt{\frac{\ell-1}{2\ell}}\, n_{4x} + \right.$$
$$\left. - \frac{1}{2}\sqrt{\frac{\ell-1}{\ell}}\, n_{6x} - \frac{1}{2}\, \check{n}_{6x} \right] \cos(\omega t) + \sqrt{g}\ \left[\begin{array}{c} \text{non} \\ \text{detected} \\ \text{terms} \end{array} \right] \sin(\omega t) \tag{26a}$$

$$v_b(t) = \sqrt{g}\ \left[\frac{1}{\sqrt{2\ell}}\, n_{1x} + \frac{1}{2\sqrt{\ell}}\, n_{2x} + \frac{1}{2}\, n_{3y} - \sqrt{\frac{\ell-1}{2\ell}}\, n_{5x} + \right.$$
$$\left. + \frac{1}{2}\sqrt{\frac{\ell-1}{\ell}}\, n_{6x} + \frac{1}{2}\, \check{n}_{6x} \right] \sin(\omega t) + \sqrt{g}\ \left[\begin{array}{c} \text{non} \\ \text{detected} \\ \text{terms} \end{array} \right] \cos(\omega t)\ . \tag{26b}$$

After filtering out the 2ω components, the detected signals present at the IF output of the mixers are

$$a(t) = \sqrt{\frac{2g}{\ell_m}}\ \left[\frac{1}{\sqrt{2\ell}}\, n_{1x} - \frac{1}{2\sqrt{\ell}}\, n_{2x} - \frac{1}{2}\, n_{3y} + \right.$$
$$\left. - \sqrt{\frac{\ell-1}{2\ell}}\, n_{4x} + \frac{1}{2}\sqrt{\frac{\ell-1}{\ell}}\, n_{6x} + \frac{1}{2}\, \check{n}_{6x} \right] \tag{27a}$$

$$b(t) = \sqrt{\frac{2g}{\ell_m}}\ \left[\frac{1}{\sqrt{2\ell}}\, n_{1x} + \frac{1}{2\sqrt{\ell}}\, n_{2x} + \frac{1}{2}\, n_{3y} + \right.$$
$$\left. - \sqrt{\frac{\ell-1}{2\ell}}\, n_{5y} + \frac{1}{2}\sqrt{\frac{\ell-1}{\ell}}\, n_{6x} + \frac{1}{2}\, \check{n}_{6x} \right]\ . \tag{27b}$$

Substituting the expression (27a) and (27a) in the definition (20), all the cross terms vanish. Hence the cross spectrum density is

$$S_{ab}(f) = \frac{g}{\ell_m}\ \left[\frac{1}{\ell}\, N_{1x} - \frac{1}{2\ell}\, N_{2x} - \frac{1}{2}\, N_{3y} + \frac{\ell-1}{2\ell}\, N_{6x} + \frac{1}{2}\, \check{N}_{6x} \right]\ . \tag{28}$$

Under the hypothesis of temperature uniformity, it holds $N_i = R_0 k_B T_0$ for all $i = 1 \ldots 6$. Consequently most of the terms of (28) cancel with one another

and there results

$$S_{ab}(f) = \frac{g}{2\ell_m}\,\check{N}_{6x}(f)\ . \tag{29}$$

The above equation states that the instrument compensates for the thermal noise, and therefore $\check{N}_{6x}(f)$ only contributes to the measured phase noise.

Finally, the gain of the double interferometer is

$$K_\varphi = \frac{gR_0P_c}{2\ell_m}\ , \tag{30}$$

which is half that of the single interferometer. This is is reasonable because in this case only half of the DUT noise is processed by each interferometer. For the same reason, the single arm noise floor is 3 dB higher than that of the single interferometer.

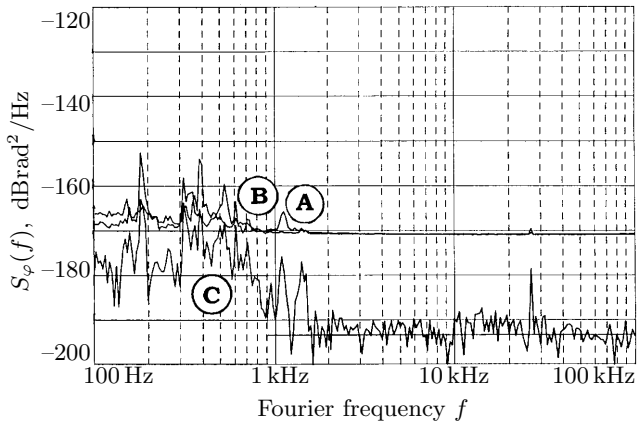

Fig. 20. Noise floor of the 100 MHz double interferometer prototype. A and B: single-arm. C: correlation

6.3 Noise properties of the double interferometer

The double interferometer shows some relevant noise properties that, although quite innatural at first sight, are well predicted by the general theory.

Noise floor. The first consequence of the thermal noise compensation mechanism is that the noise floor of the double interferometer can be lower than the thermally originated phase noise $S_{\varphi\,\mathrm{th}} = k_BT_0/P_c$.

Figure 20 shows an example of noise floor averaged on $m = 32767$ measures. This floor refers to a prototype operating at the carrier frequency $\nu_c = 100$ MHz, and fully described in [24]. In this prototype, the amplifiers

show gain $g = 40$ dB and noise figure $F = 2$ dB. The signal power at the mixer LO inputs is 8 dBm, and the DUT power is $P_c = 8$ dBm. The hybrids show losses $\ell_h = 0.8$ dB, while the mixer loss is $\ell_m = 6$ dB. Obviously, the DUT is replaced with a cable.

The single-arm noise floor (curves A and B of Fig. 20) is -172 dBrad^2/Hz, which is close to the expected value $S_{\varphi\,a}(f) = S_{\varphi\,b}(f) = 4Fk_BT_0\ell_h^2/P_c \simeq -172.3$ dBrad^2/Hz. The thermal noise calculated for the same conditions is $S_{\varphi\,\mathrm{th}} = k_BT_0/P_c \simeq -182$ dBrad^2/Hz. Yet, the measured floor (curve C) is $S_{\varphi\,0} \simeq -194$ dBrad^2/Hz, which is 12 dB lower than $S_{\varphi\,\mathrm{th}}$.

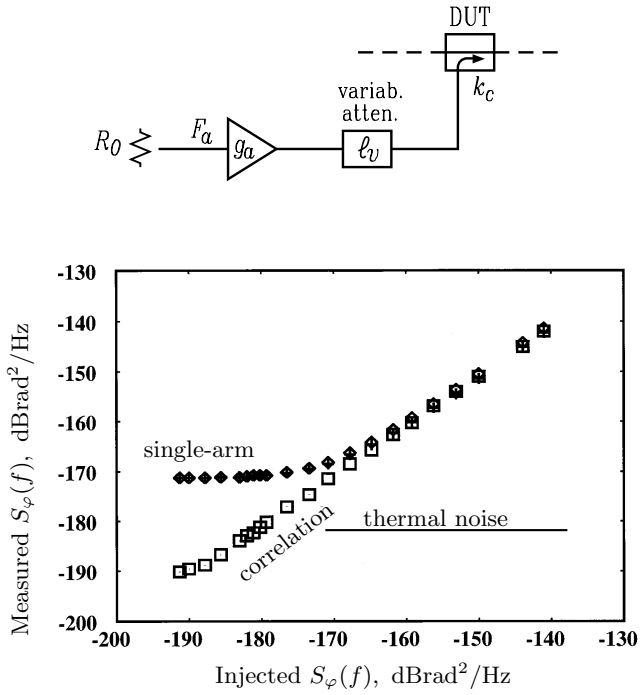

Fig. 21. Measurement of a reference noise $\check{N}_6$ injected along the DUT path

Noise measurement below the thermal floor. The possibility of measuring *extra* noise below the thermal floor can be experimented by injecting noise in the DUT path through a directional coupler, as shown in Fig. 21 (top). Neglecting the thermal noise N_6 because it is expected to be rubbed out, this circuit injects calibrated noise $\check{N}_6(f) = g_aR_0F_ak_BT_0/(\ell_vk_c)$. $\check{N}_6$ can be set to the desired value adjusting ℓ_v. The equivalent phase noise thereby injected is $S_{\varphi\,i}(f) = g_aR_0F_ak_BT_0/(\ell_vk_cP_c)$. Fig. 21 (bottom) shows the measured S_φ as a function of the injected $S_{\varphi\,i}$. Going to the left of that figure, ℓ_v

increases and the injected noise becomes negligible compared to the equivalent noise at the amplifier inputs. Therefore, the single-arm S_φ approaches the value of -172 dBrad^2/Hz previously measured in the absence of the DUT. By contrast, the correlated noise fits the straight line $S_\varphi = S_{\varphi\, i}$ even below $S_{\varphi\,\mathrm{th}} = k_B T_0 / P_c$.

Noise of an attenuator. The same noise mechanism responsible for the compensation of the shared resistor noise (R_1, R_2 and R_3) is also effective on the noise of an attenuator inserted along the DUT path. Figure 22 shows two extreme situations, in which the attenuator noise is generated separately in each arm of the double interferometer (left), or in the shared path (right). Experiments performed with 16 dB attenuators show that the instrument noise floor is the same for both the configurations, that it is lower than the thermal noise, and that it is limited by the averaging size m only [23].

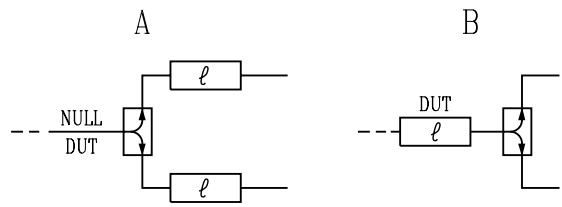

Fig. 22. Measurement schemes with independent attenuators in each arm, and with a shared attenuator. The attenuation $\ell = 16$ dB was usen in the experiments

References

1. D. B. Leeson, "A simple model of feed back oscillator noise spectrum", *IEEE Proceedings* vol. 54 no. 2 pp. 329–330, February 1966.
2. D. B. Leeson, G. F. Johnson, "Short-term stability for a Doppler radar: requirements, measurements and techniques", Proc. *IEEE-NASA Symposium on Short Term Frequency Stability* pp. 3–9, Greenbelt (MD, USA), 23–24 November 1964.
3. S. J. Goldman, *Phase noise analysis in radar systems*, John Wiley 1989, ISBN 0-471-61894-2.
4. E. S. Ferre-Pickal, J. R. Vig, J. C. Camparo, L. S. Cutler, L. Maleki, W. J. Riley, S. R. Stein, C. Thomas, F. L. Walls and J. D. White, "Draft revision of IEEE STD 1139-1988 standard definitions of physical quantities for fundamental frequency and time metrology - random instabilities", Proc. *51st Frequency Control Symposium* pp. 338–357, Orlando (FL, USA), 28–30 May 1997.
5. S. A. Maas, *Microwave mixers*, Artech House 1993, ISBN 0-89006-605-1.
6. S. A. Maas, *Nonlinear Microwave Circuits*, Artech House 1998, ISBN 0-7803-3403-5.

7. E. L. Kollberg (editor), *Microwave and millimeter-wave Mixers*, IEEE Press 1984, ISBN 0-87942-179-7
8. S. Franco, *Design with operational amplifiers and analog integrated circuits (2nd ed)*, McGraw Hill 1998, ISBN 0-07-115722-0.
9. V. N. Kuleshov, T. I. Boldyreva, "$1/f$ AM and PM noise in bipolar transistor amplifiers: sources, ways of influence, techniques of reduction", Proc. *51st Frequency Control Symposium* pp. 446–455, Orlando (FL, USA), 28–30 May 1997.
10. F. L. Walls, E. S. Ferre-Pikal, S. R. Jefferts, "Origin of $1/f$ PM and AM noise in bipolar junction transistor amplifiers", *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control* vol. 44 no. 2 pp. 326–334, March 1997.
11. J. J. Gagnepain, "Fundamental noise studies of quartz crystal resonators", Proc. *30th Frequency Control Symposium* pp. 84–91, Atlantic City (USA), 2–4 Jun 1976.
12. K. H. Sann, "The measurement of near-carrier noise in microwave amplifiers", *IEEE Transactions on Microwave Theory and Techniques* vol. 16 no. 9 pp. 761–766, September 1968.
13. E. Rubiola, V. Giordano, J. Groslambert, "Very high frequency and microwave interferometric PM and AM noise measurements", *Review of Scientific Instruments* vol. 70 no. 1 pp. 220–225, January 1999. ISSN 0034-6748.
14. R. Mongia, I. J. Bahl, P. Bhartia, *RF and microwave coupled-line circuits*, Artech House 1999, ISBN 0-89006-830-5.
15. J. Groslambert, V. Giordano, M. Brunet, E. Rubiola, "Flicker noise measurement of HF quartz resonators", Proc. *13th European Frequency and Time Forum / 1999 Frequency Control Symposium*, Besançon (France), 12–16 April 1999. To be printed.
16. E. Rubiola, J. Groslambert, M. Brunet, V. Giordano, "Flicker noise measurement of HF quartz resonators", to be published in the *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, 2000.
17. E. N. Ivanov, M. E. Tobar, R. A. Woode, "Microwave interferometry: applications to precision measurements and noise reduction techniques", *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control* vol. 45 no. 6 pp. 1526–1536, November 1998.
18. R. F. C. Vessot, R. F. Mueller, J. Vanier, "A cross-correlation technique for measuring the short-term properties of stable oscillators", Proc. *IEEE-NASA Symposium on Short Term Frequency Stability* pp. 111–118, Greenbelt (MD, USA), 23–24 November 1964.
19. F. L. Walls, S. R. Stain, J. E. Gray, D. J. Glaze, "Design considerations in state-of-the-art signal processing and phase noise measurement systems", Proc. *30th Frequency Control Symposium* pp. 269–274, Atlantic City (NJ, USA) 2–4 Jun 1976.
20. D. Fest, J. Groslambert, J. J. Gagnepain, "Individual characterization of an oscillator by means of cross-correlation or cross covariance method", *IEEE Transactions on Instrumentation and Measurement* vol. 32 no. 3 pp. 447–450, September 1983.
21. R. N. McDonough, A. D. Whalen, *Detection of signals in noise*, Academic Press 1995, ISBN 0-12-744852-7.
22. E. Rubiola, V. Giordano, and J. Groslambert, "Double correlating interferometer scheme to measure PM and AM noise", *Electronics Letters* vol. 34 no. 1, January 8th, 1998.

23. E. Rubiola, V. Giordano, "Correlation-based noise measurements below the thermal noise floor", Proc. *13th European Frequency and Time Forum / 1999 Frequency Control Symposium*, Besançon (France), 12–16 April 1999. To be printed.
24. E. Rubiola, V. Giordano, J. Groslambert, "Improved interferometric method to measure near-carrier AM and PM noise", *IEEE Transactions on Instrumentation and Measurement* vol. 48 no. 2 pp. 642–646, April 1999.

Phonon Fine Structure in the $1/f$ Noise of Metals, Semiconductors and Semiconductor Devices

Mihai N. Mihaila *

Institute of Microtechnology, PO Box 38-160, str. Erou Iancu Nicolae 32B, 72996 Bucharest, Romania

Abstract. Results pointing to phonon participation in the $1/f$ noise of metals, semiconductors and semiconductor devices are presented. A fine structure corresponding to both bulk and surface phonons is shown to exist in the $1/f$ noise of different solid-state physical systems. It is described how Phonon Density Of States (PDOS) superposition method can be used to identify surface and bulk equilibrium atomic motions as microscopic $1/f$ noise sources. A close connection between the $1/f$ noise parameter and PDOS or Eliashberg function is suggested. Examples proving that the temperature dependence of the $1/f$ noise parameter is the image of the lattice vibration spectrum are given for both metals and semiconductors. Consequently, a simple connection between the activation energy distribution and PDOS is revealed and lattice anharmonicity appears to naturally affect the frequency exponent.

1 Introduction

During his famous experiment on thermal noise of electricity in conductors, Johnson [1] observed that under a given frequency f, the noise in a vacuum diode increases while the frequency decreases. The noise intensity was found to be inversely proportional to the frequency: $S_I \sim 1/f$. Schottky [2] has called it flicker noise. In almost 75 years since then, $1/f$ noise – as it is called now- has been found in an enormous number of physical systems. In solid-state systems, such as metals, superconductors, semiconductors and semiconductor devices, $1/f$ noise is almost omnipresent. Over the years, the phenomenon proved to be of formidable ubiquity entailing a very complex, sometimes entangled, problematics. It has been vividly investigated, discussed and disputed in literature and throughoutly reviewed by experts in the field[3]-[16].

Due to its ubiquity, $1/f$ noise practically eluded all attempts to find its microscopic origin. In solid, there is a general consensus that $1/f$ noise is due to resistance(conductivity) fluctuations. This is mainly due to the fundamental experiments of Voss and Clarke [17] and Beck and Spruit [18] who found $1/f$ noise in thermal noise, both in metals and semiconductors. The main

* mihaim@imt.ro

M. Planat (Ed.): LNP 550, pp. 216–231, 2000.
© Springer-Verlag Berlin Heidelberg 2000

unsolved dispute is what is the microscopic mechanism of conductivity fluctuations: is it the carrier number or the mobility which fluctuates with a $1/f$ spectrum? The mobility fluctuation hypothesis is supported by experiments strongly suggesting the phonon involvement in $1/f$ noise generation [19]-[26]. A set of observations and experiments, all pointing to phonon participation in the $1/f$ noise, is described in this paper. Phonon-based $1/f$ noise models are beyond the scope of this paper.

2 Mobility Fluctuation $1/f$ Noise Induced by Lattice Scattering

Where $1/f$ noise comes from has been investigated since the thirties when Bernamont [27] originated the idea that one can obtain a $1/f$ spectrum by superposition of lorentzians. Van der Ziel [28] and du Pre [29] showed that superposition of lorentzians requires the existence of a distribution of the relaxation times(τ): $D(\tau) \sim 1/\tau$.

The first microscopic model of $1/f$ noise is due to McWhorter [30] who attributed the phenomenon to the tunneling of the carriers into the oxide(surface) states. If the concentration of these states is constant, one can obtain a distribution $D(\tau) \sim 1/\tau$, as required by the phenomenological models[28],[29]. McWhorter's model is, in essence, a number fluctuation model and it is applicable to semiconductors and semiconductor devices. As a corollary, the idea that $1/f$ noise is a surface effect has been introduced. In 1969, collecting a large number of existing experimental data for both metals and semiconductors, Hooge[31] found that the noise intensity is inversely proportional to the total number of the carriers in the sample(N). From this dependence, he concluded that $1/f$ noise is a bulk effect.

Quantitatively, Hooge expressed his findings by heuristic formula:

$$\frac{S_V}{V^2} = \frac{\gamma}{Nf} \ , \tag{1}$$

where S_V is the noise spectral density of a fluctuating voltage (V) developed across the terminals of a linear resistor when a current is injected into it; γ is a variable which is now generally adopted under the name of $1/f$ noise parameter.

For semiconductors, the introduction of the $1/f$ noise parameter paved the way toward mobility fluctuation hypothesis, whose roots are in the application of the relation (1) to the explanation of $1/f$ noise in concentration cells[4] and thermo e.m.f.[32]. Before long, experiments performed on semiconductors with different impurity concentrations revealed that from all scattering mechanisms only mobility fluctuation due to lattice scattering generates $1/f$ noise[19].

3 Phonon Fine Structure in the $1/f$ Noise of Semiconductor Devices

Soon after its introduction by Hooge, the $1/f$ noise parameter became a very useful mean to characterize the noisiness of semiconductor devices. Very low γ values(of the order 10^{-8}) found by Schmidt et al. [33] in near ballistic n^+nn^+ GaAs submicron diodes were "attributed to the near absence of phonon collisions". This was an indirect confirmation that $1/f$ noise was "caused by lattice phonon collisions"[33]. For a given batch of bipolar transistors, it has been found that statistical distribution of γ vs. base current(I_B) exhibits maxima at some currents. It was possible to show that this maxima correspond to phonon energies in silicon [21]. Moreover, for a given transistor, it turned out that both $\gamma(I_B)$ [35] and $S_I(I_B)$ [36] features a phonon fine structure. In this way, phonon energies in phosphorous doped silicon were determined by $1/f$ noise spectroscopy. Investigating $1/f$ noise in a GaAs Schottky tunnel diode at 77 K and 4.2 K, Carruthers [34] found a fine structure in the noise intensity vs. applied bias. Almost all structure in noise was in coincidence with a structure in the second derivative d^2I/dV^2. A peak observed at 36 mV was in good correspondence with the energy of the LO phonon at Γ in GaAs. The shape of the noise spectrum at some voltages was a lorentzian. The spectrum at 36 mV was not presented, so that no direct link between the LO phonon and the $1/f$ noise is possible.

At room temperature, the mobility of a two-dimensional electron gas (2DEG) is almost entirely determined by phonon scattering, so that a 2DEG is an ideal system to verify phonons participation in $1/f$ noise. Such measurements were done recently on InAlAs/InGaAs High Electron Mobility Transistors [37]-[39]. Figure 1 shows that at very low drain voltage, the $1/f$ noise spectral density of the drain voltage fluctuations does not smoothly follow the Ohm's law(V_{DS}^2), as required by linear response theory, but features a fine structure(indicated by arrows) which corresponds to the specific phonon energies of InAlAs/InGaAs system. The first two singularities located at 28.2 mV and 33.7 mV are in good agreement with InAs-like(28.5 meV) and GaAs-like(33.7 meV) longitudinal optical(LO) phonon energies in InGaAs, respectively. The singularity at 36.5 mV fits with a 36 meV LX intervalley phonon in InAlAs, while those at 42 mV and 48.6 mV can be assigned to 43 meV transversal optical(TO) and 48 meV(LO), both AlAs-like phonons in InAlAs, respectively. The structure in $1/f$ noise mirrors a phonon fine structure existing in both drain-source resistance and second derivative of I-V characteristic, as was shown recently [39]. Such a correlation gives strong experimental support to the phonons participation in the generation of $1/f$ noise in semiconductor devices.

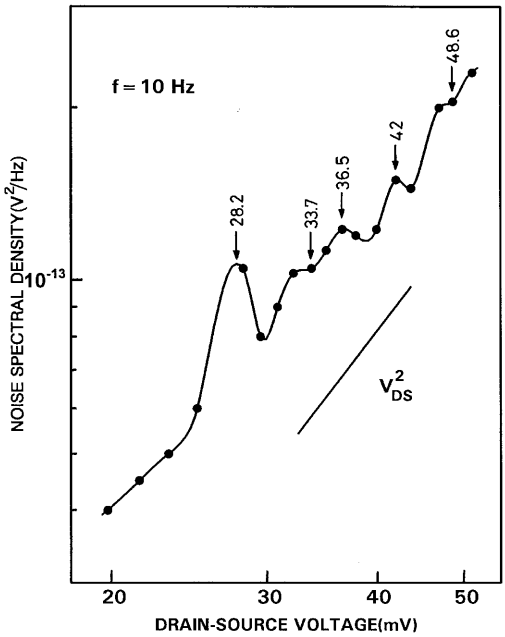

Fig. 1. Dependence of the $1/f$ noise drain voltage spectral density on the drain voltage V_{DS} for an InAlAs/InGaAs HEMT; the arrows along with the associated voltages indicate the noise fine structure. Dots are experimental data, while continuous line was obtained by a spline function fitting procedure(from [38])

4 Surface and Bulk Phonons in the $1/f$ Noise of Metals

In metallic point contacts(PC) at very low temperature and in ballistic regime, a nonmonotonic dependence of $1/f$ noise intensity on the applied bias has been observed by Akimenko et al. [22]. The structure they have observed has been attributed to both normal and Umklapp phonons. In the diffusion and thermal regimes the noise structure dissapears.

Since with the work of Eberhard and Horn [40], peaks in the temperature dependence of the $1/f$ noise parameter were often found for both continuous [41]–[51] and discontinuous [53]–[55] metal films. A close examination of Eberhard and Horn's [40] noise data for continuous noble metal films revealed not only peaks but also kinks [23]. From their noise data, the $1/f$ noise parameter has been calculated by taking into account the temperature dependence of the frequency exponent $m(1/f^m)$. In this way, the structure in γ vs. temperature for copper, silver [23] and gold [52] became more evident. To understand what the origin of this structure would be, the associated energies($k_{\mathrm{B}}T$, k_{B} the Boltzmann constant) have been compared with the specific bulk phonon

energies of each metal. A very good correspondence has been found. These results pointed to the electron-bulk phonons interaction as microscopic source of $1/f$ noise in metals.

Further experiments performed on discontinuous platinum films [53]–[55], which have both surface and bulk properties, revealed that under about 200 K, the relative noise spectral density shows a fine structure. It has been possible to associate this structure not only with bulk phonons but also with surface phonons. In this way, surface phonons were observed in the $1/f$ noise of a metal. Therefore, it has been established that the carrier-phonon interaction is the common microscopic source of $1/f$ noise at both surface and in the bulk. The observation of surface and bulk phonons in $1/f$ noise conciliates the perennial dispute whether $1/f$ noise is a surface or a bulk effect, in the sense that it offers a common physical explanation. Consequently, whether $1/f$ noise is a surface or a bulk effect is irrelevant from physical point of view.

5 $1/f$ Noise Induced by Surface and Bulk Atomic Motion

For some time, a kind of atomic motion has been suspected to be the source of $1/f$ noise especially in metals [10],[49]-[51],[56],[57] but the role of lattice vibrations(phonons) was not explicitely recognized. Usually, under 200 K, the temperature dependence of $1/f$ noise in discontinuous platinum films is very complicated. Attempting to model such a structure, we have developed what is was called Phonon Density Of States(PDOS) superposition method [53]-[55]. The PDOS method consists in the superposition of bulk and surface phonon spectra. The resulting function is used to model the temperature dependence of the normalized $1/f$ noise intensity.

An example of how this method works is shown in Fig. 2, where the noise data for the R_{AB} film and the function $F(\omega)$ are compared [55]. The function $F(\omega)$ was obtained by superposition of some surface PDOS and bulk PDOS for platinum [58]. For surface, only PDOS corresponding to the longitudinal motion of the platinum atoms in the first atomic layer, both in the ΓK and ΓM symmetry directions, were taken from Kern et al. [59]. For comparison to be possible, the phonon frequency(ω) was converted into an equivalent temperature($T = \hbar\omega/k_B$, where terms have their usual meaning). The reasonable fit between $F(\omega)$ and the noise data points to the longitudinal motion of surface atoms as a source of $1/f$ noise. In a similar way, the vertical motion of the surface atoms was also identified as source of $1/f$ noise [55]. These data showed that both longitudinal and vertical(Rayleigh wave) motion of the surface atoms as well as bulk atomic motion are microscopic sources of $1/f$ noise. In this way, the fundamental sources of $1/f$ noise were identified.

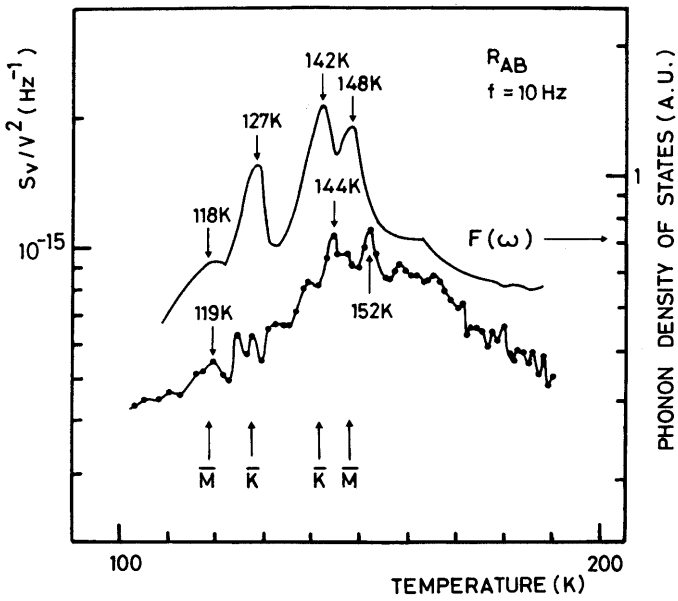

Fig. 2. Comparison between the normalized $1/f$ noise spectral density of the R_{AB} film and the function $F(\omega)$; the arrows with the associated letters indicate the symmetry directions in the platinum surface Brillouin zone (from[55])

6 Physical Significance of the $1/f$ Noise Parameter

An interesting consequence of the PDOS method is that it allowed for a physical explanation of the $1/f$ noise parameter [55],[60]. Using this method, we have found that the function obtained by superposition of platinum surface and bulk phonon densities of states fits well the noise experimental results. From these data, one can write:

$$S_v/V^2 \sim \sum F_i(\omega) \tag{2}$$

where $F_i(\omega)$ is the PDOS function of the i-th phonon mode (surface or bulk) and ω is the phonon frequency. From the relations (1) and (2), it results:

$$\gamma \sim \sum F_i(\omega) = F(\omega) \tag{3}$$

Relation (3) suggests a possible connection between γ and PDOS[F(ω)]. Our data showed that each noise peak develops in a small temperature intervals and can be almost exclusively attributed to a single phonon mode(branch).

Consequently, for a single peak one can write: $S_V/V^2 \sim F_i(\omega)$. The contribution of each atomic vibration mode could be weighted by the associated electron-phonon matrix element(α_i^2), so as in the Hooge's form we have:

$$S_v/V^2 \sim \gamma \sim \alpha_i^2 F_i(\omega) \tag{4}$$

But the product $\alpha_i^2 F_i(\omega)$ of the squared matrix element(α_i^2) for the electron-phonon interaction and the phonon density of states $F_i(\omega)$ is known as the Eliashberg function. Consequently, the formula (4) suggests a possible connection between $1/f$ noise parameter and the Eliashberg function.

Such a connection, if any, can explain –at least qualitatively– many unclear and apparently dissimilar aspects of the $1/f$ noise in solid-state physical systems. In general, according to the relation (4), large $1/f$ noise(γ) would be expected in materials with strong electron-phonon coupling, but this prediction has to be experimentally verified. The new interpretation of the $1/f$ noise parameter suggests that its temperature dependence would follow the energy dependence of the Eliashberg function or phonon density of states.

7 Image of Phonon Spectrum in $1/f$ Noise

7.1 Metals

The validity of the relation (4) has been recently verified for the case of tin, aluminum and a copper nanobridge [60]. The example of tin is given in what follows. For a comparison between $1/f$ noise parameter and the Eliashberg function to be possible, we need both noise data and the Eliashberg function for the same metal. The phonon spectrum, $F(\omega)$, is usually determined by inelastic neutron spectroscopy [58], while the Eliashberg functions for different metals were determined by both inelastic electron tunneling spectroscopy [61] and point-contact spectroscopy at very low temperature. The superconducting properties of tin(Sn) were intensively studied [61]. Also its noise properties at the superconducting transition were investigated by Clarke and Hsiang [62]. Fleetwood and Giordano [63] have found that, at room temperature, the $1/f$ noise intensity is substrate-dependent. In a further study, Fleetwood and coworkers [42] presented very interesting and reliable results for the temperature dependence of $1/f$ noise parameter in tin films on both sapphire and glass substrates. Their noise data for a tin film on a sapphire substrate are presented in Fig. 3.

A peak in the noise has been observed at about 240 K. Above this temperature the noise is substrate dependent, while for lower temperatures the substrate dependence is "much smaller or negligible" [42]. This may indicate that noise could be determined by some intrinsic properties of the metal. For comparison, the Eliasberg function $\alpha^2 F(\omega)$ for Sn, as determined by Rowell et al. [61] in a tin-tin oxide-tin(Sn-I-Sn) tunnel junction in superconducting

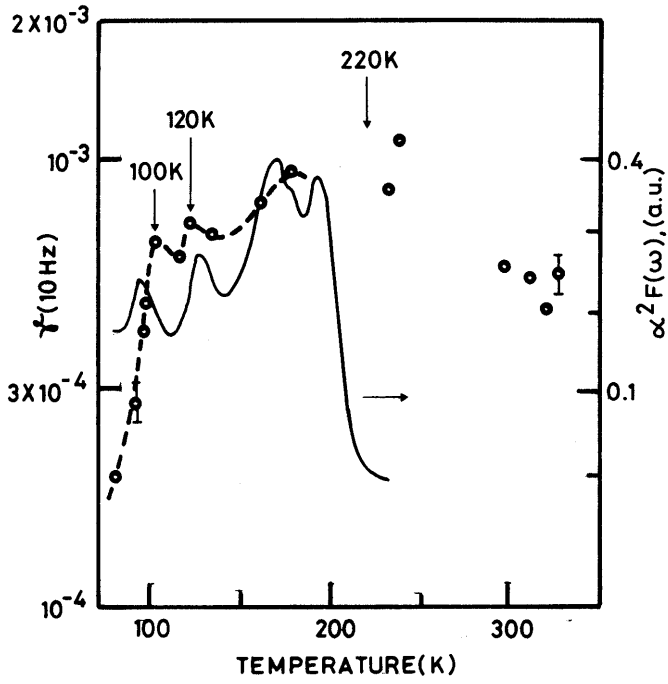

Fig. 3. Comparison between the temperature dependence of the $1/f$ noise parameter(γ) for tin(from Fleetwood et al. [42]) and the Eliashberg function($\alpha^2F(\omega)$, from Rowell et al. [61]).

state, is also presented in Fig. 3. Since $\alpha^2F(\omega)$ is a function of the tunnel electron energy(voltage across the tunnel junction), this energy was converted into an equivalent temperature: $T = \hbar\omega/k_B$, where $\hbar$ and k_B are Planck and Boltzmann constant, respectively, and ω is phonon frequency. In this way, the comparison between the noise data and Sn Eliashberg function is possible. Note the good correspondence between the two curves especially for temperature lower than 150 K. It is remarkable that the two noise peaks, indicated by arrows at 100 K and 120 K, are in reasonable agreement with some peaks located at 93 K and 127 K in the Eliashberg function, respectively. There seems also to be an asymmetrical noise peak at about 185 K which could be well described by the shape of the Eliashberg function. For temperature higher than about 200 K, $\alpha^2F(\omega)$ strongly decreases, the maximum of atomic vibration energy being of about 19 meV. The equivalent temperature of this energy is of about 220 K(indicated by arrow in Fig. 3).

It results that at least for T <150 K, in Sn film on sapphire the temperature dependence of $1/f$ noise parameter mirrors the Eliashberg function, as predicted by relation (4). As noted by Rowell et al. [61], in the case of Sn, the

electron-phonon coupling parameter α^2 is relatively slow energy dependent. Great difficulties could be encountered when α^2 is energy dependent, $\alpha^2(\omega)$, and electron-phonon strength differs from one mode of atomic vibrations to another.

7.2 Semiconductors

Starting with the work of Montgomery [64] on germanium filaments, the temperature dependence of $1/f$ noise in semiconductors has been quite intensively studied [20],[65]-[68]. Many of these data were interpreted in terms of mobility fluctuation due to phonon-carrier interaction. Processes of thermally activated kinetics were also proposed [65]. For different reasons, eq. (4) cannot be easily applied to semiconductors. Assuming a slow energy dependence for α_i^2, $\alpha_i^2 F_i(\omega)$ would feature the same structure as $F_i(\omega)$. Consequently, the relation (3) have been applied to explain some temperature dependences of γ in germanium, gallium arsenide and silicon [71].
The temperature dependence of the $1/f$ noise parameter in p-type germanium has been determined by Bisschop and Cuijpers [66] for samples of different resistivities. Their noise data for a sample of 1Ωcm are shown in Fig.4 (curve a). It is a matter of evidence that $\gamma(T)$ features a fine structure composed mainly of peaks, some of them indicated by arrows. In Fig. 4, the results of Montgomery [64] for single crystal filaments of germanium are compared with those of Bisschop and Cuijpers [66]. Striking similarities exist between these data, although they were obtained at a considerable interval of time. For temperature T higher than about 166 K ($1000/T \simeq 6\ K^{-1}$), although the shape of the curves is different, there is an almost one-by-one correspondence between the noise peaks. Even the point which in the Bisschop's data appears as an anomaly at about 200 K seems to correspond to a very well defined peak in the Montgomery's results. For T¡166 K, the noise data vary quite similar with temperature. Germanium bulk PDOS function obtained by inelastic neutron scattering [69] is also presented in Fig. 4. For comparison, the phonon frequency ω has been converted into an equivalent temperature: $T = \hbar\omega/k_B$, where the terms have their usual meaning. The noise peaks located at about 166 K and 286 K are in excellent correspondence with van Hove singularities in PDOS. For temperature lower than 166 K, namely, in the region where the PDOS is determined by the transversal acoustic phonons, some general trends in the noise behaviour are reasonably described by the PDOS shape. Comparison between the temperature dependence of the $1/f$ noise parameter.

The noise data of Chen et. al. [68] for two δ-doped, MBE-grown GaAs samples are compared with GaAs bulk PDOS function [69] in Fig. 5. Note that in both samples, the noise peaks correspond to the van Hove singularities in PDOS, which are indicated by arrows along with the associated atomic vibration modes in different symmetry directions.

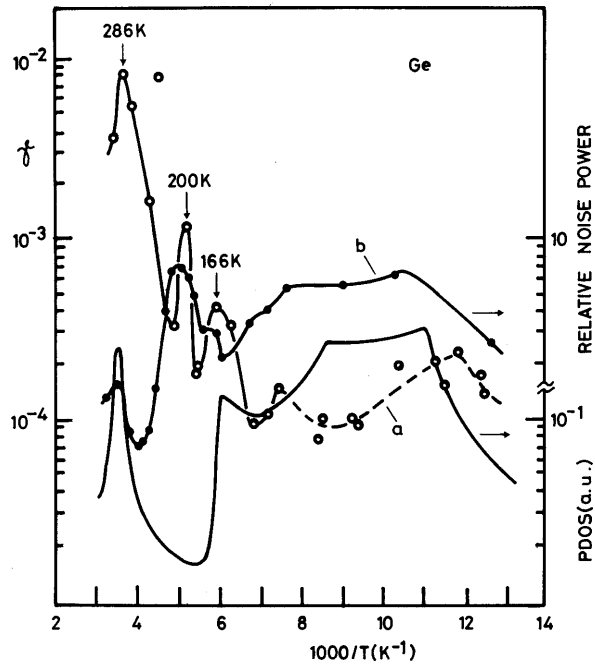

Fig. 4. Comparison between $\gamma(T)$ for germanium (Montgomery [64]–curve b; Bisschop and Cuijpers [66]–curve a) and germanium bulk PDOS obtained by inelastic neutron scattering(Dolling and Cowley [69]). Note the one-by-one correlation between the noise structure(indicated by vertical arrows) of the two curves (a and b) for $T \geq$166 K.

For T <125 K ($1000/T$=8 K^{-1}), then in the region where the PDOS is dominated by transversal acoustic(TA) phonons, there is a reasonable fit between $\gamma(T)$ and PDOS only for the sample W294 [68]. In this temperature range, the noise level in the sample W261 is much lower. According to the relation (3), $\gamma(T)$ could be determined by different combinations of F_i functions, so that some functions could lack. The shape of $\gamma(T)$ for the sample W261 indicates the absence of the function F_i of the TA$[\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ phonons, so that it is a consequence of the fact that carriers do not interact with this atomic vibration mode. A similar situation appears in the noise data of Chen [67], which are represented as γ/T vs. T in Fig. 5. Three peaks are visible in this dependence, all being located at van Hove singularities in PDOS. The two peaks located at temperatures under 200 K($1000/T = 5K^{-1}$) are very weak, indicating a weak interaction between carriers and phonon modes having energy excitation thresholds under 200 k_B(17.2 meV). But in the Dutta, Dimon and Horn(DDH) model [70], $\gamma/T \sim D(E)$, where $D(E)$ is the activation energy(E) distribution function. Because $\gamma \sim F(\omega)$, it simply results

that $\gamma/T \sim F(\omega)/T$ or, equivalently, $\gamma/T \sim F(\omega)/\omega \sim D(E)$. This observation is a hint that a connection between $F(\omega)$ and $D(E)$ could also exist [71].

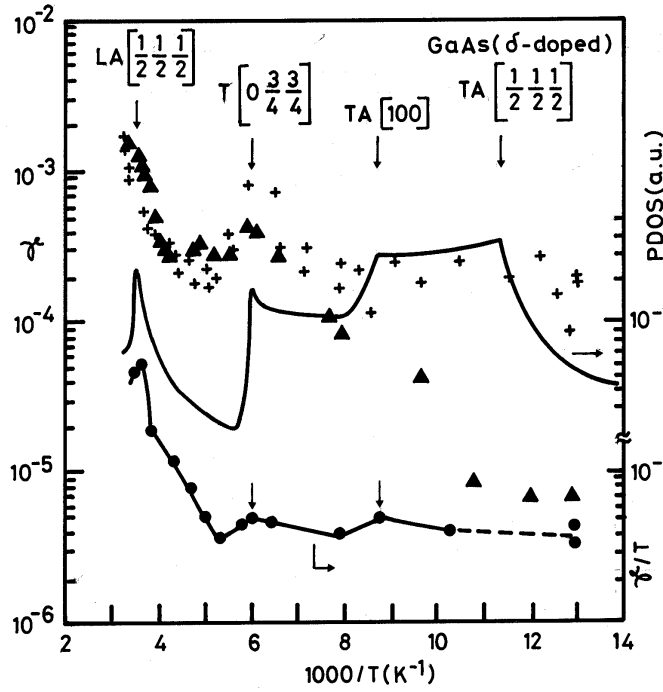

Fig. 5. Comparison between $\gamma(T)$ for δ-doped MBE-grown GaAs (Chen et al. [68]) and GaAs bulk PDOS (Dolling and Cowley [69]): black triangle-sample $W261$; $+-$sample $W294$. Figured with dots is γ/T calculated from Chen [67]. The vertical arrows indicate the van Hove singularities in the PDOS along with the associated atomic vibration modes in different symmetry directions of the bulk Brillouin zone; TA and LA are the transversal and longitudinal acoustic vibration modes, respectively.

Physical significance of the activation energy distribution. In search of the possible connection $F(\omega)/\omega \sim D(E)$, we have calculated both γ/T and $F(\omega)/\omega$ from the noise data of Palenskis and Shoblitzkas [20] for n-type silicon and bulk silicon PDOS of Dolling and Cowley [69], respectively. When represented vs. T, as in the Fig. 6, it appears that $F(\omega)/\omega$ is capable to reproduce the structure of γ/T, so that – at least for silicon- the activation energy distribution function, D(E), seems to be the image of $F(\omega)/\omega$.

In all presented cases, it appeared that temperature dependence of $1/f$ noise parameter is more or less the image of lattice vibration spectrum. This

new approach reveals the existence of some possible new, to date, hidden aspects of the $1/f$ noise. One of them is that these observations, if correct at all, strongly support the idea that **1/f noise comes from the perpetual equilibrium atomic motion**. This mechanism may well be the explanation why Voss and Clarke [17] observed $1/f$ noise in thermal noise of both semiconductors and metals. Quite surprisingly, it becomes now apparent that a connection between phonon-induced mobility fluctuations and DDH model [70] could exist. It comes about because, if one introduces $\mathrm{F}(\omega) \sim \gamma \sim \mathrm{S}_V$ instead of S_V in the analytical expression of the frequency exponent(m) [70], it results:

$$m \sim 1 + \frac{1}{\ln(f_0/f)} \left[\frac{\partial \ln F(\omega)}{\partial \ln \omega} \frac{\partial \ln \omega}{\partial \ln T} - 1 \right] , \tag{5}$$

where $\partial \ln \omega / \partial \ln T$ represents the pure **anharmonicity or lattice self-energy shift**. Hence, for the first time, lattice anharmonicity (nonlinearity) explicitely appears as a factor affecting the $1/f$ noise exponent. That may partially explain why DDH [70] phenomenological model proved so successful in explaining the temperature dependence of m in silicon on sapphire [65] and metals [6],[40],[47],[50-51],[70]. Except for the measuring frequency(f), Eq. (5) shows that m is entirely determined by lattice-specific parameters. Other factors affecting m can be the phonon occupation function and the matrix element.

8 Conclusion

The hypothesis of mobility fluctuation due to lattice scattering as microscopic source of $1/f$ noise was very briefly presented. As arguments in favour of this hypothesis, a phonon fine structure was shown to exist in the $1/f$ noise of different semiconductor devices. Moreover, both bulk and surface phonons were identified in the $1/f$ noise of metals. PDOS method was described. Using this method, surface and bulk atomic motions were identified as fundamental sources of $1/f$ noise. Also, the PDOS method allowed for a physical explanation of the $1/f$ noise parameter. In this new interpretation, a connection between the $1/f$ noise parameter and the phonon density of states/Eliashberg function would exist. The validity of this connection was exemplified by interpreting some existing data for metals (tin) and semiconductors. In all cases, it has been found that the temperature dependence of the $1/f$ noise parameter is the image of lattice vibration spectrum. It indicates that $1/f$ noise is brought about by equilibrium, perpetual atomic motion. A connection between the activation energy distribution function and the phonon density of states has been revealed. As a consequence of the new interpretation, lattice anharmonicity explicitely appears in the analytical expression of the frequency exponent. These results revealed the existence of

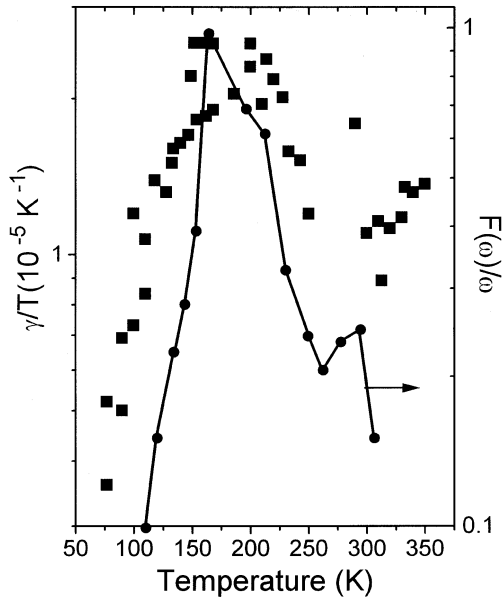

Fig. 6. Comparison between $\gamma(T)/T \sim k_B D(E)$ calculated from Palenskis and Shoblitskas' data [20] for n-type silicon and $F(\omega)/\omega$(in arbitrary units) calculated from Dolling and Cowley [69]. From these data, it seems that, for silicon, $F(\omega)/\omega$ is the image of the activation energy distribution function $D(E)$.

some (to date, hidden) unifying factors between some apparently dissimilar models of $1/f$ noise.

References

1. Johnson, J. B. (1928) Thermal agitation of electricity in conductors. Phys. Rev. 32: 97–113
2. Schottky, W. (1926) Small shot effect and flicker effect. Phys. Rev. 28: 74–103
3. van der Ziel, A. (1970) Noise in solid-state devices and lasers. Proc. IEEE 58: 1178–1206
4. Hooge, F. N. (1976) $1/f$ noise. Physica 83B: 14–23
5. van der Ziel, A. (1979) Flicker noise in electronic devices. Adv. in Electron. and Electron Phys. , L. Marton and C. Marton(editors), Academic Press, 49:225–297
6. Dutta, P., Horn, P. M. (1981) Low-frequency fluctuations in solids: $1/f$ noise. Rev. Mod. Phys. 53: 497–516
7. Hooge, F. N., Kleinpenning, T. G. M., Vandamme, L. K. J. (1981) Experimental studies on $1/f$ noise. Rep. Prog. Phys. 44: 479–532
8. Kogan, Sh. M. (1985) Curent noise with $1/f$ type spectra in solid. Sov. Phys.-Usp. 28: 285–328

9. van der Ziel, A. (1988) Unified presentation of $1/f$ noise in electronic devices: fundamental $1/f$ noise sources. Proc. IEEE 76: 233-258
10. Weissman, M. B. (1988) $1/f$ noise and other slow, nonexponential kinetics in condensed matter. Rev. Mod. Phys. 60: 537–571
11. Giordano, N. (1989) Defect motion and low frequency noise in disordered metals. Reviews of Solid State Science 3: 27–69
12. van Vliet, Carolyn M. (1991) A survey of results and future prospects on quantum $1/f$ noise and $1/f$ noise in general. Solid-St. Electron. 34: 1–21
13. Dyakonova, N. V., Levinstein, M. E., Rumyantsev. S. L. (1991) Nature of the bulk $1/f$ noise in GaAs and Si(review). Sov. Phys. Semicond. 25:1241–1265
14. Special issue on noise (1994) IEEE Trans. Electron Devices ED-41
15. Jones, B. K. (1997) Electrical noise in semiconductors: the role of phonons. Unsolved Problems of Noise, Ch. Doering, L. B. Kiss, M. Shlesinger(editors), World Scientific, 27–38
16. Weissman, M. B. (1997) Some unsolved problems in $1/f$ conductance noise. Unsolved Problems of Noise, Ch. Doering, L. B. Kiss, M. Shlesinger(editors), World Scientific, 39–56
17. Voss, Richard F., Clarke, John (1976) $1/f$ noise from systems in thermal equilibrium. Phys. Rev. Lett. 36: 42-45
18. Beck, H. G. E., Spruit, W. P. $1/f$ noise in the variance of Johnson noise. J. Appl. Phys. 49: 3384–3386
19. Hooge, F. N. , Vandamme, L. K. J. (1978) Lattice scattering causes $1/f$ noise. Phys. Lett. 66A:315–316
20. Palenskis, V., Z. Shoblitskas, Z. (1982) Origin of $1/f$ noise. Solid St. Comm. 43:761–763
21. Mihaila, M. (1984) Phonon observations from $1/f$ noise measurements. Phys. Lett. 104A:157–158
22. Akimenko, A. I., Verkin, A. B., Yanson, I. K. (1984) Point contact noise spectroscopy of phonons in metals. J. Low Temp. Phys. 54: 247–266
23. Mihaila, M. (1985) Phonon signature in the $1/f$ noise parameter of copper, silver and silicon. Phys. Lett. 107A: 465–467
24. Planat, M., Gagnepain, J. J. (1987) $1/f$ noise in quartz crystal resonators in relation with acoustic losses and frequency dispersion. Appl. Phys. Lett. 50: 510–512
25. Musha, T., Borbely, G., Shogi, M. (1990) $1/f$ phonon-number fluctuations in quartz observed by laser light scattering. Phys. Rev. Lett. 64:2394–2397
26. Koslowski, B., Baur, C. et al. (1993) Atomic scale variation of current noise in GaAs(110) detected by a scanning tunneling microscope. Surf. Science 280:106–114
27. J. Bernamont (1937) Fluctuations de potentiel aux bornes d'un conducteur metallique de faible volume parcouru par un courant. Ann. Phys.(Leipzig)7:71–140
28. van der Ziel, A. (1950) On the noise spectra of semi-conductor noise and of flicker effect. Physica XVI:359–372
29. du Pre, F. K. (1950) A suggestion regarding the spectral density of flicker noise. Phys. Rev. 78: 615
30. McWhorter, A. L. (1957) $1/f$ noise and germanium surface properties. Semiconductor Surface Physics, R. H. Kingston(editor), 207-228
31. Hooge, F. N. (1969) $1/f$ noise is no surface efect. Phys. Lett. 29A:139–140

32. Kleinpenning, T. G. M. (1974) $1/f$ noise in thermo e.m.f of intrinsic and extrinsic semiconductors. Physica 77:78–98
33. Schmidt, R. R., Bosman, G. et al. (1983) Noise in near-ballistic n+nn+ and n+pn+ gallium arsenide submicron diodes. Solid-St. Electron. 26:437–444
34. Carruthers, Tom (1971) Bias-dependent structure in excess noise in GaAs Schottky tunnel junctions. Appl. Phys. Lett. 18: 35–37
35. Mihaila, M. (1986) Lattice vibrations in silicon by $1/f$ noise spectroscopy. Noise in Phys. Syst. and $1/f$ Noise, Elsevier Science, A. D'Amico and P. Mazzetti (editors), 433–435
36. Mihaila, M. (1987) A possible fundamental property of the $1/f$ noise in condenced matter systems. Noise in Phys. Syst. and $1/f$ Noise, World Scientific, C.M. van Vliet (editor), 343–346
37. Mihaila, M., Heedt, C., Tegude, F. J. (1996) Nonlinear effects in the $1/f$ noise of lattice-matched InAlAs/InGaAs HEMT's. AIP Conf. Proc. 371, P. H. Handel, A. L Chung (editors), 127–133
38. Mihaila, M., Heedt, C. et al. (1996) Origin of $1/f$ noise in InAlAs/InGaAs HEMT's. Proc. IPRM, 368–371
39. Mihaila, M., Mihaila, A.-P. (1997) $1/f$ noise phonon spectroscopy in InAlAs/InGaAs HEMT's. Noise in Phys. Syst. anf $1/f$ Noise, World Scientific, C. Claeys and E. Simoen(editors), 51–54
40. Eberhard, J. W., and Horn, P. M., (1978) Excess $(1/f)$ noise in metals. Phys. Rev. 18B: 6681–6693
41. Kilmer, J., van Vliet, C. M. et al. (1984) Evidence of electromagnetic quantum $1/f$ noise found in gold thin films. Phys. Stat. Sol. 121b: 429–432
42. Fleetwood, D. M., Beutler, D. E. et al. (1987) The role of temperature in sample-to-sample comparisons of the $1/f$ noise of metal films J. Appl. Phys. 61: 5308–5313
43. Giordano, N. (1996-II) Low-frequency electrical noise in Ni: The effects of magnetic fluctuations. Phys. Rev. 53B: 14937–14940
44. Fleetwood, D. M., Giordano, N. (1985) Direct link between $1/f$ noise and defects in metal fields. Phys. Rev. 31B:115-7-1160
45. Scofield, J. H., Mantese J., Webb, W. W. (1986) Temperature dependence of $1/f$ noise processes in metals. Phys. Rev. 34B: 723–731
46. Zimmerman, N. M. and Webb, W. W. (1988) Microscopic scatterer displacements generate the $1/f$ resistance noise of H in Pd. Phys. Rev. Lett. 61: 889-892
47. Keener, C. D., Weissman, M. B. (1991-I) $1/f$ noise in bismuth consistent with defect motion. Phys. Rev. 44B: 9178–9184
48. Ralls, K. S, Buhrman, R. A. (1991-I) Microscopic study of $1/f$ noise in metal nanobridges. Phys. Rev. 44B:5800-5817
49. Koch, R. H., Lloyd, J. R., and Cronin, J. (1985) $1/f$ noise and grain-boundary diffusion in aluminum alloys. Phys. Rev. Lett. 55: 2487–2490
50. Briggmann, J., Dagge, K. et al. (1994) Irradiation-induced defects in thin aluminium films studied be $1/f$ noise. Phys. Stat. Sol.(a) 146:325–335
51. Van den Homberg, M. J. C., Verbruggen, A. H. et al. (1998) $1/f$ noise in mono-and polycrystalline aluminum. Phys. Rev. 57B: 53–55
52. Mihaila, M. (1986) Phononic structures in the $1/f$ noise parameter of the gold films. Noise in Phys. Syst. and $1/f$ Noise, A. D'Amico and P. Mazzetti(editors), Elsevier , 437–439

53. Mihaila, M., Stepanescu, A., and Masoero, A. (1991) Phonon fingerprints in the $1/f$ noise of discontinuous platinum films. Noise in Phys. Syst. and $1/f$ Noise, T. Musha, S. Sato and M. Yamamoto(editors), Ohmsha Ltd., 17–22
54. Mihaila, M., Stepanescu, A., Masoero, A. (1995) Longitudinal motion of surface atoms generates $1/f$ noise in discontinuous platinum films. Noise in Phys. Syst. and $1/f$ Fluctuations, V. Bareikis and R. Katilius (editors), World Scientific, 307–310
55. Mihaila, M., Mihaila, A.-P. (1997) Enhanced $1/f$ noise induced by atomic Rayleigh waves in discontinuous platinum films. Unsolved Problems of Noise, Ch. Doering, L. B. Kiss, M. F. Shlesinger(editors), World Scientific, 81–86
56. Verbruggen, A. H., Koch, R. H., and Umbach, C. P. (1987-I) Correlations between $1/f$ noise and grain boundaries in thin gold films. Phys. Rev. 35B:5864–5867
57. Ochs, E., Seeger., A. et al. (1998) Electrical noise in nanocrystalline films. Phys. Stat. Sol.(rapid research note)
58. Dutton, D. H., Brockhouse, B. N., Miiller, A. P. (1972) Crystal dynamics of platinum by inelastic neutron scattering. Can. J. Phys. 50: 2915–2927
59. Kern, Klaus, David, Rudolf et al. (1986) Surface phonon dispersion of platinum (111). Phys. Rev. B33: 4334–4337
60. Mihaila, M. (1999) Possible connection between $1/f$ noise parameter and the Eliashberg function, van der Ziel Symposium on Quantum $1/f$ Noise, P. H. Handel and A.L. Chung(editor), AIP Press, 466:48–54
61. Rowell, J. M., McMillan, W. L., Feldmann, W. L. (1971) Superconductivity and lattice dynamics of white tin. Phys. Rev. 3B: 4065–4073
62. Clarke, J., Hsiang, T. Y. (1976) Low-frequency noise in tin and lead films at the superconducting transition. Phys. Rev. 13B: 4790–4800
63. Fleetwood, D. M. and Giordano, N. (1982) Experimental study of excess low-frequency noise in tin. Phys. Rev. 25B:1427–1430
64. Montgomery, H. C. (1952) Electrical noise in semiconductors. Bell System Tech. J. 31:950–975
65. Black, R. D., Restle, P. J., Weissman, M. B. (1983) Hall effect, anisotropy, and temperature-dependence measurements of $1/f$ noise in silicon on sapphire. Phys. Rev. 28B:1935–1943
66. Bisschop, J., Cuijpers, J. L. (1983) Experimental temperature dependence of $1/f$ fluctuations in germanium and silicon. Physica B123: 6–10
67. Chen, X. Y. (1997) Temperature dependence of $1/f$ noise. Unsolved Problems of Noise, Ch. Doering, L. B. Kiss, M. F. Shlesinger(editors), World Scientific, 111-116
68. Chen, X. Y., Koenraad, P. M. et al. (1997-II) $1/f$ noise in d-doped GaAs analyzed in terms of mobility fluctuations. Phys. Rev. B55: 5290–5296; Chen, X. Y. (1997) Ph. D. Thesis, Eindhoven Univ. of Technology
69. Dolling, G., Cowley, R. A. (1966) The thermodynamic and optical properties of germanium, silicon, diamond and gallium arsenide. Proc. Phys. Soc., London 88:463–494
70. Dutta, P, Dimon, P., Horn, P. M. (1979) Energy scales for noise processes in metals. Phys. Rev. Lett. 43:646–649
71. Mihaila, M. (1999) Image of lattice vibration spectrum in the $1/f$ noise of semiconductors. 15th Int. Conf. on Noise in Phys. Syst. and $1/f$ Fluctuations (Hong Kong),Ch. Surya (editor), Bentham Press, 146-149

The General Nature of Fundamental $1/f$ Noise in Oscillators and in the High Technology Domain

Peter H. Handel *

Department of Physics and Astronomy, University of Missouri - St. Louis, St. Louis, MO 63121, United States

Abstract. The ubiquitous fundamental $1/f$ noise spectrum is derived ontologically from quantum electrodynamics for any current, cross section or process rate as a universal macroscopic quantum fluctuation process, and as the most important infrared divergence phenomenon. It is present both in space and in time and is described in both the frequency and time domains. For small and ultrasmall devices and solid state samples the conventional quantum $1/f$ effect is introduced, and for larger sizes the coherent quantum $1/f$ effect applies. This new aspect of quantum mechanics is described by simple practical relations and by a general interpolation formula. The connection with numbers of harmonics and subharmonics characterizing the basic non-linearity of the particle-field interaction as well as other nonlinearities, is elucidated. Finally, the general epistemological explanation of the ubiquitous $1/f$ phenomenon is derived in the form of a sufficient $1/f$ noise criterion, and is applied to various nonlinear systems, including the particle-field system of quantum-electrodynamics as a particular example. Application of the frequency mixing $1/f$ frequency noise experiments of Planat and collaborators is considered as an equivalent way to understand the derivation of the fundamental $1/f$ spectrum. This new approach could reveal how our earlier qualitative interpretation of lattice-dynamical quantum $1/f$ effects below the lowest transversal acoustic phonon frequency of the sample in terms of subharmonics can be verified.

1 Introduction

Fluctuations with a spectral density proportional to $1/f$ are found in a large number of systems in science, technology and everyday life. These fluctuations are known as $1/f$ noise in general. They have first been noticed by Johnson [1] in early amplifiers, have limited the performance of vacuum tubes in the thirties and forties, and have later hampered the introduction of semiconductor devices.

$1/f$ noise is present as a limitation in most modern high-technology devices. It is present in the resonance frequency of quartz resonators, in SAW devices, in junction and MIS infrared detectors, in SQUIDs, in electron beams, etc. Outside the domain of electrophysics $1/f$ noise is present in the rate of radioactive decay, in the flow rate of sand in an hourglass, in the flux of cars

* handel@jinx.umsl.edu

M. Planat (Ed.): LNP 550, pp. 232–264, 2000.
© Springer-Verlag Berlin Heidelberg 2000

on an expressway, in the frequency of sunspots, in the light output of quasars, in the flow rate of the Nile over the last 2000 years, in the water current velocity fluctuations at a depth of 3100 meters in the Pacific ocean, and in the loudness and pitch fluctuations of classical music. It has also been found below 10^{-8} Hz in the angular velocity of the earth's rotation, and below 10^{-4} Hz in the relativistic neutron flux in the terrestrial atmosphere [2].

The present paper is focused on the general origin of fundamental $1/f$ noise as a universal form of chaos, and on the cause of its ubiquity. It starts with a special case of the general $1/f$ noise phenomenon, the Quantum $1/f$ Effect (with its conventional and coherent contributions) which is as fundamental as time and space. The paper then presents the general case of $1/f$ fluctuations as a necessary consequence of the mathematical homogeneity of the dynamical (or physical) equations describing the motion of an arbitrary chaotic or stochastic nonlinear system. A sufficient criterion is derived, which indicates if an arbitrary system governed by a given system of differential equations will exhibit $1/f$ noise. The criterion is then applied to several particular systems, and is used to predict the fundamental quantum $1/f$ effect as a special case. Elsewhere the nature of the entropy increase implied by quantum $1/f$ noise is clarified based on the new quantum information theory and the negative quantum entropy concept. Again, in another paper, the characteristic functional is derived, and the fractal dimension of this oldest and most basic form of quantum chaos is discussed. In this paper we focus on the connection between the quantum $1/f$ theory and the number theoretical approach in Sec. 11.

2 Conventional Quantum 1/f Effect

This effect [3]– [8] is present in any cross section or process rate involving charged particles or current carriers. The physical origin of quantum $1/f$ noise is easy to understand. Consider for example Coulomb scattering of current carriers, e.g., electrons on a center of force. The scattered electrons reaching a detector at a given angle away from the direction of the incident beam are described by DeBroglie waves of a frequency corresponding to their energy. However, some of the electrons have lost energy in the scattering process, due to the emission of bremsstrahlung. Therefore, part of the outgoing DeBroglie waves is shifted to slightly lower frequencies. When we calculate the probability density in the scattered beam, we obtain also cross terms, linear both in the part scattered with and without bremsstrahlung. These cross terms oscillate with the same frequency as the frequency of the emitted bremsstrahlung photons. The emission of photons at all frequencies results therefore in probability density fluctuations at all frequencies. The corresponding current density fluctuations are obtained by multiplying the probability density fluctuations by the velocity of the scattered current carriers. Finally, these current fluctuations present in the scattered beam will

be noticed at the detector as low frequency current fluctuations, and will be interpreted as fundamental cross section fluctuations in the scattering cross section of the scatterer. While incoming carriers may have been Poisson distributed, the scattered beam will exhibit super-Poissonian statistics, or bunching, due to this new effect which we may call quantum $1/f$ effect. The quantum $1/f$ effect is thus a many-body or collective effect, at least a two-particle effect, best described through the two-particle wave function and two-particle correlation function.

Let us estimate the magnitude of the quantum $1/f$ effect semiclassically by starting with the classical (Larmor) formula $2q^2\mathbf{a}^2/3c^3$ for the power radiated by a particle of charge q and acceleration $\mathbf{a}$. The acceleration can be approximated by a delta function $\mathbf{a}(t) = \Delta\mathbf{v}\ \delta(t)$ whose Fourier transform $\Delta\mathbf{v}$ is constant and is the change in the velocity vector of the particle during the almost instantaneous scattering process. The one-sided spectral density of the emitted bremsstrahlung power $4q^2(\Delta\mathbf{v})^2/3c^3$ is therefore also constant. The number $4q^2(\Delta\mathbf{v})^2/3hfc^3$ of emitted photons per unit frequency interval is obtained by dividing with the energy hf of one photon. The probability amplitude of photon emission $\left[4q(\Delta\mathbf{v})^2/3hfc^3\right]^{1/2} e^{i\gamma}$ is given by the square root of this photon number spectrum, including also a phase factor $e^{i\gamma}$. Let ψ be a representative Schrödinger catalogue wave function of the scattered outgoing charged particles, which is a single-particle function, normalized to the actual scattered particle concentration. The beat term in the probability density $\rho = |\psi|^2$ is linear both in this bremsstrahlung amplitude and in the non-bremsstrahlung amplitude. Its spectral density will therefore be given by the product of the squared probability amplitude of photon emission (proportional to $1/f$) with the squared non-bremsstrahlung amplitude which is independent of f. The resulting spectral density of fractional probability density fluctuations is obtained by dividing with $|\psi|^4$ and is therefore

$$|\psi|^{-4}\, S_{|\psi|^2}(f) = 8q^2(\Delta\mathbf{v})^2/3hfNc^3 = 2\alpha A/fN = j^{-2}S_j(f), \tag{1}$$

where $\alpha = e^2/\hbar c = 1/137$ is the fine structure constant and the expression $\alpha A = 4q^2(\Delta\mathbf{v})^2/3hc^3$ is known as the infrared exponent in quantum field theory, and is known as the quantum $1/f$ noise coefficient, or Hooge constant, in electrophysics.

The spectral density of current density fluctuations is obtained by multiplying the probability density fluctuation spectrum with the squared velocity of the outgoing particles. When we calculate the spectral density of fractional fluctuations in the scattered current j, the outgoing velocity simplifies, and therefore (1) also gives the spectrum of current fluctuations $S_j(f)$, as indicated above. The quantum $1/f$ noise contribution of each carrier is independent, and therefore the quantum $1/f$ noise from N carriers is N times larger; however, the current j will also be N times larger, and therefore in (1) a factor N was included in the denominator for the case in which the cross section fluctuation is observed on N carriers simultaneously.

The fundamental fluctuations of cross sections and process rates are reflected in various kinetic coefficients in condensed matter, such as the mobility μ and the diffusion constant D, the surface and bulk recombination speeds s, and recombination times τ, the rate of tunneling j_t and the thermal diffusivity in semiconductors. Therefore, the spectral density of fractional fluctuations in all these coefficients is given also by (1).

When we apply it to a certain device, we first need to find out which are the cross sections σ or process rates which limit the current I through the device, or which determine any other device parameter P, and then we have to determine both the velocity change $\Delta\mathbf{v}$ of the scattered carriers and the number N of carriers simultaneously used to test each of these cross sections or rates. Then (1) provides the spectral density of quantum $1/f$ cross section or rate fluctuations. These spectral densities are multiplied by the squared partial derivative $(\partial I/\partial\sigma)^2$ of the current, or of the device parameter P of interest, to obtain the spectral density of fractional device noise contributions from the cross sections and rates considered. After doing this with all cross sections and process rates, we add the results and bring (factor out) the fine structure constant α as a common factor in front. This yields excellent agreement with the experiment [9] in a large variety of samples, devices and physical systems.

(1) was derived in second quantization, using the commutation rules for boson field operators. For fermions one repeats the calculation replacing in the derivation the commutators of field operators by anticommutators, which yields [6], [7].

$$\rho^{-2}S_\rho(f) = j^{-2}S_j(f) = \sigma^{-2}S_\sigma(f) = 2\alpha A/f(N-1). \tag{2}$$

This causes no difficulties, since $N \gg 2$ for particle correlations to be defined, and is practically the same as (9), since usually $N >> 1$. Equations (1) and (2) suggest a new notion of physical cross sections and process rates which contain $1/f$ noise, and express a fundamental law of physics, important in most high-technology applications [7].

We conclude that the conventional quantum $1/f$ effect can be explained in terms of interference beats between the part of the outgoing DeBroglie waves scattered without bremsstrahlung energy losses above the detection limit (given in turn by the reciprocal duration T of the $1/f$ noise measurement) on one hand, and the various parts scattered with bremsstrahlung energy losses; but there is more to it than that: exchange between identical particles is also important. This, of course, is just one way to describe the reaction of the emitted bremsstrahlung back on the scattered current. This reaction, itself an expression of the nonlinearity introduced by the coupling of the charged-particle field to the electromagnetic field, thus reveals itself as the cause of the quantum $1/f$ effect, and implies that the effect can not be obtained with an independent boson model. The effect, just like the classical turbulence-generated $1/f$ noise [12], is a result of the scale-invariant nonlinearity of the equations of motion describing the coupled system of matter and

field. Ultimately, therefore, this nonlinearity is the source of the $1/f$ spectrum in both the classical and quantum form of the theory. We can say that the quantum $1/f$ effect is an infrared divergence phenomenon, this divergence being the result of the same nonlinearity. The quantum $1/f$ effect is, in fact, the first time-dependent infrared radiative correction. Finally, it is also deterministic in the sense of a well determined wave function, once the initial phases γ of all field oscillators are given. In quantum mechanical correspondence with its classical turbulence analog [12], the new effect is therefore a quantum manifestation of classical chaos which we can take as the definition of a certain type of quantum chaos.

3 Derivation of the Coherent Quantum $1/f$ Noise Effect

The coherent quantum $1/f$ effect is a quantum fluctuation effect present in any extended current due to the definition of the physical electron as a system composed of the bare particle plus its electomagnetic field which is in a coherent state and has therefore indefinite energy. This in turn causes the state to be non-stationary. The non-stationarity is expressed by the coherent quantum $1/f$ effect. An elementary physical derivation is given in Sec. 7.

The present derivation is based on the well-known new propagator $G_s(x'-x)$ derived relativistically [13] in 1975 in a new picture required by the infinite range of the Coulomb potential. The corresponding nonrelativistic form [14] was provided by Zhang and Handel:

$$\begin{aligned} -i\left\langle \Phi_0|T\psi_{S'}(x')\psi_S^+(x)|\Phi_0\right\rangle &\equiv \delta_{SS'}G_S(x'-x) \\ &= (i/V)\sum_{\mathbf{p}}\{\ \exp\ i[\mathbf{p}(\mathbf{r}-\mathbf{r}') - \mathbf{p}^2(t-t')/2m]/\hbar\}n_{\mathbf{p},s} \\ &\times\{-i\mathbf{p}(\mathbf{r}-\mathbf{r}')/\hbar + i(m^2c^2+\mathbf{p}^2)^{1/2}(t-t')(c/\hbar)\}^{\alpha/\pi}. \end{aligned} \tag{3}$$

Here $\alpha = e^2/\hbar c = 1/137$ is Sommerfeld's fine structure constant, $n_{\mathbf{p},s}$ the number of electrons in the state of momentum $\mathbf{p}$ and spin s, m the rest mass of the fermions, $\delta_{ss'}$ the Kronecker symbol, c the speed of light, $x = (\mathbf{r}, t)$ any space-time point and V the volume of a normalization box. T is the time-ordering operator which orders the operators in the order of decreasing times from left to right and multiplies the result by $(-1)^P$, where P is the parity of the permutation required to achieve this order. For equal times, T normal-orders the operators, i.e., for $t = t'$ the left-hand side of (3) is $i\left\langle \Phi_0|\psi_S^+(x)\psi_{S'}(x')|\Phi_0\right\rangle$. The state Φ_0 of the N electrons is described by a Slater determinant of single-particle orbitals.

The resulting spectral density coincides with the result $2\alpha/\pi f N$, first derived [10] directly from the coherent state of the electromagnetic field of

a physical charged particle. The connection with the conventional quantum 1/f effect was suggested later [11].

To calculate the current autocorrelation function we need the density correlation function, which is also known as the two-particle correlation function. The two-particle correlation function is defined by

$$\begin{aligned}&\left\langle \Phi_0|T\psi_S^+(x)\psi_S(x)\psi_{S'}^+(x')\psi_{S'}(x')|\Phi_0\right\rangle \\ &\quad=\left\langle \Phi_0|\psi_S^+(x)\psi_S(x)|\Phi_0\right\rangle \left\langle \Phi_0|\psi_{S'}^+(x')\psi_{S'}(x')|\Phi_0\right\rangle \\ &\quad-\left\langle \Phi_0|T\psi_{S'}(x')\psi_S^+(x)|\Phi_0\right\rangle \left\langle \Phi_0|T\psi_S(x)\psi_{S'}^+(x')|\Phi_0\right\rangle .\end{aligned} \tag{4}$$

The first term can be expressed in terms of the particle density of spin s, $n/2 = N/2V = \left\langle \Phi_0|\psi_S^+(x)\psi_S(x)|\Phi_0\right\rangle$, while the second term can be expressed in terms of the Green function (1) in the form

$$\begin{aligned}&A_{SS'}(x-x')\equiv\left\langle \Phi_0|\psi_S^+(x)\psi_{S'}^+(x')\psi_{S'}(x')\psi_S(x)|\Phi_0\right\rangle \\ &=(n/2)^2+\delta_{SS'}G_S(x'-x)G_S(x-x').\end{aligned} \tag{5}$$

The "relative" autocorrelation function $A(x-x')$ describing the normalized pair correlation independent of spin is obtained by dividing by n^2 and summing over s and s'

$$\begin{aligned}&A(x-x')=1-(1/n^2)\sum_s G_S(x-x')G_S(x'-x)=1-(1/N^2)\\ &\sum_s\sum_{\mathbf{pp}'}\{\exp\ i[(\mathbf{p}-\mathbf{p}')(\mathbf{r}-\mathbf{r}')-(\mathbf{p}^2-\mathbf{p}'^2)(t-t')/2m]/\hbar\}n_{\mathbf{p},s}n_{\mathbf{p}',s}\\ &\times\{\mathbf{p}(\mathbf{r}-\mathbf{r}')/\hbar-(m^2c^2+\mathbf{p}^2)^{1/2}(t-t')(c/\hbar)\}^{\alpha/\pi}\\ &\times\{\mathbf{p}'(\mathbf{r}-\mathbf{r}')/\hbar-(m^2c^2+\mathbf{p}'^2)^{1/2}(t-t')(c/\hbar)\}^{\alpha/\pi}.\end{aligned} \tag{6}$$

Here we have used (1). We now consider a beam of charged fermions, e.g., electrons, represented in momentum space by a sphere of radius p_F, centered on the momentum $\mathbf{p}_0$ which is the average momentum of the fermions. The energy and momentum differences between terms of different $\mathbf{p}$ are large, leading to rapid oscillations in space and time which contain only high-frequency quantum fluctuations. The low-frequency and low-wavenumber part A_l of this relative density autocorrelation function is given by the terms with $\mathbf{p}=\mathbf{p}'$

$$\begin{aligned}&A_l(x-x')=1-(1/N^2)\\ &\times\sum_s\sum_{\mathbf{pp}'}n_{\mathbf{p},s}\{\mathbf{p}(\mathbf{r}-\mathbf{r}')/\hbar-(m^2c^2+\mathbf{p}^2)^{1/2}(t-t')(c/\hbar)\}^{2\alpha/\pi}\\ &\approx 1-(1/N)|\mathbf{p}_0(\mathbf{r}-\mathbf{r}')/\hbar-mc^2\tau/\hbar|^{2\alpha/\pi}\quad\text{for}\quad p_F\ll|p_{03}-mc^2\tau/z|\end{aligned} \tag{7}$$

Here we have used the mean value theorem, considering the $2\alpha/\pi$ power as a slowly varying function of $\mathbf{p}$ and neglecting $\mathbf{p}_0$ in the coefficient of $\tau=t-t'$,

with $z \equiv |\mathbf{r} - \mathbf{r}'|$. Using the identity [15], with arbitrarily small cutoff ω_0, we obtain from (7) with $\theta \equiv |\mathbf{p}_0(\mathbf{r} - \mathbf{r}')/\hbar - mc^2\tau/\hbar|$ the exact form

$$A_l(x - x') = 1 + [(2\alpha/\pi N) \int_{\omega_0}^{\infty} (mc^2/\hbar\omega)^{2\alpha/\pi} \cos(\theta\omega) d\omega/\omega]$$
$$\times \{\cos\ \alpha + (2\alpha/\pi) \sum_{n=0}^{\infty} (\theta\omega_0)^{2n-2\alpha/\pi} [(2n)!(2n - 2\alpha/\pi)]^{-1}\}^{-1} \tag{8}$$

This indicates a $\omega^{-1-2\alpha/\pi}$ spectrum and a $1/N$ dependence of the spectrum of fractional fluctuations in density n and current j, if we neglect the curly bracket in the denominator which is very close to unity for very small ω_0. The fractional autocorrelation of current fluctuations δj is obtained by multiplying (5) on both sides with ep_0/m, and dividing by $(enp_0/m)^2$ which is the square of the average current density j, instead of just dividing by n^2. It is the same as the fractional autocorrelation for quantum density fluctuations. Then (8) for the coherent Quantum Electrodynamical chaos process in electric currents can be written also in the form

$$S_{\delta j/j}(k) \simeq [(2\alpha/\pi\omega N)(mc^2/\hbar\omega)]^{2\alpha/\pi} \simeq 2\alpha/\pi\omega N = 0.00465/\omega N \tag{9}$$

Being observed in the presence of a constant applied field, these fundamental quantum current fluctuations are usually interpreted as mobility fluctuations [3]. Most of the conventional quantum $1/f$ fluctuations in physical cross sections and process rates are also mobility fluctuations, but some are also in the recombination speed or tunneling rate.

4 Sufficient Criterion for Fundamental $1/f$ Noise

In spite of the practical success of our quantum $1/f$ theory in explaining electronic $1/f$ noise in most high tech devices, and in spite of the conceptual success of our earlier classical turbulence approach to $1/f$ noise, the question about the origin of nature's omnipresent $1/f$ spectra remained unanswered. During the last three decades, we have claimed repeatedly that nonlinearity is a general cause of $1/f$ noise. The present paper proves that nonlinearity always leads to a $1/f$ spectrum if homogeneity is also present in the equation(s) of motion. Specifically, if the system is described in terms of the dimensionless vector function $\mathbf{Y}(x,t)$ by the mth order nonlinear differential equation

$$\partial\mathbf{Y}/\partial t + \mathbf{F}(\mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/\partial x_1 ... \partial\mathbf{Y}/\partial x_n, \partial^2\mathbf{Y}/\partial x_1^2 ..... \partial^m\mathbf{Y}/\partial x_n^m) = 0 \tag{10}$$

a $1/f$ spectrum is obtained if the nonlinear function $\mathbf{F}$ satisfies the homogeneity condition

$$\mathbf{F}[\lambda\mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/(\lambda\partial x_1) ... \partial\mathbf{Y}/(\lambda\partial x_n), \partial^2\mathbf{Y}/(\lambda\partial x_1)^2 ..... \partial^m\mathbf{Y}/(\lambda\partial x_n)^m]$$
$$= \lambda^{-p}\mathbf{F}(\mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/\partial x_1) ... \partial\mathbf{Y}/\partial x_n, \partial^2\mathbf{Y}/\partial x_1^2 ..... \partial^m\mathbf{Y}/\partial x_n^m), \tag{11}$$

for any real number λ. The order of homogeneity is the number $-p$.

Performing a Fourier transformation of (9) with respect to the vector $\mathbf{x}(x_1, x_2, ...x_n)$, we get in terms of the Fourier-transformed wavevector $\mathbf{k}$ the nonlinear integro-differential equation

$$\partial \mathbf{y}(\mathbf{k},t)/\partial t + \mathbf{G}[\mathbf{k}, \mathbf{y}(\mathbf{k},t), k_1\mathbf{y}(\mathbf{k},t), ...k_n\mathbf{y}(\mathbf{k},t), \quad k_1^2\mathbf{y}(\mathbf{k},t).....k_n^m\mathbf{y}(\mathbf{k},t)] = 0, \tag{12}$$

where $\mathbf{y}(\mathbf{k},t)$ is the Fourier transform of $\mathbf{Y}(\mathbf{x},t)$. Due to (11) , the nonlinear integro-differential operator $\mathbf{G}$ satisfies the relation

$$\mathbf{G}[\lambda\mathbf{k}, \mathbf{y}, \lambda k_1\mathbf{y}...\lambda k_n\mathbf{y}, (\lambda k_1)^2\mathbf{y}...(\lambda k_n)^m\mathbf{y}] = \lambda^p\mathbf{G}[\mathbf{k}, \mathbf{y}, k_1\mathbf{y}...k_n\mathbf{y}, k_1^2\mathbf{y}.....k_n^m\mathbf{y}], \tag{13}$$

where the integration differentials $d\mathbf{k}$, $d\mathbf{k}'$, etc., are excepted from replacement with $\lambda d\mathbf{k}$, $\lambda d\mathbf{k}'$, etc. Equation (12) can thus be rewritten in the form

$$d\mathbf{y}/d(t/\lambda^p) + G[\lambda\mathbf{k}, \mathbf{y}, \lambda k_1\mathbf{y}...\lambda k_n\mathbf{y}, (\lambda k_1)^2\mathbf{y}.....(\lambda k_n)^m]\mathbf{y}] = 0 \tag{14}$$

Taking $\lambda = 1/k$, where $k = |\mathbf{k}| = (k_1^2 + ... + k_n^2)^{1/2}$, and setting $k^p t = z$, we notice that k has been eliminated from the dynamical equation, and only $\mathbf{k}/k$ is left. This means that there is no privileged scale left for the system in $\mathbf{x}$ or $\mathbf{k}$ space, other than the scale defined by the given time t, and expressed by the dependence on z. We call this property of the dynamical system "sliding-scale invariance".

In certain conditions, instabilities of a solution of (9) may generate chaos, or turbulence. In a sufficiently large system described by the local dynamical equation (9), in which the boundary conditions become immaterial, homogeneous, isotropic turbulence, (chaos) can be obtained, with a spectral density determined only by (9). The stationary autocorrelation function $A(\tau)$ is defined as an average scalar product, the average being over the turbulent ensemble

$$A(\tau) = \langle \mathbf{Y}(\mathbf{x},t)\mathbf{Y}(\mathbf{x},t+\tau)\rangle = \int \langle \mathbf{y}(\mathbf{x},t)\mathbf{y}(\mathbf{x},t+\tau)\rangle d^n k = \int u(\mathbf{k},z)d^n\mathbf{k}. \tag{15}$$

Here we have introduced the scalar

$$u(\mathbf{k},z) = \langle \mathbf{y}(\mathbf{k},t)\mathbf{y}(\mathbf{k},t+\tau)\rangle \tag{16}$$

of homogeneous, isotropic chaos (turbulence), which depends only on $|\mathbf{k}|$ and $z = k^p\tau$. All integrals are from minus infinity to plus infinity. The chain of integro-differential equations for the correlation functions of any order obeys the same sliding-scale invariance which we have noticed in the fundamental dynamical equation above. Therefore, in isotropic, homogeneous, conditions,

u can only depend on k and z. Furthermore, the direct dependence on k must reflect this sliding-scale invariance, and is therefore of the form

$$u(\mathbf{k}, z) = \mathbf{k}^{-n} v(z) \tag{17}$$

Indeed, only this form insures that $u(\mathbf{k}, z)d^n\mathbf{k}$ and therefore also the corresponding integrals and multiple convolutions in $\mathbf{k}$ space have the necessary sliding-scale invariance.

According to the Wiener-Khintchine theorem, the spectral density is the Fourier-transform of $A(t)$,

$$S_y(f) = \int e^{2\pi i f\tau} A(\tau) d\tau = (1/f) \int e^{2\pi i t'} \int \mathbf{k}'^{-n} v(z) d^n\mathbf{k}' dt' = C/f \tag{18}$$

where we have set $f\tau = t'$, $\mathbf{k}^n = f\mathbf{k}'^n$, $z = \mathbf{k}^n\tau = \mathbf{k}'^n t'$, and the integral

$$C = \int e^{2\pi i t'} \int \mathbf{k}'^{-n} v(z) d^n\mathbf{k}' dt' = \int e^{2\pi i t'} \int \mathbf{k}''^{-n} v(\mathbf{k}"^n) d^n\mathbf{k}" dt' \tag{19}$$

is independent of f. We have defined the vector $\mathbf{k}" = t'^{1/n}\mathbf{k}$.

The general form of our criterion considers a system described in terms of the integro-differential system of equations

$$\Phi[t, \mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/\partial t, \partial\mathbf{Y}/\partial x_1 ... \partial\mathbf{Y}/\partial x_n, \partial^2\mathbf{Y}/\partial t^2, \partial^2\mathbf{Y}/\partial x_1^2, ...\partial^m\mathbf{Y}/\partial x_n^m] = 0 \tag{20}$$

where the vector function Φ may be nonlinear in any of its arguments. If a number θ exists such that (20) implies

$$\Phi[\lambda^\theta t, \lambda\mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/\lambda^\theta\partial t, \partial\mathbf{Y}/\lambda\partial x_1 ... \partial\mathbf{Y}/\lambda\partial x_n, \partial^2\mathbf{Y}/\lambda^{2\theta}\partial t^2, \partial^2\mathbf{Y}/\lambda^2\partial x_1^2 .....\partial^m\mathbf{Y}/\lambda^m\partial x_n^m] = 0, \tag{21}$$

for any real number λ, the power spectral density of any chaotic solution for the vector function $\mathbf{Y}$ defined by (10) is proportional to 1/f.

Here we have assumed that there are no boundary conditions associated with (21), or that any boundary conditions included would satisfy the same homogeneity conditions.

In conclusion, nonlinearity + homogeneity = $1/f$ noise, provided the system is chaotic. The ultimate cause of the ubiquitous $1/f$ noise in nature is the omnipresence of nonlinearities (no matter how weak) and homogeneity. The latter is finally related to rotational (or Lorentz) invariance and therefore to the isotropy of space (or space-time). All our four specific theories of $1/f$ chaos in nonlinear systems are just special cases to which this criterion is applicable. They include our magneto-plasma theory of turbulence for current carriers in intrinsic symmetric semiconductors [12] (1966), our similar theory for metals [12] (1971), the quantum 1/f theory [3]–[11] (pure QED, 1975), and

the spectral theory of Musha's highway traffic turbulence results [16] (1989). Applied to the motion of a nonlinearly interacting chain of atoms, it predicts no $1/f$ spectrum. Starting from a wrong defining equation of the chain, both our criterion and direct calculation allowed for $1/f$ noise in a special case [16], but the correct defining equation does not fulfill the criterion, and no $1/f$ spectrum is expected. However, $1/f$ fluctuations in phonon number, in frequency, and in phase are predicted by the criterion, are derived directly [17] with the quantum $1/f$ theory, and have been experimentally verified [17],[18], in piezoelectric crystals.

5 Application to QED: Quantum $1/f$ Effect as a Special Case

The nonlinearity causing the $1/f$ spectrum of turbulence in both semiconductors and metals is caused by the reaction of the field generated by charged particles and their currents back on themselves. The same nonlinearity is present in quantum electrodynamics (QED), where it causes the infrared divergence, the infrared radiative corrections for cross sections and process rates, and the quantum $1/f$ effect. We shall prove this on the basis of our sufficient criterion for $1/f$ spectral density in chaotic systems.

Consider a beam of charged particles propagating in a well-defined direction which we shall call the x direction, so that the one-dimensional Schrödinger equation describes the longitudinal fluctuations in the concentration of particles. Considering the non-relativistic case which is encountered in most quantum $1/f$ noise applications, we write in second quantization the equation of motion for the Heisenberg field operators ψ of the in the form

$$i\hbar\frac{\partial\psi}{\partial t} = (1/2m)[-i\hbar\nabla - (e/c)A]^2\psi, \tag{22}$$

With the non-relativistic form $\mathbf{J} = -i\hbar\psi^*\nabla\psi/m +$ hermitian conjugate, and with

$$A(x,y,z,t) = (\hbar/2cmi)\int\frac{\psi^*\nabla\psi - \psi\nabla\psi^*}{|x-x'|}dx' \tag{23}$$

we obtain

$$i\hbar\frac{\partial\psi}{\partial t} = (1/2m)[-i\hbar\nabla - (e\hbar/2c^2mi)\int\frac{\psi^*\nabla\psi - \psi\nabla\psi^*}{|x-x'|}dx']^2\psi. \tag{24}$$

At very low frequencies or wave numbers the last term in rectangular brackets is dominant on the r.h.s., leading to

$$i\hbar\frac{\partial\psi}{\partial t} = (-1/2m)[(e\hbar/2c^2m)\int\frac{\psi^*\nabla\psi - \psi\nabla\psi^*}{|x-x'|}dx']^2\psi. \tag{25}$$

For x replaced by λx, and x' replaced by $\lambda x'$, we obtain

$$i\hbar\frac{\partial\psi}{\partial t} = (-1/2m)[(e\hbar/2c^2m)\int\frac{\psi^*\nabla/\lambda\psi - \psi\nabla/\lambda\psi^*}{|x-x'|}\lambda^3 dx']^2\psi$$
$$= \lambda^2 H\psi = \lambda^{-p}H\psi. \tag{26}$$

This satisfies our homogeneity criterion with $p = -2$. Our sufficient criterion only requires homogeneity, with any value of the weight p, for the existence of a $1/f$ spectrum in chaos. Therefore, we expect a $1/f$ spectrum of quantum current-fluctuations, i.e., of cross sections and process rates in physics, as derived in detail earlier [3]-[8],[10]-[11]. This is in agreement with the well-known, and experimentally verified, results of the Quantum $1/f$ Theory.

In conclusion, we realize that, both in classical and quantum mechanical nonlinear systems, the limiting behavior at low wave numbers is usually expressed by homogeneous functional dependences, leading to fundamental $1/f$ spectra on the basis of our criterion.

6 Derivation of the Conventional Quantum $1/f$ Noise Effect in Second Quantization

The simplified description of quantum $1/f$ noise was presented above in the elementary terms of Schrödinger's statistical catalogue model, without using second quantization. This approach is natural in view of the close connection between this new effect and diffraction which is usually treated without second quantization, in the statistical catalogue model based on the single-particle solution of the Schrödinger equation, normalized to the number of particles N. Just as the superposition of elementary phase-shifted waves allows for the simplest and most intuitive description of diffraction through a slit, the description of quantum $1/f$ noise in terms of interference beats between slightly frequency-shifted scattered partial waves with bremsstrahlung energy losses will always provide the simplest and most elementary quantitative derivation of the quantum $1/f$ effect, easily accessible even at the undergraduate level.

Below we now present the derivation of the Quantum $1/f$ Effect in a general form which determines the scattered current j from the observation of a sample of N outgoing particles. The minimal outgoing sample for defining particle-particle correlations in the scattered wave consists of two particles, and therefore the effect can be calculated for the case of two outgoing particles.

We start with the expression of the Heisenberg representation state $|S\rangle$ of N identical bosons of mass M emerging at an angle θ from some scattering process with various undetermined bremsstrahlung energy losses reflected in

their one-particle waves $\phi_i(\xi_i)$

$$|S\rangle = (N!)^{-1/2} \prod_i \int d^3\xi_i \phi_i(\xi_i)\psi^+(\xi_i)|0\rangle = \prod_i \int d^3\xi_i \phi_i(\xi_i)|S^0\rangle, \qquad (27)$$

where $\psi^+(\xi_i)$ is the field operator creating a boson with position vector ξ_i and $|0>$ is the vacuum state, while $|S^0>$ is the state with N bosons of position vectors ξ_i with $i = 1.....N$. All products and sums in this Section run from 1 to N, unless otherwise stated.

To calculate the particle density autocorrelation function in the outgoing scattered wave, we need the expectation value of the operator

$$O(\mathbf{x}_1, \mathbf{x}_2) = \psi^+(\mathbf{x}_1)\psi^+(\mathbf{x}_2)\psi(\mathbf{x}_2)\psi(\mathbf{x}_1), \qquad (28)$$

known as the operator of the pair correlation. This operator corresponds to a density autocorrelation function. The presence of two-particle coordinates in the operator O does not mean that we are considering two-particle interactions, it only means that the expectation value which we are calculating depends on the relative position of the particles. Using the well known commutation relations for boson field operators

$$\begin{aligned}
&\psi(\mathbf{x})\psi^+(\mathbf{y}) - \psi^+(\mathbf{y})\psi(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{y}),\\
&\psi(\mathbf{x})\psi(\mathbf{y}) - \psi(\mathbf{y})\psi(\mathbf{x}) = 0,\\
&\psi^+(\mathbf{x})\psi^+(\mathbf{y}) - \psi^+(\mathbf{y})\psi^+(\mathbf{x}) = 0, \qquad (29)
\end{aligned}$$

we first calculate the matrix element:

$$\begin{aligned}
&N!\langle S^0|O|S^0\rangle =\\
&\sum_{\mu\nu}{}' \sum_{mn}{}' \delta(\eta_\nu - \mathbf{x}_1)\delta(\eta_\mu - \mathbf{x}_2)\delta(\xi_n - \mathbf{x}_1)\delta(\xi_m - \mathbf{x}_2) \sum_{(i,j)} \prod_{ij}{}' \delta(\eta_j - \xi_i), \qquad (30)
\end{aligned}$$

where $|S^0>$ is the state with well defined particle coordinates. Here the prime excludes $\mu = \nu$ and $m = n$ in the summations and excludes $i = m$, $i = n$, $j = \mu$ and $j = \nu$ in the product. The summation $\sum_{(i,j)}$ runs over all permutations of the remaining $N-2$ values of i and j. On the basis of this result we now calculate the complete matrix element

$$\begin{aligned}
&\langle S|O|S\rangle = [1/N(N-1)] \sum_{\mu\nu}{}' \sum_{mn}{}' \int d^3\eta_\mu \int d^3\eta_\nu \int d^3\xi_m \int d^3\xi_n\\
&\phi^*_\mu(\eta_\mu)\phi^*_\nu(\eta_\nu)\phi_m(\xi_m)\phi_n(\xi_n)\delta(\eta_\nu - \mathbf{x}_1)\delta(\eta_\mu - \mathbf{x}_2)\delta(\xi_n - \mathbf{x}_1)\delta(\xi_m - \mathbf{x}_2) =\\
&[1/N(N-1)] \sum_{\mu\nu}{}' \sum_{mn}{}' \phi^*_\mu(\mathbf{x}_2)\phi^*_\nu(\mathbf{x}_1)\phi_m(\mathbf{x}_1)\phi_n(\mathbf{x}_2). \qquad (31)
\end{aligned}$$

The one-particle states are spherical waves emerging from the scattering center located at $x = 0$:

$$\phi(\mathbf{x}) = (C/x)e^{i\mathbf{K}\mathbf{x}}[1 + \sum_{kl} b(\mathbf{k}, l)e^{-iqx}a^+_{\mathbf{k},l}]. \qquad (32)$$

Here C is an amplitude factor, $\mathbf{K}$ the boson wave vector magnitude, $b(\mathbf{k}, l)$ the bremsstrahlung amplitude for emission of photons of wave vector $\mathbf{k}$ and polarization l, while $a^+_{\mathbf{k},l} =$ is the corresponding photon creation operator, allowing the emitted photon state to be created from the vacuum if (32) is inserted into (31). The momentum magnitude loss $\hbar q = Mck/K \equiv 2\pi Mf/K$ is necessary for energy conservation in the bremsstrahlung process. The reader can derive this expression of $q = \delta p$ from the canonical equation $v = \partial H/\partial p$ which yields $\delta p = \delta H/v = M\delta H/p = M\delta H/\hbar K$. Substituting (32) into (31), we obtain

$$\langle S|O|S\rangle = |C/x|^4\{N(N-1) + 2(N-1) \times \sum_{\mathbf{k}l} |b(\mathbf{k}, l)|^2[1 + \cos\ q(x_1 - x_2)]\}, \tag{33}$$

where we neglected a small term of higher order in $b(\mathbf{k}, l)$. To perform the angular part of the summation in (33), we calculate the current expectation value of the state in (32), and compare it to the well known cross section without and with bremsstrahlung

$$\mathbf{j} = (\hbar\mathbf{K}/Mx^2)[1 + \sum_{kl} |b(\mathbf{k}, l)|^2] = \mathbf{j}_0[1 + \alpha A \int df/f], \tag{34}$$

where the quantum fluctuations have disappeared, $\alpha = e^2/\hbar c$ is the fine structure constant, $\alpha A = (2\alpha/3\pi)(\Delta\mathbf{v}/c)^2$ is the fractional bremsstrahlung rate coefficient, also known in QED as the infrared exponent, and where the $1/f$ dependence of the bremsstrahlung part displays the well-known infrared catastrophe, i.e., the emission of a logarithmically divergent number of photons in the low frequency limit. Here $\Delta\mathbf{v}$ is the velocity change $\hbar(\mathbf{K} - \mathbf{K}_0)/M$ of the scattered boson, and $f = ck/2\pi$ the photon frequency. Equation (33) thus gives

$$\langle S|O|S\rangle = |C/x|^4\{N(N-1) + 2(N-1)\alpha A \int [1 + \cos\ q(x_1 - x_2]df/f\} \tag{35}$$

which is the pair correlation function, or density autocorrelation function along the scattered beam with $df/f = dq/q$. The spatial distribution fluctuations along the scattered beam will also be observed as fluctuations in time at the detector, at any frequency f. According to the Wiener-Khintchine theorem, we obtain the spectral density of fractional scattered particle density ρ, (or current j, or cross section σ) fluctuations in terms of frequency f or wave number q by dividing the coefficient of the cosine by the constant term $N(N-1)$:

$$\rho^{-2}S_\rho(f) = j^{-2}S_j(f) = \sigma^{-2}S_\sigma(f) = 2\alpha A/fN, \tag{36}$$

where N is the number of particles or current carriers used to define the current j whose fluctuations we are studying. Quantum $1/f$ noise is thus

a fundamental $1/N$ effect. The exact value of the exponent of f in (36) can be determined by including the contributions from all real and virtual multiphoton processes of any order (infrared radiative corrections), and turns out to be $\alpha A - 1$, rather than -1, which is important only philosophically, since $\alpha A << 1$. The spectral integral is thus convergent at $f = 0$.

For fermions we repeat the calculation replacing in the derivation of (33) the commutators of field operators by anticommutators, which finally yields in the same way

$$\rho^{-2} S_\rho(f) = j^{-2} S_j(f) = \sigma^{-2} S_\sigma(f) = 2\alpha A / f(N-1), \tag{37}$$

The denominator causes no difficulties, since $N >> 2$ for particle correlations to be defined, and which is practically the same as (37), since usually $N >> 1$. Equations (36) and (37) suggest a new notion of physical cross sections and process rates which contain $1/f$ noise, and express a fundamental law of physics, important in most high-technology applications [9].

We conclude that the conventional quantum $1/f$ effect can be explained in terms of interference beats between the part of the outgoing DeBroglie waves scattered without bremsstrahlung energy losses above the detection limit (given in turn by the reciprocal duration T of the $1/f$ noise measurement) on one hand, and the various parts scattered with bremsstrahlung energy losses; but there is more to it than that: exchange between identical particles is also important. This, of course, is just one way to describe the reaction of the emitted bremsstrahlung back on the scattered current. This reaction thus reveals itself as the cause of the quantum $1/f$ effect, and implies that the effect can not be obtained with the independent boson model. The effect, just like the classical turbulence-generated $1/f$ noise [12], is a result of the scale-invariant nonlinearity of the equations of motion describing the coupled system of matter and field. Ultimately, therefore, this nonlinearity is the source of the $1/f$ spectrum in both the classical and quantum form of the author's theory. We can say that the quantum $1/f$ effect is an infrared divergence phenomenon, this divergence being the result of the same nonlinearity. The new effect is, in fact, the first time-dependent infrared radiative correction. Finally, it is also deterministic in the sense of a well determined wave function, once the initial phases γ of all field oscillators are given. In quantum mechanical correspondence with its classical turbulence analog, the new effect is therefore a quantum manifestation of classical chaos which we can take as the definition of a certain type of quantum chaos.

We turn now to the connection to the coherent Quantum $1/f$ Effect. The coherent state in a conductor or semiconductor sample is the result of the experimental efforts directed towards establishing a steady and constant current, and is therefore the state defined by the collective motion, i.e. by the drift of the current carriers. It is expressed in the Hamiltonian by the magnetic energy E_m, per unit length, of the current carried by the sample.

In very small samples or electronic devices, this magnetic energy

$$E_m = \int (B^2/8\pi) d^3x = [nevS/c]^2 \ln(R/r) \tag{38}$$

is much smaller than the total kinetic energy E_k of the drift motion of the individual carriers

$$E_k = \sum mv^2/2 = nSmv^2/2 = E_m/s. \tag{39}$$

Here we have introduced the magnetic field B, the carrier concentration n, the cross sectional area S and radius r of the cylindrical sample (e.g., a current carrying wire), the radius R of the electric circuit, and the "coherence ratio"

$$s = E_m/E_k = 2ne^2S/mc^2 \ln(R/r) \simeq 2e^2N'/mc^2, \tag{40}$$

where $N' = nS$ is the number of carriers per unit length of the sample and the natural logarithm $\ln(R/r)$ has been approximated by one in the last form. We expect the observed spectral density of the mobility fluctuations to be given by a relation of the form

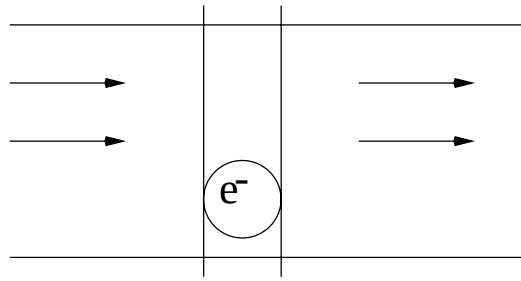

Fig. 1. To define the parameter s, a slice as thick as the classical electron radius is considered. The number of carriers in it is s

$$(1/\mu^2)S_\mu(f) = [1/(1+s)][2\alpha A/fN] + [s/(1+s)][2\alpha/\pi fN], \tag{41}$$

which can be interpreted as an expression of the effective quantum $1/f$ constant if the number N of carriers in the (homogeneous) sample is brought to the numerator of the left hand side. In this equation $\alpha A = 2\alpha(\Delta \mathbf{v}/c)^2/3\pi$ is the usual nonrelativistic expression of the infrared exponent, present in the familiar form of the conventional quantum $1/f$ effect [3]-[8]. This equation is limited to quantum $1/f$ mobility (or diffusion) fluctuations, and does not include the quantum $1/f$ noise in the surface and bulk recombination cross sections, in the surface and bulk trapping centers, in tunneling and injection processes, in emission or in transitions between two solids.

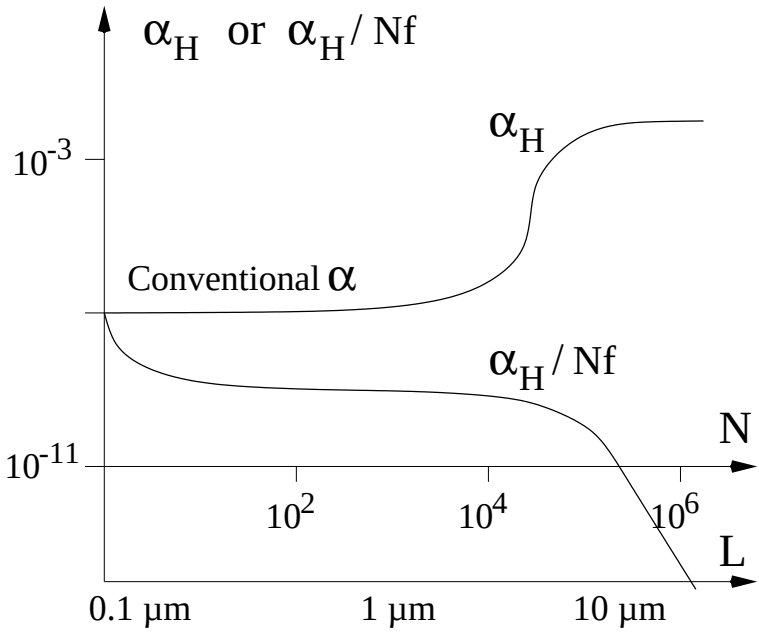

Fig. 2. The quantum $1/f$ parameter α_H and the resulting spectral density $S_j = \alpha_H/Nf$ as a function of the number of carriers in the sample N or of the cross-section size L

Note that the coherence ratio s introduced here equals the unity for the critical value $N' = N" = 2.10^{12}$/cm., e.g. for a cross section $S = 2.10^{-4}$ cm^2 of the sample when $n = 10^{16}$. This critical linear concentration corresponds to having 2 electrons in every cross-sectional "salami slice" of thickness equal to the classical radius of the electron $r = e^2/mc^2$ (see Fig. 1). For small samples with $N' << N"$ only the first term survives, while for $N' > N"$ the second term in (41) is dominant. Fig. 2 shows the resulting quantum $1/f$ constant and the fractional mobility fluctuation spectrum α_H/fN, as a function of sample size, when the cross-sectional dimension of the sample is increased.

7 Physical Derivation of Coherent Quantum $1/f$ Noise Effect

This effect arises in a beam of electrons (or other charged particles propagating freely in vacuum) from the definition of the physical electron as a bare particle plus a coherent state of the electromagnetic field. It is caused by the energy spread characterizing any coherent state of the electromagnetic field oscillators, an energy spread which spells non-stationarity, i.e., fluctuations. To find the spectral density of these inescapable fluctuations which are known

to characterize any quantum state which is not an energy eigenstate, we use an elementary physical derivation based on Schrödinger's definition of coherent states, which supplements the rigorous derivation which was given in Sec. 3 from a well-known quantum-electrodynamical propagator. The chaotic character of these fluctuations was discussed in Sec. 5.

The coherent quantum $1/f$ effect will be derived in three steps: first we consider a hypothetical world with just one single mode of the electromagnetic field coupled to a beam of charged particles; considering the mode to be in a coherent state, we calculate the autocorrelation function of the quantum fluctuations in the particle-density (or concentration) which arise from the nonstationarity of the coherent state. Then we calculate the amplitude with which this one mode is represented in the field of an electron, according to electrodynamics. Finally, we take the product of the autocorrelation functions calculated for all modes with the amplitudes found in the previous step.

Let a mode of the electromagnetic field be characterized by the wave vector $\mathbf{q}$, the angular frequency $\omega = cq$ and the polarization λ. Denoting the variables $\mathbf{q}$ and λ simply by q in the labels of the states, we write the coherent state of amplitude $|z_q|$ and phase $arg\ z_q$ in the form

$$|z_q\rangle = \exp[-(1/2)|z_q|^2] \exp[z_q a_q^+]|0\rangle = \exp[-(1/2)|z_q|^2] \sum_{n=0}^{\infty} (z_q^n)/n!|n\rangle. \tag{42}$$

Here a_q^+ is the creation operator which adds one energy quantum to the energy of the mode. Let us use a representation of the energy eigenstates in terms of Hermite polynomials $H_n(x)$

$$|n\rangle = (2^n n! \sqrt{\pi})^{-1/2} \exp[-x^2/2] H_n(x) e^{in\omega t}. \tag{43}$$

This yields for the coherent state $|z_q >$ the representation

$$\begin{aligned} \psi_q(x) &= \exp[-(1/2)|z_q|^2] \exp[-x^2/2] \sum_{n=0}^{\infty} \frac{[z_q e^{i\omega t}]^n}{[n!(2^n\sqrt{\omega})]^{1/2}} H_n(x) \\ &= \exp[-|z_q|^2/2] \exp[-x^2/2] \exp[-z_q^2 e^{-2i\omega t} + 2x z_q e^{i\omega t}] \end{aligned} \tag{44}$$

In the last form the generating function of the Hermite polynomials was used. The corresponding autocorrelation function of the probability density function, obtained by averaging over the time t or the phase of z_q, is, for $|z_q| << 1$,

$$\begin{aligned} P_q(\tau, x) &= \langle |\psi_q|_t^2 |\psi_q|_{t+\tau}^2 \rangle \\ &= \{1 + 8x^2|z_q|^2[1 + \cos\ \omega\tau] - 2|z_q|^2\} \exp[-x^2/2] \end{aligned} \tag{45}$$

Integrating over x from $-\infty$ to $+\infty$, we find the autocorrelation function

$$A^1(\tau) = 2^{-1/2}\{1 + 2|z_q|^2 \cos\ \omega\tau\}. \tag{46}$$

This result shows that the probability distribution contains a constant background with small superposed oscillations of frequency ω. Physically, the small oscillations in the total probability describe self-organization or bunching of the particles in the beam. They are thus more likely to be found in a measurement at a certain time and place than at other times and places relative to each other along the beam. Note that for $z_q = 0$ the coherent state becomes the ground state of the oscillator which is also an energy eigenstate, and therefore stationary and free of oscillations.

We now determine the amplitude z_q with which the field mode q is represented in the physical electron. One way to do this is to let a bare particle dress itself through its interaction with the electromagnetic field, i.e. by performing first order perturbation theory with the interaction Hamiltonian

$$H' = A_\mu j^\mu = -(e/c)\mathbf{v}.\mathbf{A} + e\phi, \tag{47}$$

where $\mathbf{A}$ is the vector potential and ϕ the scalar electric potential. Another way is to Fourier expand the electric potential e/r of a charged particle in a box of volume V. In both ways we obtain the well-known result

$$|z_q|^2 = \pi(e/q)^2(\hbar c q V)^{-1}. \tag{48}$$

Considering now all modes of the electromagnetic field, we obtain from the single - mode result of (46)

$$\begin{aligned} A(\tau) = C\prod_q\{1 + 2|z_q|^2\cos\ \omega_q\tau\} &= C\{1 + \sum_q 2|z_q|^2\cos\ \omega_q\tau\} \\ &= C\{1 + 4(V/2^3\pi^3)\int d^3q|z_q|^2\cos\ \omega_q\tau\} \end{aligned} \tag{49}$$

Here we have again used the smallness of z_q and we have introduced a constant C. Using (48) we obtain

$$\begin{aligned} A(\tau) &= C\{1 + 4\pi(V/2^3\pi^3)(4\pi/V)(e^2/\hbar c)\int(dq/q)\cos\ \omega_q\tau\} \\ &= C\{1 + 2(\alpha/\pi)\int\cos(\omega\tau)d\omega/\omega\}. \end{aligned} \tag{50}$$

Here $\alpha = e^2/\hbar c$ is Sommerfeld's fine structure constant. The first term in curly brackets is unity and represents the constant background, or the d.c. part of the current carried by the beam of particles through vacuum. The autocorrelation function for the relative (fractional) density fluctuations, or for the current density fluctuations in the beam of charged particles is obtained therefore by dividing the second term in curly brackets by the first term. The constant C drops out when the fractional fluctuations are considered. According to the Wiener-Khintchine theorem, the coefficient of cos $\omega\tau$ is the spectral density of the fluctuations, $S|\psi|^2$ for the particle concentration, or S_j for the current density $j = e(k/m)|\psi|^2$

$$S_{|\psi|^2}\langle|\psi|^{-2}\rangle = S_j\langle j\rangle^{-2} = 2(\alpha/\pi f N) = 4.6.10^{-3}f^{-1}N^{-1} \tag{51}$$

Here we have included the total number N of charged particles which are observed simultaneously in the denominator, because the noise contributions from each particle are independent. This result is related to the conventional Quantum $1/f$ Effect considered in the next section. A similar calculation yields the gravidynamical quantum $1/f$ effect (QGD $1/f$ effect) by substituting gravitons for the photons considered so far as infraquanta.

8 Derivation of Mobility Quantum $1/f$ Noise in $n^+ - p$ Diodes

For a diffusion limited $n^+ - p$ junction the current is controlled by diffusion of electrons into the p - region over a distance of the order of the diffusion length $L = (D_n \tau_n)^{1/2}$ which is shorter than the length w_p of the p - region in the case of a long diode. If $N(x)$ is the number of electrons per unit length and D_n their diffusion constant, the electron current at x is

$$I_{nd} = -eD_n dN/dx, \tag{52}$$

where we have assumed a planar junction and taken the origin $x = 0$ in the junction plane. Diffusion constant fluctuations, given by kT/e times the mobility fluctuations, will lead to local current fluctuations in the interval Δx

$$\delta \Delta I_{nd}(x,t) = I_{nd} \Delta x \delta D_n(x,t)/D_n. \tag{53}$$

The normalized weight with which these local fluctuations representative of the interval Δx contribute to the total current I_d through the diode at $x = 0$ is determined by the appropriate Green function and can be shown to be $(1/L)\exp(-x/L)$ for $w_p/L >> 1$. Therefore the contribution of the section Δx is

$$\delta \Delta I_d(x,f) = (\Delta x/L)\exp(-x/L) I_{nd} \delta D_n(x,t)/D_n, \tag{54}$$

with the spectral density

$$S_{\Delta I_d}(x,f) = (\Delta x/L)^2 \exp(-2x/L) I_{nd}^2 S_{D_n}(x,f)/D_n^2. \tag{55}$$

For mobility and diffusion fluctuations the fractional spectral density is given by $\alpha_{H_{nd}}/fN\Delta x$, where $\alpha_{H_{nd}}$ is determined from quantum $1/f$ theory according to Secs 2 and 6. With (52) we obtain then

$$S_{\Delta I_d}(x,f) = (\Delta x/L)^2 \exp(-2x/L)(eD_n dN/dx)^2 \alpha_{H_{nd}}/fN. \tag{56}$$

The electrons are distributed according to the solution of the diffusion equation, i.e.

$$N(x) = [N(0) - N_p]\exp(-x/L); \quad dN/dx = -\{[N(0) - N_p]/L\} \exp -x/L. \tag{57}$$

Substituting (56) into and simply summing over the uncorrelated contributions of all intervals Δx, we obtain

$$S_{I_d}(f) = \alpha_{H_{nd}}(eD_n/L^2)^2 \times \int_0^{w_p} \frac{[N(0)-N_p]^2 e^{-4x/L}}{[N(0)-N_p]e^{-x/L}+N_p}dx. \tag{58}$$

We note that $eD_n/L^2 = e/\tau_n$. With the expression of the saturation current $I_0 = e(D_n/\tau_n)^{1/2}N_p$ and of the current $I = I_0[\exp(eV/kT)-1]$, we can carry out the integration

$$S_{I_d}(f) = \alpha_{H_{nd}}(eI/f\tau_n)\int_0^1 a^2u^3du/(au+1) = \alpha_{H_{nd}}(eI/f\tau_n)F(a) \tag{59}$$

Here we have introduced the notations

$$u = \exp(-x/L), \quad a = \exp(eV/kT)-1, \quad F(a) = 1/3 - 1/2a + 1/a^2 - (1/a^3)\ln(1+a). \tag{60}$$

Equation (59), obtained by van der Ziel and Anderson, gives the diffusion noise as a function of the quantum $1/f$ noise parameter aH_{nd}. A similar result can be derived for the quantum $1/f$ fluctuations of the recombination rate r in the bulk of the p - region, the only difference being the presence of $\alpha_{H_{nr}}$ instead of $\alpha_{H_{nd}}$ in (59). The total noise is the given by (59) with $\alpha_{H_{nr}}$ replaced by the sum $\alpha_{H_{nd}} + \alpha_{H_{nr}}$

$$S_{I_d} = (\alpha_{H_{nd}} + \alpha_{H_{nr}})(eI/f\tau_n)F(a) \tag{61}$$

As we have seen in Secs. 2 and 3, quantum $1/f$ noise is a low-frequency fluctuation process present in elementary cross sections σ and process rates Γ, given by the fractional spectral density

$$S_\sigma(f)/\sigma^2 = S_\Gamma(f)/\Gamma^2 = (4\alpha/3\pi fN)(\Delta \mathbf{v}/c)^2 \tag{62}$$

for conventional quantum $1/f$ noise which is applicable to small devices, and

$$S_\sigma(f)/\sigma^2 = S_\Gamma(f)/\Gamma^2 = (2\alpha/\pi fN) = \alpha_c/fN \tag{63}$$

for coherent state quantum $1/f$ noise which is applicable to large devices in which the energy of the carrier drift motion is predominantly magnetic, rather than kinetic. The two forms of quantum $1/f$ noise are closely related infrared divergence phenomena which arise due to the interaction of electrons and soft photons.

Both in n^+p diodes [9] and metal-insulator-semiconductor (MIS) devices the current will be determined by cross sections si such as recombination and scattering cross sections (by phonons and lattice defects), as well as by other process rates (e.g., band to band and trap-assisted tunneling, particularly

important in long-wavelength $Hg_{1-x}Cd_xTe$ detectors). The spectral density of the resulting quantum $1/f$ current fluctuations is therefore

$$S_I(f) = \sum_i [\partial I/\partial \sigma_i]^2 S_{\sigma_i}(f) + \sum_i [\partial I/\partial \Gamma_i]^2 S_{\Gamma_i}(f), \tag{64}$$

where the spectral densities on the right hand side will be given by (62) or (63), depending on the size of the device.

The current density I of (64) contains a diffusion term I_d, a term I_r caused by recombination in the space charge region, a surface recombination term I_s, a tunneling term I_t and a photovoltaic term caused by the creation of electron hole pairs by photons:

$$\begin{aligned} I = I_d + I_r + I_s + I_t + q\eta\Phi = q\eta_i\{(n_i/n_0)(D_n/\tau_n)^{1/2}(e^{qV/kT} - 1) \\ + (W/\tau)(e^{qV/2kT} - 1) + s\} + I_t + q\eta\Phi. \end{aligned} \tag{65}$$

Here n_i is the intrinsic concentration, n_0 the concentration of acceptors on the p side, D_n and τ_n the diffusion constant and lifetime of minority carriers on the p side, W the width of the depletion region, $t = \tau_{p0} + \tau_{n_0}$ the Shockley-Hall-Read lifetime, V the applied voltage, s the surface recombination speed, η the quantum efficiency and Φ the incident flux of photons. With the exception of the last term, the terms in (65) are known as dark current components.

The first term in curly brackets in (65) gives the diffusion current density I_d, and yields a noise term [9] as shown in (61)

$$S_{I_d}(f) = (\alpha_d + \alpha_r)(qI_d/f\tau_n)F(a); \quad \alpha_d = (4\alpha/3\pi)(h/m^*bc)^2 \exp(-\theta/2T) \tag{66}$$

in (64) for small devices, and

$$S_{I_d}(f) = \alpha_C(qI_d/f\tau_n)F(a) \tag{67}$$

for large devices, with $F(a) = -1/3 - 1/2a + 1/a^2 - (1/a^3)\ln(1 + a)$ and $a = \exp(qV/kT) - 1$. Here b is the lattice constant, m^* the effective mass of the electrons, and q the Debye temperature. These results are obtained by adding the scattering and recombination quantum $1/f$ cross section fluctuation spectra for all points on the p side of a long n^+p diode with the appropriate Green function weight $[\partial I/\partial\sigma_i]^2$, as prescribed by (64). The spatial dependence $\exp(-x/L_d)$ of this Green function is given by the diffusion equation applied to electrons on the p side, with L_d being the electron diffusion length. If the length w_p of the p side is shorter than L_d, we have to replace $F(a)$ by $F(a) - F[a\ \exp(-w_p/L_d)]$ in (66) and (67), and also to include the quantum $1/f$ noise of the recombination cross sections at the p contact. Similarly we obtain [9]

$$\begin{aligned} &S_{I_r} = \alpha_r q I_r \tanh(qV/2kT)/f\tau \\ &= (4\alpha/3\pi c^2)[2q(V_{\text{diff}} - V) + 3kT]\{(m_n^*)^{1/2} + (m_p^*)^{1/2}\}^{-2}. \end{aligned} \tag{68}$$

The corresponding expression for the surface recombination current is given by (68) with half of the surface potential jump U added to the diffusion potential V_{diff}. The tunneling current I_t and its noise is the same as in MIS devices,and will be discussed below. In the final form of (64) the fine structure constant a appears as a general factor.

9 Quantum $1/f$ Noise in SQUIDS

As we have seen in Secs. 2 and 6 above, any cross section or process rate defined for electrically charged particles must fluctuate in time with a $1/f$ spectral density according to quantum electrodynamics, as a consequence of infrared-divergent coupling to low-frequency photons. This fundamental effect leads to quantum $1/f$ noise observed in many systems with a small number of carriers, and is also present in the cross sections and process rates which determine the resistance and tunneling rate in Josephson junctions, providing a lower limit of the observed $1/f$ noise.

In a Josephson junction the normal resistance R_n of the barrier is proportional to a scattering cross section or transition rate experienced by the electron in quasiparticle tunneling and by the Cooper pairs below the critical current I_c. Therefore

$$\begin{aligned} R_n^{-2} S_{R_n}(f) &= (4\alpha/3\pi)[(\Delta \mathbf{v})^2/c^2 N f] \\ &= (8\alpha/3\pi)(v_F^2/c^2 N f) = 4.10^{-14}/f\Omega, \end{aligned} \tag{69}$$

where we have approximated $(\Delta \mathbf{v})^2$ with $2v_F^2$, v_F being the Fermi velocity, and the number of carriers N simultaneously present in the barrier of volume Ω (in m^3) by 10^7, for barriers wider than 10^{-7} cm.

Assuming a linear relationship between the critical current I_c and $G_n = R_n^{-1}$, we obtain , from (69) for R_n, the spectral density of voltage fluctuations

$$\begin{aligned} S_V(f) &= (4/f)10^{-12}(T/3K)R_S g(V)[R_S + R_J]^2 \\ &\quad [I_c R_n (I^2/I_c^2 - 1)^{-1/2} + g(V)V]^2 \Omega^{-1} \end{aligned} \tag{70}$$

where $R_J(V)$ is the junction resistance, R_s the shunt resistance, and $g(V) = R_n/R_J$.

The noise caused in a SQUID by the source considered above can be obtained as the sum of the noise contributions from the two junctions.

The above quantum $1/f$ results of (69) and (70) are in good quantitative agreement with the experimental data.

In conclusion, the fundamental quantum $1/f$ fluctuations of the cross sections and transition rates which determine the normal resistance have been evaluated in this Section for the case of a Josephson junction. Considering the velocity change in the quantum $1/f$ formula equal to twice the Fermi velocity and the concentration of carriers in the barrier 10^{19} cm^{-3}, a spectral density of fractional fluctuations in the normal resistance of the barrier of

$4.10^{-14}/f$ was obtained for a Josephson junction with a volume of the barrier of 10^{-12}cm^3. These fluctuations are inversely proportional to the barrier volume and result in voltage fluctuations both directly and through the dependence of the critical current on the normal resistance, in good agreement with the experimental data.

10 Quantum $1/f$ Noise in Bulk Acoustic Wave and SAW Quartz Resonators

Frequency standards contain a main resonant mode which can be described as a harmonic oscillator with losses

$$d^2x/dt^2 + \gamma dx/dt + \omega_0^2 x = F(t). \tag{71}$$

The quantum 1/f fluctuations are present in the loss coefficient γ. They are given by an expression of the form

$$S_{\delta\gamma/\gamma}(f) = \Lambda/f \tag{72}$$

where Λ is a quantum $1/f$ coefficient characterizing the elementary loss process. The resonance frequency is given by

$$\omega_r^2 = \omega_0^2 + \gamma^2 \tag{73}$$

The quantum $1/f$ fluctuations in the resonance frequency are given by $\omega_r \delta\omega_r = -2\gamma\delta\gamma$, or

$$\delta\omega_r/\omega_r = -2(\gamma/\omega_r)^2 \delta\gamma/\gamma = -(1/2Q^2)(\delta\gamma/\gamma) \tag{74}$$

where $Q = \omega_r/2\gamma$ is the quality factor. Squaring, averaging and particularizing (74) for the unit frequency interval, we obtain the Q^{-4} law [21]

$$S_y(f) = \left\langle (\delta\omega_r/\omega_r)^2 \right\rangle_f = (1/4Q^4) \left\langle (\delta\gamma/\gamma)^2 \right\rangle_f = \Lambda/4fQ^4 \tag{75}$$

where y is the fractional frequency fluctuation $\delta\omega_r/\omega_r$. This law was found in approximative form empirically by Gagnepain and Uebersfeld for the case of quartz resonators. The coefficient Λ is calculated for an ideal quartz crystal as follows.

Phonon scattering in the resonator is known to limit the short- and medium-term frequency stability in all quartz resonators [1]. Phonon scattering can occur on other phonons (particularly at higher temperatures) or on crystal defects (favored by default at low temperatures). In both cases this process is shown to yield a $1/f$ spectrum of resonator frequency fluctuations through the conventional Quantum $1/f$ Effect. As was first shown on this basis with the help of a simple harmonic oscillator model [2], bulk acoustic wave (BAW) and surface acoustic wave (SAW) quartz resonators ought to

have a $Q-4$ dependence of their FM power spectrum. This has been experimentally verified by Gagnepain and Uebersfeld for BAW resonators [3] when they first noticed their $1/Q^{4.4}$ law, and by Parker for the SAW case [4]. Although the quantum $1/f$ effect provided the historical basis for the derivation of the Q^{-4} law [2] as being caused by fluctuations in the dissipation rate of the quartz resonator, the exact mechanism through which the quantum $1/f$ effect modulates the dissipation rate remained unknown from 1978 to 1991.

Finally, the bridge directly connecting $1/f$ noise in frequency standards to the quantum $1/f$ effect was discovered [5]–[7]. Here is how it works. The rate Γ of photon-interactions which remove phonons from the main quartz resonator mode is modulated by the quantum $1/f$ effect, therefore exhibiting observable quantum fluctuations, while its expectation value remains constant. Indeed, whenever a phonon is removed from the main resonator mode, the time-derivative of the polarization vector of the quartz crystal $d\mathbf{P}/dt = \dot{\mathbf{P}}$, is suddenly jolted, suffering a step-like modification as a function of time. From Maxwell's equations we know, however, that $\dot{\mathbf{P}}$ is added to the current $\mathbf{J}$ and that such a modification of the current causes radiation. Solving Maxwell's equations we find that as a result of the phonon removal there is a constant energy of $(1/4\pi\epsilon_0)4(\Delta\dot{\mathbf{P}})/3c^3$ radiated away per unit frequency, i.e. per Hertz at any frequency f. Dividing this result by the energy of a photon hf, we find that there is thus a probability of $2\alpha(\Delta\ \dot{\mathbf{P}})^2/(3\pi f e^2 c^2)$ for the emission (radiation) of a bremsstrahlung photon of frequency f. Here α is Sommerfeld's fine structure constant, a dimensionless universal constant constructed from Planck's constant, the charge e of the electron, and the speed of light c. SI units are used here, while Gaussian units were used above and in [7].

Since there is a probability $2\alpha(\Delta\ \dot{\mathbf{P}})^2/(3\pi f e^2 c^2) \ll 1$ for the emission of a photon of frequency f, the quartz crystal suffers a reaction, or a recoil, in its quantum state, causing the phonon-emission rate Γ to perform quantum oscillations with frequency f and with two-sided spectral density S' of fractional fluctuations given by the same expression $S'_{\delta\Lambda/\Lambda}(f) = 2\alpha(\Delta\dot{\mathbf{P}})^2/3\pi f e^2 c^2$. This is the quantum $1/f$ effect. The one-sided spectrum is thus

$$S_{\delta\Lambda/\Lambda}(f) = 4\alpha(\Delta\dot{\mathbf{P}})^2/3\pi f e^2 c^2 \tag{76}$$

This means that any radiation caused or implied by a quantum transition from one state to another comes with a price: it reacts back on the system, causing the rate of that transition to be modulated by exhibiting observable macroscopic quantum fluctuations with a spectral density of fractional rate fluctuations identical to the photon emission probability accompanying the transition considered. No knowledge of quantum mechanics is therefore needed in order to apply the quantum $1/f$ effect, provided one manages to divide the energy radiated by the energy hf of one photon of frequency f, i.e., provided one accepts just the reality of Planck's constant h and Planck's relation between photon energy and frequency. Knowledge of electrodynamics is needed, however, in order to calculate the energy radiated in a transition.

Our application of the quantum $1/f$ theory to bulk acoustic wave (BAW) and surface acoustic wave (SAW) quartz resonators has initially been limited to the case of very high Q resonators in which the phonons are coherent throughout the resonator volume. Below the quantum $1/f$ theory is applied to the general case of an arbitrary coherence length of the phonons, which may be large or small compared with the size of the quartz resonator. This allows to extend the theory for the first time to low-Q resonators in which the phonons are localized in a part of the resonator volume. Our theory is also extended to include defect scattering along with the phonon scattering case exclusively published so far.

Spatial incoherence of phonon loss rate fluctuations in low and intermediate Q resonators. The treatment [5]-[7] we have provided so far, assumes that the photons are spread over the whole crystal, and that therefore there is coherence of the quantum $1/f$ phonon loss rate fluctuations throughout the resonator volume. In the limit of very high Q resonators, and of high resonator frequency, the mean free path and the coherence length ϵ of the phonons exceed the size of the resonator and our assumption is justified. However, in the low Q and low frequency case the coherence length is small compared with the dimensions of the resonator crystal. The resonator volume is then composed of many incoherent regions of volume ϵ^3 which fluctuate independently. This is applicable in particular to SAW resonators, as suggested empirically by Parker et al. [9],[10]. We are here borrowing their notation of the size of the coherent volume elements by ϵ. Considering the $\nu = V/\epsilon^3$ independently fluctuating regions similar, we replace (76) by

$$S_{\delta\Omega/\Omega}(f) = \sum_{i=1}^{\nu} \langle (\delta\Gamma_i/\Gamma))^2 \rangle_f = \nu^{-1} \langle (\delta\Gamma_i/\Gamma_i))^2 \rangle_f$$
$$= 4\alpha(\Delta\dot{\mathbf{P}}_i)^2/3\pi f \nu e^2 c^2. \tag{77}$$

Here we assumed that $\Gamma = \nu\Gamma_i$ and that $\langle (\delta\Gamma_i/\Gamma_i)^2 \rangle_f = \nu S_{\delta\Gamma/\Gamma(f)}$ is independent of i. With $\nu = V/\epsilon$ we finally obtain

$$S_{\delta\Gamma/\Gamma}(f) = 4\alpha\epsilon^3(\Delta\dot{\mathbf{P}}_i)^2/3\pi f V e^2 c^2 = \Lambda/f \tag{78}$$

in the incoherent domain i.e., for sufficiently large V and small Q. The corresponding fluctuations in the frequency ω of the quartz resonator are derived from the equations

$$\omega^2 = \omega_0^2 - 2\Gamma^2, \quad \omega\delta\omega = -2\Gamma\delta\Gamma; \quad S_{\delta\omega/\omega} = (1/4Q^4) S_{\delta\gamma/\gamma}(f) \tag{79}$$

in which ω_o would be the natural frequency of the unloaded quartz resonator mode in the absence of the dissipation Γ. Therefore the spectral density of fractional frequency fluctuations $S_{\delta\omega/\omega}(f)$ will display the same $1/V$ dependence in the incoherent regime, down to a volume $V \approx \epsilon$. Below this volume we expect a proportionality of $S_{\delta\omega/\omega}(f)$ with the volume. Consequently,

$S_{\delta\omega/\omega}(f)$ will first increase proportional to $1/V$ in the incoherent region, and will decrease proportional to V when V is lowered to values below ϵ. This line of thought, however, neglects the fact that resonators operate usually in the lowest vibration mode, and that therefore a decrease in volume will correspond to a certain increase in the frequency ω_0. The latter is also connected with the speed of the sound c_s and the acoustic attenuation length $l = 1/\alpha_{ac} = 2c_S Q/\omega_0 \approx \epsilon$.

The phonon mean free path is about 40 Å for bulk wave phonons in quartz at room temperature and 500 Å at liquid nitrogen temperatures. This approximates the phonon coherence length ϵ very well. For SAW phonons the corresponding coherence length values may be 4 times lower due to the smaller velocity of the surface wave and due to its stronger scattering.

In a SAW quartz resonator the wave is localized within about two coherence lengths ϵ from the surface. Therefore, in the incoherent regime $V = 2\epsilon A$ is a good approximation. Consequently, we expect an increase of $S_{\delta\omega/\omega}(f)$ proportional with $1/A$ when the resonant area A is decreased, down to very small areas of the order of ϵ^2.

The incoherence encountered here is similar to the incoherence introduced spatially by the very small coherence length of electrons in semiconductors, of the order of 30 Å. The quantum $1/f$ effect and therefore the measured $1/f$ noise, has no spatial correlations in semiconductors down to this very small length scale. No other $1/f$ noise "models" can explain this experimental fact. The quantum $1/f$ contributions from each electron are independent, which causes a factor $1/N_e$ to be present in all quantum $1/f$ effect formulas. Here N_e is the number of electrons present in the sample, which had the transition considered as their last interaction.

In the following subsection we consider the quantum $1/f$ effect in a resonator volume equal or smaller than ϵ in all directions, the result being applicable whether or not it represents the whole resonator mode.

Spatial coherence of phonon loss rate fluctuations in high-Q resonators. Let the resonator volume V be smaller than the phonon coherence length ϵ in all directions. Then $\nu = 1$ and we can use (1). In (76)–(78), $(\Delta\dot{\mathbf{P}})^2$ is the square of the dipole moment rate change associated with the process causing the removal of a phonon from the main oscillator mode: scattering of a main resonator mode phonon on a themal phonon of higher frequency $\langle\omega\rangle \approx kT/\hbar$. After this, we will include also the case of defect scattering below. To calculate $(\Delta\dot{\mathbf{P}})^2$, we write the energy W of the interacting mode $\langle\omega\rangle$ in the form

$$W = n\hbar\langle\omega\rangle = 2(Nm/2)(dx/dt)^2 = (Nm/e^2)(e\ dx/dt)^2$$
$$= (m/Ne^2)\epsilon^2(\dot{\mathbf{P}})^2; \tag{80}$$

The factor two includes the potential energy contribution. Here m is the reduced mass of the elementary oscillating dipoles, e their charge, g a po-

larization constant of the order of the unity, and N their number in the resonator. Applying a variation $\Delta n = 1$ we get

$$\Delta n/n = 2|\Delta\dot{\mathbf{P}}||\dot{\mathbf{P}}|, \quad or \quad \Delta\dot{\mathbf{P}} = \dot{\mathbf{P}}/2n. \tag{81}$$

Solving (80) for $\dot{\mathbf{P}}$ and substituting into (81), we obtain

$$\Delta\dot{\mathbf{P}} = (N\hbar\langle\omega\rangle/n)^{1/2}(e/2g). \tag{82}$$

Substituting $\dot{\mathbf{P}}$ into (76), we get

$$\Gamma^{-2}S_\Gamma(f) = N\alpha\hbar\langle\omega\rangle/3n\pi mc^2fg^2 \equiv \Lambda/f. \tag{83}$$

This result is applicable to the fluctuations in the loss rate Γ of the resonator volume.

From (4) the corresponding resonance frequency fluctuations of the quartz is given for $V \leq \epsilon$ by

$$\omega^{-2}S_\omega(f) = (1/4Q^4)(\Lambda/f) = N\alpha\hbar\langle\omega\rangle/12n\pi mc^2fg^2Q^4; \tag{84}$$

where $Q = \omega/2\Gamma$ is the quality factor of the single-mode resonator considered, and $\langle\omega\rangle$ is not the circular frequency of the main resonator mode, but rather the practically constant frequency of the average interacting (thermal) phonon. Indeed, there are an average $n_\omega = kT/\hbar\omega$ phonons present in any mode of frequency ω. For the case of quartz resonators we have used the interacting thermal mode of average frequency $\langle\omega\rangle$ to calculate the quantum $1/f$ effect. The corresponding $\Delta\dot{\mathbf{P}}$ in the main resonator mode of frequency ω_0 has to be also included, but is negligible because of the very large number n of phonons present in the main resonator mode and entering in the denominator of (82)-(84).

Considering also (78), (84) can be written in general with $N \equiv VN/V$ in the form

$$S(f) = \beta' V/fQ^4, \text{ for } V \leq \epsilon, \tag{85}$$

and

$$S(f) = \beta'\epsilon^2/fVQ^4, \text{ for } V \geq \epsilon, \tag{86}$$

where, with an intermediary value $\langle\omega\rangle = 10^8/s$, with $n = kT/\hbar\langle\omega\rangle$, $T = 300K$ and $kT = 4.10^{-21}\ J$

$$\begin{aligned}\beta' &= (N/V)\alpha\hbar\langle\omega\rangle/12n\pi g^2mc^2 \\ &= 10^{22}(1/137)(10^{-27}10^8)^2/12kT\pi 10^{-27}\ 9\ 10^{20} \approx 1\end{aligned} \tag{87}$$

For the case of defect scattering, a two-phonon process takes place. A phonon from the main resonator mode scatters on a defect and a phonon of comparable frequency emerges into another mode with much smaller phonon occupation number $n_\omega = kT/\hbar\omega$. In this case we have to replace $\langle\omega\rangle$ by ω and

$n_{\langle\omega\rangle}$ with n_ω, which gives a β-value which is $(\langle\omega\rangle/\omega)^2$ smaller, i.e. $10^4 - 10^6$ times smaller. In general, therefore, writing $\Gamma = \Gamma' + \Gamma''$, we obtain for the combined phonon and defect scattering case, in general,

$$\beta = \beta'[\Gamma'^2 + (\langle\omega\rangle/\omega)^2\Gamma''^2]/\Gamma^2. \tag{88}$$

Although the defect scattering term is small at room temperature, it may become dominant at low temperatures, when the phonon scattering rate Γ' becomes much smaller than the defect scattering rate Γ".

The form of (85)–(88) shows that the level of $1/f$ frequency noise depends not only as Q^{-4} as previously proposed [21], but also on the oscillation frequency or the volume of the active region [17]. This theory qualitatively fits the data of Gagnepain who varied the Q-factor with temperature in the same quartz resonator (but not frequency or volume), the data of Walls who considered several quartz resonators which differ in volume and frequency, and the data of Parker for SAW resonators, with their relatively low Q values.

The theory also provides the basis for predicting from first principles, without adjustable parameters, how to improve the $1/f$ level of resonators, beyond just improving the Q-factor, which has been known for many years, and which has been related [21] to fluctuations in the dissipation. Since the $1/f$ noise level depends on the active volume, in the coherent regime one should use the lowest overtone and smallest diameter consistent with other circuit parameters. In the incoherent (low Q) case the opposite should be considered.

The case of other frequency standards [22]-[25] has been studied by Handel and Walls.

11 A Different Approach to $1/f$ Noise from Frequency Mixing Experiments

As we have seen, the quantum $1/f$ theory provides us with the ontologically most fundamental example of nonlinear system exhibiting a $1/f$ spectrum. Indeed, we have shown how this basic nonlinear system of particle and field derives its nonlinearity from the infrared-divergent coupling between them, each of them being linear in the absence of interaction. It is the reaction of the emitted field back on the source particle, which causes the nonlinearity. Combined with the homogeneity required by dimensional consistency in simple physical systems, this leads to the $1/f$ spectrum as we have already seen in Secs. 2–5. In the special case of high stability and high Q quartz resonators, the whole quartz crystal takes the role of an accelerated "particle" showing a time-dependent current $\dot{\mathbf{P}}$ From the epistemological vantage point offered to us by the sufficient criterion applicable to the special quantum $1/f$ case, we recognize that both the particular distribution of photon frequencies and their amplitudes contribute to the fundamental $1/f$ spectrum. We also realized in Sec. 5 that this spectrum is just caused by the combination of

nonlinearity and homogeneity, independent of the specific system considered. Therefore we believe that the simple numerical consideration of the distribution of frequencies and of their associated amplitudes is sufficient to derive the fundamental $1/f$ spectrum and its magnitude from first principles.

We are dealing with two physical entities that are the "parents" of the $1/f$ noise: the particle and the field. If the charged particle moves uniformly, i.e., the current I is constant, the $1/f$ spectrum contains all beat frequencies between this zero-frequency of the current and the frequencies of the electromagnetic oscillators of the universe which make up the field. This corresponds to coherent quantum $1/f$ effect, but can also correspond to the conventional case if the sample exhibiting $1/f$ noise is very small. If the particle oscillates with the frequency fo, the beat frequencies will generate the $1/\Delta f$ spectrum of sidelines on both sides of the frequency f_0. This is interpreted as quantum frequency fluctuations of the given particle-oscillator. This can transform itself also into $1/f$ phase noise if the particle is part of an electronic system.

In fact, the quantum $1/f$ effect is much more complex than the simplistic 1st order model presented here so far. The shapeless $1/f$ spectrum with no lower limit does not make sense physically in our real world that is shaped by a subtle perfection characterized by perturbed, spoiled or broken symmetries. The actual, physical, $1/f$ spectrum is not infrared divergent and is beautifully shaped by infrared radiative corrections of all orders. It shows a $2\alpha A f^{\alpha A-1}$ dependence of the spectral density of fractional quantum $1/f$ current fluctuations. This comes about through the inclusion of multiphotonic processes of any order, both real and virtual. Therefore, the physical $1/f$ spectrum includes all inter-combination frequencies with the frequency fo of the particle-oscillator. This combines with the right amplitude all harmonics of the frequencies of each of the field oscillators of the universe among themselves and with the particle's frequency f_o.

If the charged particle-oscillator would be allowed to interact only with one of the electromagnetic oscillators of the universe, of frequency f_1 (with well-defined wave vector and polarization), the sideline spectra would be reduced to just two frequencies, $f_0 - f_1$ and $f_0 + f_1$ until we allow for infrared radiative corrections of all orders. This is not a $1/\Delta f$ spectrum, but the dependence of the amplitude of the two sidebands on f_1 is given by $(f_1)^{-3/2}$. [Note: This insures the presence of a $1/\Delta f$ spectrum when all oscillators of the universe are switched on again, because the number of such oscillators is $8\pi f^2 df/3c^2$. The physical result, including infrared radiative corrections, but with only one oscillator allowed to interact, will therefore show a line-spectrum of sidelines at frequencies $f_0 - nf_1$ and $f_o + nf_1$ where n is a natural number. The intensity would decrease strongly from line to line when n increases, because of the factor α^2 which enters into the partial process rate whenever one additional photon is attached to the corresponding Feynman diagram.

If, on the other hand, we allow a current of frequency f_0 and amplitude I_0 to interact in a nonlinear circuit with a current of frequency f_1 and amplitude I_1, we obtain outgoing currents of all frequencies $f_{mn} = |mf_0 \pm nf_1|$[26],[27]. Their amplitudes $I(|mf_0 \pm nf_1|)$ will be given for $m = 1$ and $n = 1$ by the most simple Manley-Rowe relations

$$I(|f_0 \pm f_1|) = [|f_0 \pm f_1|/f_0]^{1/2} I_0 = [|f_0 \pm f_1|/f_1]^{1/2} I_1, \tag{89}$$

Denoting the power $I_0^2/2$ with P_0 and $I_1^2/2$ with P_1, we obtain the Manley-Rowe relations in their original form

$$P(|f_0 \pm f_1|) = [|f_0 \pm f_1|/f_0] P_0 = [|f_0 \pm f_1|/f_1] P_1, \tag{90}$$

where $P(|f_0 \pm f_1|)$ is the power obtained at the frequency $P(|f_0 \pm f_1|)$. Considering, as Planat et al. did [26],[27] , f_0 and f_1 close to each other, and interposing a low-pass filter which rejects the $|f_0 + f_1|$ frequency component, we obtain (90) with the minus sign only. In this case P_1 is negative and a low-frequency output power $P(|f_0 - f_1|)$ is generated. If the generator of power P_1 is replaced by a resistor or any other white noise current generator, of power density P_1', we obtain a noise current output with a low frequency $(f = |fo - f_1|)$ spectral density

$$P(f) = (f/f_0) P_0 = (f/f_1) P_1' d\omega_1 = [(f/f_0) P_0 (f/f_1) P_1' d\omega_1]^{1/2}. \tag{91}$$

By connecting this noise current to the plate of a capacitor C whose other plate is grounded, the noise current gets integrated. This corresponds to a $1/f$ spectrum of the voltage on the capacitor,

$$P''(f) = \frac{1}{C^2 f} [P_0 P_1' d\omega_1 / (f_1 f_0)]^{1/2} \tag{92}$$

provided the rates P_0 and P_1 do not depend on f. This remains to be verified.

Finally, if this $1/f$ noise voltage can be obtained as described here, and if it is applied to the voltage-controlled oscillator which generates the power P_0, its frequency fo will be modulated to display $1/f$ frequency noise. Such a feedback would close the feedback loop indicated earlier. The feedback loop can also be used to provide the integration of the intermediary current noise.

12 Discussion

The derivations of conventional and coherent quantum 1/f noise in Secs. 2 and 3 correspond to different physical situations. These two situations have been discussed on the first page of the 1966 turbulence paper [12], at the beginning of this long journey which led us from the classical hydromagnetic or plasma turbulence to quantum $1/f$ noise and the general sufficient criterion. The discussion of these two situations was repeated identically [10] for the

quantized form of our turbulence theory, i.e., for the two related quantum $1/f$ effects in 1985. It shows us that conventional quantum $1/f$ noise is observed in small samples, for which most of the drift energy of the current carriers is included in the sum of their individual kinetic energies $mv^2/2$. For larger samples, and larger values [20] of the parameter s measuring this proportion numerically, most of the drift energy of the carriers is in their collective magnetic energy $LI^2/2$. The transition between the two situations is given by a physical interpolation formula [12], and is the focus a present research effort discussed by Handel and Zhang [20].

Our criterion shows how homogeneity provides the ingredient leading from nonlinearity to $1/f$ noise. Physically, the homogeneity is required both by the physical requirement of dimensional homogeneity of terms in the equations of physics, and by the invariance of the three-dimensional space with respect to rotations, i.e., by the isotropy of space, which requires x_1, x_2, and x_3 to enter in the same way into the basic laws of nature. In general, we conclude that the ubiquity of the $1/f$ spectrum is caused by the omnipresence of nonlinearities, no matter how small, and by the simultaneous requirement of rotational and Lorentz invariance which shape the world of classical and relativistic physics respectively. In general, we conclude that ontologically, i.e., from the construction of our world with quarks and leptons, quantum $1/f$ noise theory gives the cause of fundamental $1/f$ noise, while epistemologically, i.e., in the world of general notions, the combination of nonlinearity and homogeneity required by our general sufficient criterion is the ultimate cause of all fundamental $1/f$ noise, including the ontologically primordial quantum $1/f$ noise as a special case. Mathematically, this happens in all fundamental $1/f$ spectra on the basis of the idempotence of $1/f^{1-\epsilon}$ with respect to convolutions in the limit $\epsilon \sim 0$, with $\epsilon = \alpha A$ in the case of quantum $1/f$ noise. In practice, however, the idempotent property of $1/f$ does not allow us to distinguish which systems will show $1/f$ fluctuations, while the general sufficient criterion, first presented at the Symposium on $1/f$ Noise and Chaos In Tokyo, March 1991, allows us to easily recognize the systems which generate $1/f$ spectra, if their mathematical definition is given in terms of a dynamical system of nonlinear integro-differential equations, or in simpler terms.

We note that our sufficient criterion explains the ubiquity of $1/f$ noise through a homogeneity which can be established sometimes even without knowing the exact form of the dynamical equation(s) governing a nonlinear system. The derivation of the criterion shows that it is obviously connected with (actually based on) the idempotent property of the $1/f$ spectrum with respect to the convolution operation. Therefore the $1/f$ spectrum corresponds to an accumulation point in Hilbert space, as was first demonstrated [12] directly in 1966 and 1971. Due to the divergence of the integral of $1/f$ at $f = 0$, this author reformulated this accumulation point property in dimensional analysis terms before submitting his paper [19] for publication in 1980; in this form, the argument is more elegant and avoids the divergence at $f = 0$.

The author acknowledges the support of the Air Force Office of Scientific Research, of Michael Grace and of the National Science Foundation.

References

1. Johnson J. B. (1925) The Schottky effect in low frequency circuits. Phys Rev 26:71-85
2. Press W. H. (1978) Comments astrophysics. Space Physics 7:103
3. Handel P. H. (1975) $1/f$ noise - an "infrared" phenomenon'. Phys. Rev. Lett. 34:1492-1494 (1975); Nature of $1/f$ noise. Phys. Rev. Lett. 34:1495-1498
4. Handel P. H. (1980) Quantum Approach to $1/f$ Noise. Phys. Rev. 22A: 745
5. Handel P. H. and Wolf D.(1982) Characteristic Functional of Quantum $1/f$ Noise. Phys. Rev. A26: 3727-30
6. Handel P. H. (1989) The Quantum $1/f$ Effect and the General Nature of $1/f$ Noise. Archiv für Elektronik und Übertragungstechnik (AEÜ) 43:261-270
7. Handel P. H. (1982) Starting Points of the Quantum $1/f$ Noise Approach, Submitted to Physical Review B. Fundamental Quantum $1/f$ Fluctuation of Physical Cross Sections and Process Rates, submitted to Phys. Rev. Letters.
8. Sherif T. S. and Handel P. H.(1982) Unified Treatment of Diffraction and $1/f$ Noise. Phys. Rev. A26:596-602
9. van der Ziel A. (1988) Unified Presentation of 1/f Noise in Electronic Devices; Fundamental 1/f Noise Sources (Review paper). Proc. IEEE 76: 233-258; The Experimental Verification of Handel's Expressions for the Hooge Parameter, Solid-State Electronics 31:1205-1209 (1988) Semiclassical Derivation of Handel's Expression for the Hooge Parameter. J. Appl. Phys. 63:2456-2455 (1988); 64:903-906 (1988); Birbas A.N. et al., J. Appl. Phys. 64:907-912 (1988); van der Ziel A. et. al., Extensions of Handel's 1/f Noise Equations and their Semiclassical Theory, Phys. Rev. B 40:1806-1809 (1989); Tacano M., Proc. XI. Int. Conf. on Noise in Physical Systems and 1/f Fluctuations, Musha T., Sato S. and Mitsuaki Y. Editors, Ohmsha Publ. Co., Tokyo 1991, pp. 167-170; M. Tacano, Proc. Fifth van der Ziel Conference Quantum 1/f Noise and other Low Frequency Fluctuations, AIP Conference proceedings Vol. 282, Handel P. H. and Chung A. L. Editors, 1992; see also Quantum 1/f Bibliography by Handel P. H., unpublished.
10. Handel P. H. (1983), Any Particle Represented by a Coherent State Exhibits 1/f Noise in Noise in Physical Systems and 1/f Noise, (Proceedings of the VIIth International Conference on Noise in Physical Systems and 1/f Noise) edited by Savelli M.,Lecoy G. and Nougier J. P. (North - Holland, Amsterdam), p. 97.
11. Handel P.H. (1986), Coherent States Quantum 1/f Noise and the Quantum 1/f Effect in Noise in Physical Systems and 1/f Noise (Proceedings of the VIIIth International Conference on Noise in Physical Systems and 1/f Noise) edited by D'Amico A. and Mazzetti P., Elsevier, New York, p. 469.
12. Handel P. H. (1966), Instabilities, Turbulence and Flicker-Noise in Semiconductors I, II and III, Zeitschrift für Naturforschung 21a: 561-593 ; Handel P. H. (1971) Turbulence Theory for the Current Carriers in Solids and a Theory of 1/f Noise, Phys. Rev. A3:2066.
13. Zwanziger D. (1973) Phys. Rev. D7: 1082 ; Phys. Rev. Lett. 30:934; Phys. Rev. D11:3481 and 3504 (1975); T.W.B. Kibble, Phys. Rev. 173:1527; 174:1882; 175:1624 (1968); J. Math. Phys. 9: 315 (1968).

14. Zhang Y. and Handel P.H. (1992), Proc. Fifth van der Ziel Conference. Quantum 1/f Noise and other Low Frequency Fluctuations. AIP Conference proceedings Vol. 282, P.H. Handel and A.L. Chung Editors.
15. Gradshteiyn J.S. and Ryzhik I.M. (1965) Table of Integrals, Series and Products Sec. 3.761, No. 9 and No. 7, Academic Press, New York .
16. Handel P. H. (1991), Proc. XI. Int. Conf. on Noise in Physical Systems and 1/f Fluctuations, Musha T., Sato S. and Mitsuaki Y. Editors, Ohmsha Publ. Co., Tokyo, pp. 151-157.
17. Walls F.L., Handel P.H., Besson R. and Gagnepain J.J. (1992), A New Model relating Resonator Volume to 1/f Noise in BAW Quartz Resonators, Proc. 46. Annual Frequency Control Symposium, pp.327-333.
18. Musha T., Borbely G. and Shoji M. (1990), Phys. Rev. lett. 64:2394.
19. Handel P.H., Sherif T., van der Ziel A., van Vliet K.M. and Chenette E.R.: Towards a More General Understanding of 1/f Noise, submitted to Physics Letters.
20. Handel P.H. and Zhang Y. (1993), Proc. XII. Int. Conf. on Noise in Physical Systems and 1/f Fluctuations, Handel P., and Chung A. Editors, Amer. Inst. of Phys. Conf. Proceedings Vol. 285, St. Louis, pp. 172-175.
21. Handel P.H. (1979), Nature of 1/f Frequency Fluctuations in Quartz Crystal Resonators, Solid State Electron. 22:875-876.
22. Hellwig H. (1993), A Look Into the Crystal Ball, the Next 25 Years in 25th Annual Precise Time and Time Interval Applications and Planning Meeting, NASA Conference Publ. 3267.
23. Hellwig W. (1975), Atomic Frequency Standards: a Survey, Proc. of IEEE 63:212-229.
24. Walls F.L. and Hellwig H. (1976), A New Kind of Passively Operating Hydrogen Frequency Standard, Proc. 30th Annual Frequency Control Symposium, Ft. Monmouth, NJ, June 1976, pp. 473-480.
25. Handel P.H. and Walls F.L. (1994): Analysis of Quantum 1/f Effects in Frequency Standards, Proc. 1994 IEEE Frequency Control Symp., Boston, MA, L. Maleki Editor, IEEE Press, IEEE Catalog No. 94CH3446-2, pp. 539-540.
26. Planat M., Dos Santos S., Ratier N., Cresson J. and Perrine S. (1988), Close to Resonance Interaction of Radiofrequency Waves in a Schottky diode Mixer, AIP Proc. Vol. 466, VII. van der Ziel Symp. on Quantum 1/f Noise and Other Low Frequency Fluctuations in Electronic Devices, Aug. 7-8, 1998, Univ. of Missouri, St. Louis, MO, P.H. Handel and A.L. Chung, Editors, pp. 177-187.
27. Planat M. and Eckert C. (2000) On the frequency and amplitude spectrum and the fluctuations at the output of a communication receiver. IEEE Trans. Ultrason. Ferroelect. Freq. Contr. (to appear in the Special Issue on Frequency Control and Precision Timing)

$1/f$ Frequency Noise in a Communication Receiver and the Riemann Hypothesis

Michel Planat *

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, 32 Avenue de l'Observatoire, 25044 Besançon Cedex, France

Abstract. A mixer cascaded with a low pass filter is the central element of any communication receiver. It is also used to register minute frequency fluctuations of an external oscillator (RF) under test versus the frequency of a local oscillator (LO). Such a scheme may also be viewed as the basic model of an electronic oscillator, with the amplifier noise at the RF input and the resonator signal at the LO input. We have investigated experimentally the whole spectrum of frequencies and amplitudes of beat signals and their frequency fluctuations at the IF output of the mixer + filter set-up. We have found evidence that all the dynamics follows arithmetical rules. The frequency of the beat signal is defined from a diophantine approximation of the frequency ratio of input oscillators; the amplitude is defined globally from the position of resolved fractions with respect to the equally spaced graduation; and for the frequency fluctuations, a transition from white frequency noise to $1/f$ frequency noise is observed close to resonance. This is explained on the basis of number theory in relation to the Riemann problem concerning the distribution of prime numbers. More precisely it is shown that diophantine signal processing is at work in the receiver and that this may be understood from the Littlewood and Franel–Landau formulation of the Riemann hypothesis.

1 The Communication Receiver

1.1 Theoretical background

Although it may have a quite complex architecture, the basic piece of a communication receiver is close to the one proposed by Amstrong for broadcast reception. That technique revives today in wireless applications. The receiver is designed to compare the information carying external oscillator (called RF) to a local oscillator (called LO) of about the same high frequency through a nonlinear mixing element. For narrow band FM demodulation (which is typical for broadcasting and wireless analogic applications) one uses a discriminator of which the role is first to differentiate the signal, that is convert frequency modulation (FM) to amplitude modulation (AM) and second to detect its low frequency envelope: this is called baseband filtering. For more general wideband or narrowband FM demodulation one uses a low frequency filter instead of the discriminator to remove the high frequency signals generated after the mixer.

* planat@lpmo.univ-fcomte.fr

M. Planat (Ed.): LNP 550, pp. 265–287, 2000.
© Springer-Verlag Berlin Heidelberg 2000

In the closed loop operation a voltage controlled LO is used to track the frequency of the RF (hence the generic words: FMFB for a FM feedback loop or PLL for a phase locked loop). Phase modulation is becoming most frequently used for digital signals because low bit error rates can be obtained for relatively poor signal to noise ratio, in comparison to frequency modulation [1].

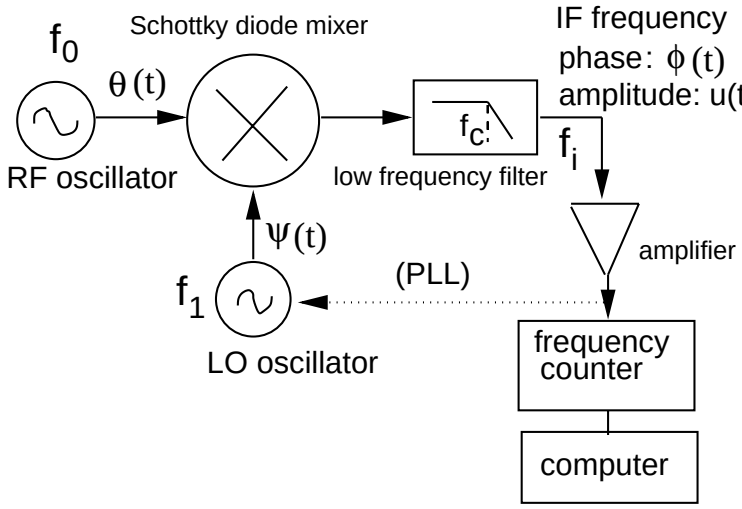

Fig. 1. Schematic of the electronic set-up used in the experiment

In this paper we will consider that the basic piece of the receiver includes a multiplier (or mixer) cascaded with a lowpass filter (in the closed loop operation, this is the PLL). Such a set up (Fig. 1) is widely used in frequency metrology to register minute frequency fluctuations of an external RF oscillator under test versus the frequency of a local LO oscillator. Mixing operation results from the frequency conversion in the time varying conductance of the diode, or the transconductance from gate to drain in a field effect transistor. Here we use Schottky barrier diodes in a doubly balanced structure to produce good isolation between the inputs [2].

The ideal mixer multiplies the RF signal of frequency f_0 to be received by the reference sinusoid of frequency f_1 of the LO shifting in both to the sum frequency and the difference frequency $f_0 \pm f_1$. The down conversion at the intermediate frequency (called IF) is desired so that a low pass filter of cut-off frequency f_c is included. In such a configuration the set mixer + filter essentially behaves as a phase detector, that is the instantaneous voltage at the output is the sine of the phase difference at the inputs:

$$u(t) = a \ \sin(\theta(t) - \psi(t)), \tag{1}$$

where $\theta(t)$ and $\psi(t)$ are the instantaneous phases of the RF and LO oscillators and a (in Volts per radian) is the phase detector sensitivity. The non linear dynamics of the set-up in the closed loop configuration is well described by introducing the phase difference $\phi(t) = \theta(t) - \psi(t)$. Using $\dot{\theta} = \omega_0$ and $\dot{\psi}(t) = \omega_1 + A\ u(t)$ with $\omega_0 = 2\pi\ f_0$, $\omega_1 = 2\pi\ f_1$ and A (in rad. Hz/Volt) the sensitivity of the voltage controlled oscillator (called VCO) one obtains the non linear differential eq.

$$\dot{\phi}(t) + K\ \sin(\phi(t)) = \omega_i, \tag{2}$$

with $\omega_i = \omega_0 - \omega_1$ as the bare frequency shift between the input oscillators and the frequency $K = aA$ as the coupling coefficient (also called open loop gain). Equation (2) is integrable but its solution looks complex [4]. If the frequency shift ω_i does not exceed the open loop gain K, the average frequency$\langle\dot{\phi}\rangle$ vanishes after a finite time to reach the stable steady state $\phi(\infty) = 2l\pi + \sin^{-1}(\omega_i/K)$, l integer. In this phase tracking range of width $2K$ the RF and LO oscillators are also frequency locked. Outside the modelocking zone there is a sech shape beat signal of frequency

$$\omega_B = \langle\dot{\phi}(t)\rangle = K(u^2 - 1)^{1/2} \quad \text{that is } \omega_B = (\omega_i^2 - K^2)^{1/2}, \tag{3}$$

with $u = \omega_i/K$ converging to the open loop frequency ω_i as $u \ll 1$.

Due to the nonlinearity of the mixer the practical operation of the phase detector also involves the harmonic interactions

$$u(t) = \sum_{p,q} a^{(p,q)}\ \sin(q\theta(t) - p\psi(t)), \tag{4}$$

leading to the intermodulation frequencies

$$f_i^{(p,q)} = \mid pf_0 - qf_1 \mid \quad \text{with } p \text{ and } q, \text{ integers} \tag{5}$$

in the bandwidth of the filter [4]. Standard models account for non linear interactions resulting from the LO excitation only, assuming the RF excitation is negligible. We found in section 2 an alternative way of thinking the spectrum of frequency mixings based on number theory and specifically on continued fraction expansions (CFE) of frequency ratios of signals at the input of the mixer.

From this form we can expect the device will produce a rational approximation of the frequency ratio $\nu = f_1/f_0$. We will show below that it behaves as a diophantine approximator. This is in contradiction with the conventional view of filtering that we remind from now. In such a linear approach the loop filter is represented from the Laplace transform $H_2(P)(P = \frac{\mathrm{d}}{\mathrm{dt}})$ so that instead of (1) we get

$$u(t) = a\ H_2(P)\ \sin\ \phi(t), \tag{6}$$

and the nonlinear dynamics is described from

$$\dot{\phi}(t) + K\ H_2(P)\ \sin\ \phi(t) = \omega_i. \tag{7}$$

The loop filter is typically a low pass filter of overall transfer function

$$H_2(P) = \frac{1 + \frac{P}{A}}{1 + \frac{P}{B}}, \tag{8}$$

leading to the describing equation

$$\ddot{\phi}(t) + \dot{\phi}(t)(B + KB/A\ \cos\phi(t)) + BK\ \sin\phi(t) = B\omega_i. \tag{9}$$

The linear description of the PLL leads to the closed loop transfer function [1]

$$H(P) = \frac{1 + \dfrac{P}{a_0\omega_n}}{(\dfrac{P}{\omega_n})^2 + 2\eta\dfrac{P}{\omega_n} + 1}, \tag{10}$$

with $\omega_n = (KB)^{1/2}$, $\eta = 1/2(B/K)^{1/2} + (KB)^{1/2}/2A$ and $a_0 = A/(K)B^{1/2}$. This closed loop transfer function is similar to the one in a FMFB loop. There is a complex pole pair of natural frequency ω_n, a damping factor η and a real zero at $P = -a_0\omega_n$.

The nonlinear operation is much more complicated allowing for the so-called cycle skipping phenomenon, i.e. the trajectory in phase space may leave the main attractive region through successive cycles of 2π phase error. This beat signal looks similar to the one described in (3) when the initial frequency offset is outside the phase-locked range K. As shown below we found experimentally a $1/f$ frequency spectrum associated to that phenomenon [4]. It represents one of the main motivations for the arithmetical approach.

A full understanding of the cycle skipping phenomenon and the associated $1/f$ frequency spectrum needs an account of the intermodulation products which are present at the output of the receiver. Using (4) instead of (1), the PLL (7) is generalized as:

$$\dot{\phi}^{(p,q)}(t) + q\ H_2(P)\ \sum_{r,s} K^{(r,s)} \sin\ \phi^{(r,s)} = \omega_i^{(p,q)}, \tag{11}$$

where $\omega_i^{(p,q)} = p\omega_0 - q\omega_1$, and $K(r,s)$ is the effective gain at the harmonic (r,s) and $H_2(P)$ is the open loop gain as in (8).

Equation (11) can be rewritten for two individual modes (p,q) and (m,n) and the sum over (r,s) removed. After integration we obtain:

$$\dot{\phi}^{(p,q)}(t) + q\ H_2(P)\ \sum_{r,s} K^{(r,s)} \sin(\frac{s}{q}\ \phi^{(p,q)} - \frac{\omega_0 t}{q}(qr - ps) + \phi_0^{(r,s)}) = \omega_i^{(p,q)}, \tag{12}$$

where $\phi_0^{(r,s)}$ is a reference phase at harmonic (r,s). We didn't undertake the task to study the differential equation (12) in detail because the arithmetical approach below will be far more efficient. However some hints concerning the non linear interactions are given now. Let us first consider the case of a first order filter $H_2(P) = 1$. If one neglects the interaction of one product with the other so that $\frac{p}{q} = \frac{r}{s}$, one recovers an equation similar to (7) at each harmonic. In the general case it is observed from (12) that the RF signal acts as periodic perturbation for the standard PLL loop. In order to illustrate the net effect of introducing such perturbations, we consider the realistic case of the interaction of the fundamental $(p, q) = (1, 1)$ with the main odd subharmonic at $(r, s) = (1, 3)$. In such a case (12) simplifies as:

$$\dot{\phi(t)} + K\ \sin(\phi(t)) + K^{(3)} \sin(3\phi(t) + 2\omega_0 t) = \omega_i, \tag{13}$$

where we used $\phi = \phi^{(1,1)}$ and $K = K^{(1,1)}$ and $K^{(3)} = K^{(1,3)}$. In ref. [3] we described an asymptotic perturbation analysis for deriving mode-locked zones in (13). If the condition $\dot{\phi}(t) = C^{\text{te}}$ holds over each cycle provided the counting time is large enough in comparison to the period, we can get a recurrence formula mapping the phase ϕ_{n+1} at time $t + 2\pi/\omega_0$ to that ϕ_n at time t with the following form:

$$\phi_{n+1} = \phi_n + 2\pi\omega - c\ \sin\phi_n - d\ \sin(3\phi_n), \tag{14}$$

with $\omega = \omega_i/\omega_0$, $c = 2\pi K/\omega_0$ and $d = 2\pi K^{(3)}/\omega_0$. Eq. (14) extends the well known Arnold map.

Such an equation is studied by introducing the winding number $\nu_{\text{av}} = \lim_{n\leftarrow\infty}(\phi_n - \phi_0)/(2\pi n)$. The curve ν_{av} versus ω is a devil's staircase with steps attached to rational values of $\omega = p/q$ and their width increasing with the coupling coefficient c. In ref. [5] we presented an approach to plot the mode-locked zones (called Arnold tongues) based on the introduction of an Hölder exponent. In the present case $d \ll c \ll 1$ so that we stay very far from the critical value c=1 where the tongues may recover and chaotic states may be present. This means that the cycle skipping phenomenon and the $1/f$ frequency noise are not related to the type of chaos predicted from the Arnold map. If we turn to the second order filtering effect in (8), we get instead of (14) a two dimensional map extending the well known standard map [6].

1.2 Experiments

We derived above a non linear differential model (12) taking into account the full intermodulation spectrum at the output of the demodulator. We now present the experimental motivation for that work and the reasons why it should be further improved to get a full consideration of the frequency and amplitude spectra and of the associated frequency fluctuations. All the

experiments below are given for the PLL in the open loop configuration. They can be regarded as an asymptotic low coupling limit for (12). We used a $f_0 = 10$ MHz RF oscillator and a LO oscillator of frequency varying from $f_1 =$ 100 kHz up to $f_1 = 10$ MHz, interacting through a wideband doubly balanced Schottky diode mixer (Fig. 1). The low frequency filter can be approximated to a second order type one with $f_n \sim \omega_n/2\pi = 375$ kHz. It is observed in Fig. 2 that there are many subharmonic modes $\nu = p/q$ in addition to the fundamental one at $\nu = 1/1$. The main ones correspond to p and q odd; intermodulation products at p or q even are strongly rejected due to the doubly balanced phase bridge structure of the mixer. Each mode is resonant and extends over a basin which is controlled from the low frequency cut off of the loop filter, the full dynamics of the subharmonics and the thermal noise of the input oscillators. In order to get valuable insights into these seemingly complicated dynamics, we need to derive an arithmetical approach of the filtering process. The transfer function $H_2(P)$ derived in (8) is clearly insufficient to describe the experimental plots. Instead of this a diophantine approach of the filtering process will be presented below.

We now turn to records of frequency fluctuations versus time. These experiments are performed using period counting of the beat frequency f_i over an integration time τ. The low frequency beat signal opens a gate of the reciprocal counter at one of its zero crossings. The register counts the period of both the high frequency counter clock signal and the beat signal. The sampling duration τ determines the integral number of cycles of the counter oscillator accumulated before the gate closes. The number of cycles of the beat signal accumulated during this time interval is finally divided by τ to yield the average frequency $f_i(\tau)$. The average beat frequency being continuously monitored, the dead time, that is the time elapsed between the end of one recording of f_i and the beginning of the next, is zero. By this way the outcome is a collection of $f_i^{(l)}(\tau)$ where l stands for the l^{th} measurement of the averaged beat frequency $f_i(\tau)$. The type of frequency noise being present may be characterized by introducing the Allan deviation which is defined as the mean squared value of the relative frequency deviation between adjacent samples of length τ . Using $y^{(l)}(\tau) = \frac{f_i^{(l)}(\tau)}{f_i}$ it may be rewritten as [7], [8]:

$$\sigma_y^2(\tau) = \frac{1}{2}\left\langle (y^{(l+1)}(\tau) - y^{(l)}(\tau))^2 \right\rangle, \tag{15}$$

where $l = (1, 2, \ldots, N)$ stands for the N measurements and $\langle\ \rangle$ means the averaging. This method is widely used to characterize the frequency stability of a test oscillator of frequency $f_0(t)$ against a highly stable reference oscillator of constant frequency f_1 .Using $f_i = f_0(t) - f_1$ at the fundamental mode, we get a relative frequency stability defined as $\frac{\delta f_0}{f_0} = \frac{\delta f_i}{f_i} . \frac{f_i}{f_0} \sim \sigma_y(\tau)$, and the sensitivity f_0/f_i of the frequency measurement set-up may be choosen very large. In the subsequent measurements it is shown that the frequency stability of the test oscillator (the frequency synthesizer) is of the order 10^{-11}.

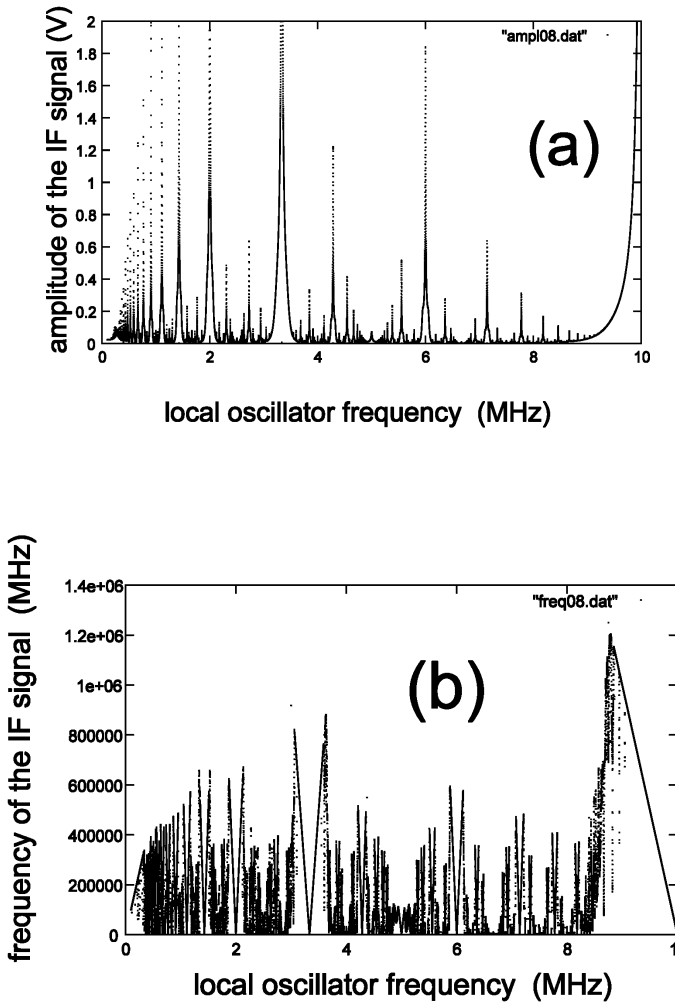

Fig. 2. The amplitude spectrum (top) and the frequency spectrum (bottom)

The graph of the Allan variance versus the sampling time generally obeys power laws τ^{-m} (m integer) over a restricted range of time scales. A dual measure of frequency noise is the power law which is followed by the power spectral density of frequency fluctuations $S_y(f) = f^p$ (p integer). In the hypothesis of stationarity, the time and frequency domains are related with a simple law $m = p + 1$ (if $-3 < p \leq 1$) and $m = 2$ (if $p \geq 1$) [7]. Therefore white frequency noise $p = 0$ leads to the Allan deviation $\sigma_y(\tau) \sim \tau^{-1/2}$ and the flicker noise $p = -1$ leads to a flicker floor $m = 0$ in the Allan deviation. In cases when m is not integer, the frequency and time domains may also be related using the Fourier transform, but with less confidence in their meaning.

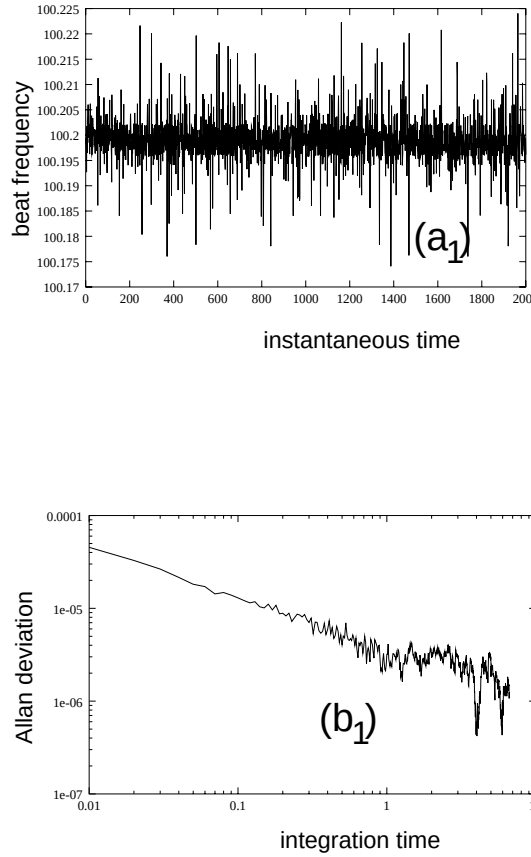

Fig. 3. Instantaneous frequency fluctuations of the beat signal between two input oscillators at $f_1 = 5.0206$ MHz and $f_0 = f_1 + 100$ Hz versus time in units a number of periods (a_1). The corresponding Allan deviation (b_1) corresponds to white frequency noise

For the experiments we used a highly stable LO oscillator with frequency $f_1 = 5.0206$ MHz. The beat frequency fluctuations $f_i^{(l)}(\tau)$ and the associated Allan deviation $\sigma_y(\tau)$ observed at the fundamental mode $p/q = 1/1$ in three situations. Instantaneous period measurements could be performed in the three cases due to the low frequency of beat signals. The first case $f_i \sim 100$ Hz shown in Fig. 3 lies away from the resonance and the corresponding Allan deviation scales as $\sigma_y(\tau) \sim \tau^{-1/2}$ which is typical of white frequency noise. The second case $f_i \sim 4.4$ Hz is closer to resonance and the observed Allan deviation Fig. 4 is slowly divergent with scaling law $\sigma_y(\tau) \sim \tau^{1/4}$. Time fluctuations show frequency jumps of magnitude 1 mHz and 2 mHz which are typical of a random telegraph signal (RTS). Finally if the two input oscillators have $f_i \sim 0.5$ Hz, we found that the RTS converts into a more

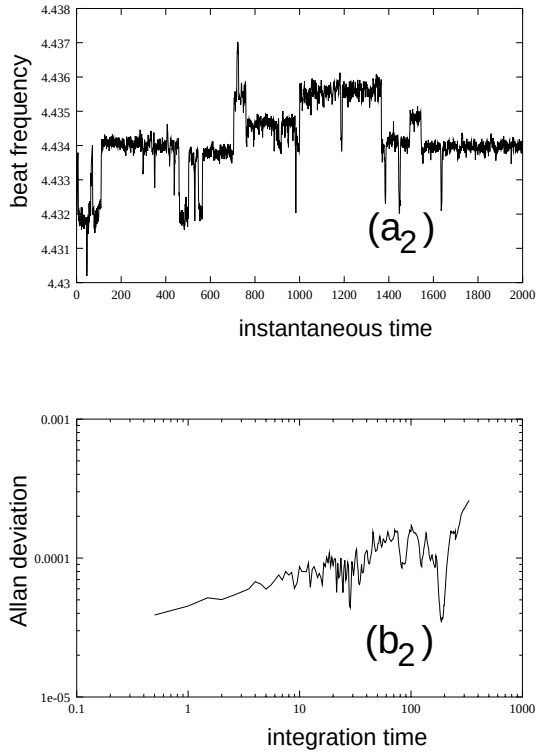

Fig. 4. Instantaneous frequency fluctuations of the beat signal between two input oscillators at $f_1 = 5.0206$ MHz and $f_0 = f_1 + 4.43$ Hz versus time in units a number of periods (a_2). This regime shows a RTS type noise. The corresponding Allan deviation Allan deviation diverges approximately as $\tau^{1/4}$ (b_2)

irregular time dynamics with an approximately constant Allan deviation as shown in Fig. 5, which is typical of a pure $1/f$ noise.

Amplitude fluctuations of the beat signals have been recorded in addition to frequency fluctuations. It has been observed that the conversion from the RTS type signal to the $1/f$ type signal near the resonance is associated to unpredictible amplitude jumps of the shot noise type. The low frequency amplitude and frequency noises were found recently in bipolar transistors; they were found of $1/f$ type and correlated to each other [9]. These features will be explained later on the basis of the arithmetical approach.

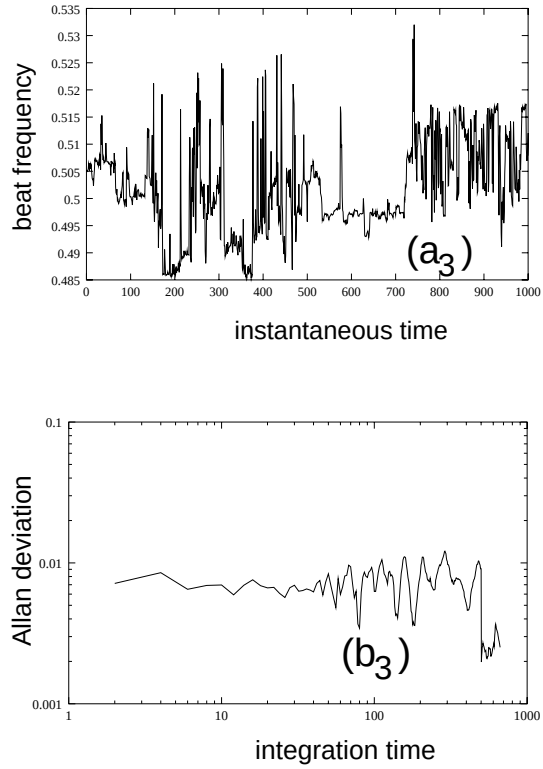

Fig. 5. Instantaneous frequency fluctuations of the beat signal between two input oscillators at $f_1 = 5.0206$ MHz and $f_0 = f_1 + 0.5$ Hz versus time in units a number of periods (a_3). The corresponding Allan deviation (b_3) corresponds to $1/f$ frequency noise

2 Arithmetic of Amplitude–Frequency Relationships

2.1 The frequency of beat signals from a diophantine approximation

As shown in Sect. 1, the practical operation of the communication receiver always involves many downconverted products at all intermediate frequencies in the bandwidth of the filter. We show that the understanding of the frequency spectrum of beat signals can be based on number theory and specifically on continued fraction expansions (CFE) of frequency ratios of signals at the input of the mixer. Let $\nu = f_1/f_0$ the LO to RF frequency ratio and $\mu = f_i/f_0$ the IF to RF frequency ratio, the three frequencies operation may be rewritten as:

$$\mu^{(p_i,q_i)} = q_i \mid \nu - \frac{p_i}{q_i} \mid . \tag{16}$$

From this form we can guess the device will produce a rational approximation p_i/q_i of the real number ν allowed by the physical constraints, in particular the non linear interaction and the synchronization between the frequencies which take place in the mixer, the properties of the filter, the finite resolution of countings and the unavoidable thermal noise which is present in the set-up. One clearly shows that the approximation is not a decimal one, otherwise we should not understand the hierarchy of interacting basins at intermodulation ratios p_i/q_i. The approximation which is performed is the best one possible accounting for the finite resolution, it is of a diophantine type. Best rational approximations of a real number ν are given from CFE, that is:

$$\nu = a_0 + 1/\{a_1 + 1/\{a_2 + ... + 1/\{a_i + ...\}\}\}; \tag{17}$$

and the a_i's are positive integers (called partial quotients) which are readily obtained from the formulas: $a_0 = [\nu], \alpha_0 = \{\nu\}$ and if $i \geq 1$, $a_i = [1/\alpha_{i-1}], \alpha_i = \{1/\alpha_{i-1}\}$, where $[\nu]$ denotes the integral part of ν and $\{\nu\} = \nu - [\nu]$ is its fractional part. Successive best approximants of ν, that is convergents p_i/q_i are obtained by truncating Eq. (17) at some stage i.

Let us remind the difference between the decimal and the diophantine approximation. From the decimal approximation one builds a series p_i/q_i which verifies the approximation $\left|\nu - \frac{p_i}{q_i}\right| \leq \frac{1}{q_i}$. On the other hand the CFE lead to convergents p_i/q_i which satisfy the far better approximation

$$\left|\nu - \frac{p_i}{q_i}\right| \leq \frac{1}{q_i q_{i+1}} \leq \frac{1}{a_{i+1} q_i^2}. \tag{18}$$

The approximation constant $C(\nu) = \lim_{i\to\infty} q_i^2 \left|\nu - \frac{p_i}{q_i}\right|$ is known as the Markoff constant [10]. It can be shown that the worst numbers (in terms of the diophantine approximation) are $\nu = (\sqrt{5} - 1)/2$, $\nu = \sqrt{2} - 1$, $\nu = (\sqrt{221} - 11)/10$,... They have periodic CFE which are given respectively by $\nu = [0; \bar{1}]$, $\nu = [0; \bar{2}]$, $\nu = [0; \overline{2, 1, 1, 2}]$... corresponding to Markoff constants $C = 1/\sqrt{5}$, $C = 1/(2\sqrt{2})$, $C = 5/\sqrt{221}$...respectively. The symbol $[.; ...]$ means the CFE as in (17) and the bar shows the period of the expansion. They can be predicted by studying the minima of binary quadratic forms. There is also much geometry behind [10].

From (18) one expects that whenever a large partial quotient is reached, the diophantine approximation becomes very efficient and the CFE is truncated at level i. Let us see the effect of restricting CFE at

$$a_i \leq a_{\max}. \tag{19}$$

Any rational number p_i/q_i has two CFE given by $p_i/q_i = [a_0; a_1, a_2..., a_i]$ and $p_i/q_i = [a_0; a_1, a_2..., a_i - 1, 1]$. Let us consider truncation at a given value $a_{\max}$ and numbers ν which are very close to p_i/q_i on the right (respectively on the left); they should be approximated as $p_{i+1}/q_{i+1} = [a_0; a_1, a_2, ...a_i, a_{i+1}]$

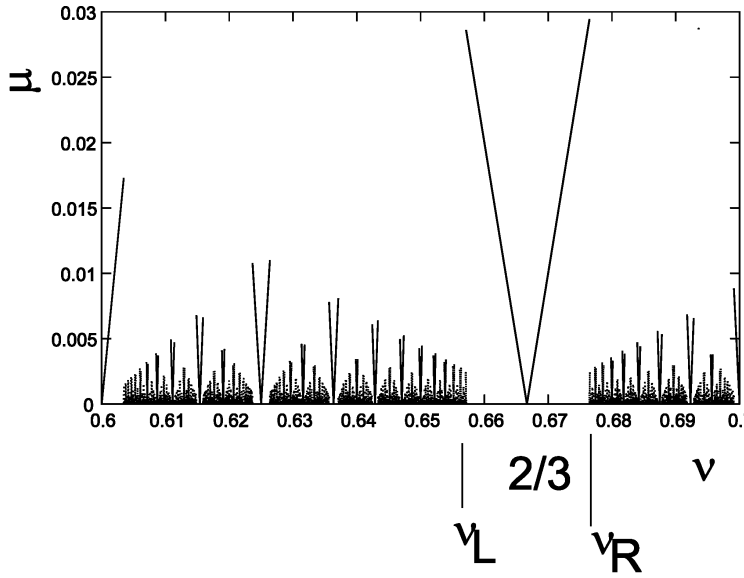

Fig. 6. The intermodulation spectrum resulting from the truncation at $a_{\max} = 10$ in the continued fraction expansion of ν. The larger basin in the window $0.6 < \nu < 0.7$ is at the rational $p_i/q_i = 2/3 = [0; 1, 2]$ with $\nu_R = [0; 1, 2, 10] = 21/31$ and $\nu_L = [0; 1, 1, 1, 10] = 21/32$

(respectively $p_{i+1}/q_{i+1} = [a_0; a_1, a_2, ...a_i - 1, 1, a_{i+1}]$) with $a_{i+1} \geq a_{\max}$. But as a result of the truncation at $a_{\max}$ those numbers will be approximated from the set of two straight lines $\mu = q_i \left|\nu - \frac{p_i}{q_i}\right|$. The last numbers in the basin of p_i/q_i will be at $\nu_R = [a_0; a_1, a_2, ...a_i, a_{\max}]$ and $\nu_L = [a_0; a_1, a_2, ...a_i - 1, 1, a_{\max}]$ given by the formulas:

$$\left|\nu_R - \frac{p_i}{q_i}\right| = \frac{1}{q_i q_{i+1}} \text{ and } \left|\nu_L - \frac{p_i}{q_i}\right| = \frac{1}{q_i(q_i a_{\max} + q_i - q_{i-1})}. \tag{20}$$

It is seen that a significant asymmetry results in most cases. Fig. 6 illustrates the result obtained by taking $a_{\max} = 10$ and the window $0.6 < \nu < 0.7$. In order to check the validity of the above arithmetical approach we registered the intermodulation spectrum for the device in Fig. 1, using a constant frequency RF signal and a variable frequency LO signal. Acquisition of data was performed using a digital counter with an integration time $\tau_0 = 0.1$ s. Results are shown in Fig. 7(a). A good fit of the data is obtained accounting for the physical device: products with p_i or q_i even were strongly rejected due to the doubly balanced phase bridge structure of the diode mixer. The theoretical curve is shown in Fig. 7(b). The filtering rate was found to govern the truncation level of CFE in the loop. Using more selective filters at 58 kHz (and 3 kHz) we found truncations at values $a_{\max} = 9$ (and 65) also allowing a good fit of the experimental curves.

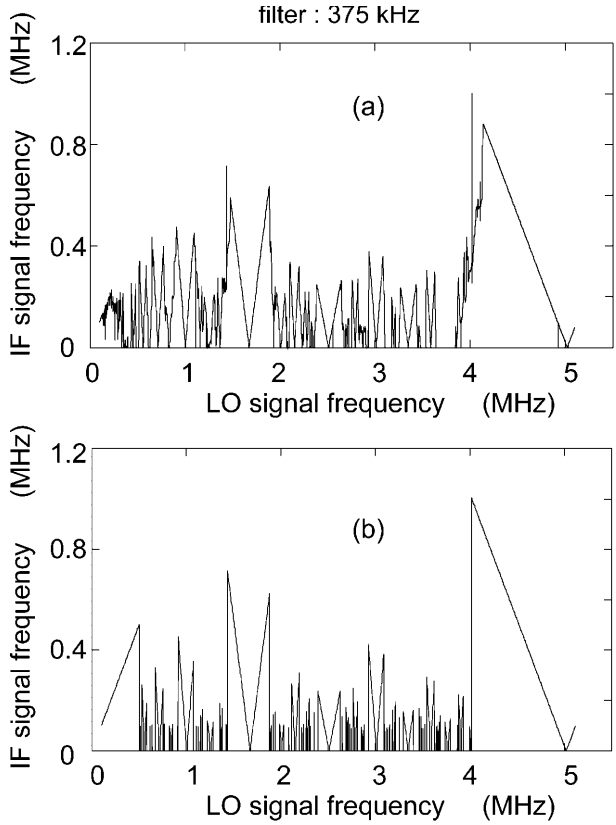

Fig. 7. (a) The intermodulation spectrum for the loop in Fig. 1 and cut-off frequency $f_c = 375$ kHz and the RF constant frequency $f_0 = 5.0206$ MHz. (b) The theoretical spectrum. For the fit we chose $a_{\max} = 3$ for the fundamental mode 1/1, $a_{\max} = 2$ for products p_i/q_i with p_i and q_i odd and $a_{\max} = 20$ for products with p_i or q_i even

We just saw the intermodulation spectrum is described from CFE and a resolution increasing with the degree of filtering. It may also be convenient to represent the error μ versus ν on a tree with the number of leaves given from a Farey criterion

$$q_i \leq q_{\max}. \tag{21}$$

The Farey series $F_{q_{\max}}$ of order $q_{\max}$ is defined as the set of irreductible fractions p_i/q_i between 0/1 and 1/1 whose denominators do not exceed $q_{\max}$. Thus F_5 is $\{0/1, 1/5, 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 1/1\}$. Between two leaves ending on the real axis at p/q and p'/q' there is one at $(p+p')/(q+q')$ provided q, q' and $q + q'$ do not exceed $q_{\max}$. It arises at the node $\nu = (p + p')/(q + q'), \mu = 1/(q + q')$ as shown in Fig. 8. The number of fractions

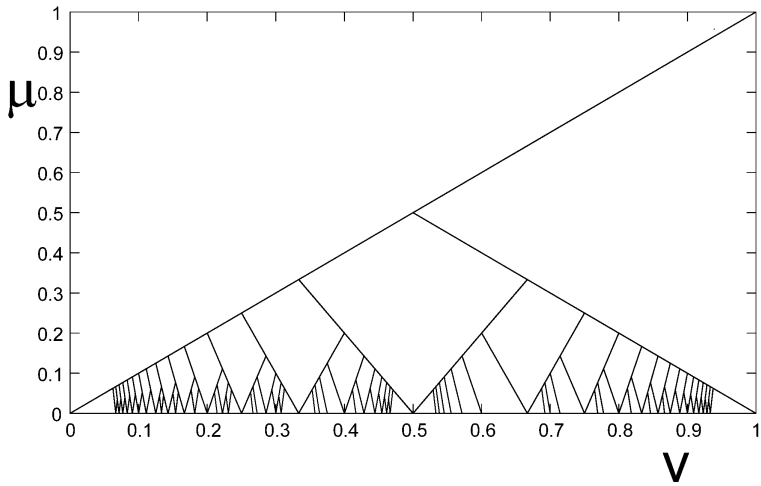

Fig. 8. The increased resolution spectrum as calculated from finite continued fraction expansions of ν from $i_{\max} = 1$ to $i_{\max} = 3$

up to $q_{\max}$ is as follows

$$n(q_{\max}) = \sum_{k=1}^{q_{\max}} \phi(k) = \frac{3\ q_{\max}^2}{\pi^2} + 0(q_{\max} log(q_{\max})), \tag{22}$$

where $\phi(k)$ is the Euler function [5]. Such fractions are called well resolved and n^{-1} is called the resolution. In contrast to criterion (21) a badly resolved experiment is governed by the truncation trick (19) or by truncating CFE at a given height:

$$i \leq i_{\max}. \tag{23}$$

We will show below their effect in producing $1/f$ type noises.

2.2 The amplitude of beat signals and the Franel–Landau shift

As shown in Sect. 2.1 the frequency of beat signals in the demodulator may be interpreted from the diophantine approximation of frequency ratios ν of signals at the input of the mixer. Although this definition accounts for the Farey structure of the underlying tree, it is local and linear: that is nearby frequencies lie on a straight line defined from the truncated CFE of ν; nonlinearity only happens from jumps whenever a partial quotient a_i has been added or removed so that the approximation of ν is controled from a new leave of the tree. Linearity is ensured provided the device operates on a prescribed subharmonic p_i/q_i; this is precisely the property one needs if one uses the demodulator to compare a test oscillator against a reference one.

In contrast to the frequency, the amplitude of beat signals is non linear as shown in Fig. 2b : thus the beat signal is resonant at the subharmonic p_i/q_i, leading to a strong sensitivity of the demodulator to amplitude modulations (AM). It is thus very critical to understand the amplitude of beat signals.

The amplitude $a^{(p_i,q_i)}$ of the beat signal at the subharmonic p_i/q_i can still be defined from number theory. This is achieved by comparing the position of fraction p_i/q_i with respect to the point i/n of the equally spaced scale with n given in (22). The shift is given from:

$$\delta_i = \left| \frac{i}{n} - \frac{p_i}{q_i} \right|, \tag{24}$$

and the corresponding amplitude $a^{(p_i,q_i)} = A\,\delta_i$ where A (in Volts) is a reference voltage depending on the sensitivity of the set-up. The shift δ_i was introduced in 1924 by Franel and Landau in the context of a conjecture equivalent to Riemann's hypothesis [5], [11]. The hypothesis relies on the position of zeros of the function $\zeta(s)$ which was introduced by Riemann in 1859 on his celebrated paper on prime numbers. According to Riemann's hypothesis there are an infinite number of non trivial zeros all located on the axis $s = 1/2$. The conjecture will be set forth in section (2.4).

Fig. 9 (curve 1) shows a plot of the shift δ_i versus $i(i = 1, n(q_{\max}))$ for $q_{\max} = 150$ (and $n(150) \sim \frac{3}{\pi^2} q_{\max}^2 = 6839$) as predicted from (22). An efficient MAPLE procedure was implemented in order to compute Farey fractions and the associated shifts. A computing time increasing like n^2 (30 s for $q_{\max} = 100$) was obtained. It can be compared to the experimental plot in Fig. 2a, if we account for the rejection of even intermodulation products due to the balanced structure of the mixer.

As already mentioned in section (1) and will be further extended in our study of frequency fluctuations in section (2.3) our ability to fit the experimental curves depends on our ability to perturb the above result to account for the lack of resolution in the CFE by truncating it at a selected partial quotient $a_{\max}$ or at a given height $i_{\max}$. To our knowledge no systematic theory of finite CFE has been developed until know. See [12] for preliminary results and the companion paper by J. Cresson. Here we mention some analytical calculations of Franel–Landau shifts for CFE of low height $i_{\max}$ from 1 to 3.

At first order $i_{\max} = 1$ the RF to LO frequency ratio is $\nu = 1/a_1, a_1 \leq n^{(1)}(q_{\max}) = q_{\max}$ so that the shift is as follows:

$$\delta^{(1)}(q_{\max}) = \left| \frac{q_{\max} - a_1 + 1}{q_{\max}} - \frac{1}{a_1} \right|. \tag{25}$$

At second order $i_{\max} = 2$, $\nu = a_2/(1+a_1a_2)$ with a_1 taking all values between a_1 and $q_{\max} - 1$ that is $a_1 = (1, q_{\max} - 1)$ and $a_2 = (1, \left[\frac{q_{\max}-1}{a_1}\right])$. The number

of ill resolved fractions is

$$n^{(2)}(q_{\max}) = 1 + \sum_{a_1=1}^{q_{\max}-1} \left[\frac{q_{\max}-1}{a_1} \right], \tag{26}$$

with the shift as

$$\delta^{(2)}(q_{\max}) = \left| \frac{a_2 + \sum_{k=a_1+1}^{q_{\max}-1} \left[\frac{q_{\max}-1}{k} \right]}{n^{(2)}(q_{\max})} - \frac{a_2}{1+a_1a_2} \right|. \tag{27}$$

The calculation is still more complicated at order $i_{\max} = 3$. In such a case we get

$$n_3(q_{\max}) = q_{\max} + \sum_{a_1=1}^{\left[\frac{q_{\max}-1}{2}\right]} \sum_{a_2=1}^{\left[\frac{q_{\max}-a_1-1}{a_1}\right]} \left[\frac{q_{\max}-a_1}{1+a_1a_2} \right]. \tag{28}$$

There are one term fractions $p_3/q_3 = 1/a_1$ which are located at

$$i = q_{\max} - a_1 + 1 \ \text{ if } \ a_1 = \left(\left[\frac{q_{\max}-1}{2} \right], q_{\max} \right), \tag{29}$$

or at the position

$$i = n^{(3)}(q_{\max}) - a_1 + 1 - \sum_{k=1}^{k=a_1-1} \sum_{l=1}^{\left[\frac{q_{\max}-1-k}{k}\right]} \left[\frac{q_{\max}-k}{1+kl} \right]$$
$$\text{if } \ a_1 = \left(1, \left[\frac{q_{\max}-1}{2} \right] \right), \tag{30}$$

and there are fractions with three terms $\frac{p_3}{q_3} = \frac{1+a_2a_3}{a_1+a_3+a_1a_2a_3}$ which are located at the position

$$i = n^{(3)}(q_{\max}) - (a_1+a_3) - \sum_{l=1}^{a_2} \left[\frac{q_{\max}-a_1}{1+a_1l} \right] - \sum_{k=1}^{a_1} \sum_{l=1}^{\left[\frac{q_{\max}-1-k}{k}\right]} \left[\frac{q_{\max}-k}{1+kl} \right] \tag{31}$$

with the following range of values

$$a_1 = \left(1, \left[\frac{q_{\max}-1}{2} \right] \right), \quad a_2 = \left(1, \left[\frac{q_{\max}-1-a_1}{a_1} \right] \right)$$
$$\text{and } \ a_3 = \left(1, \left[\frac{q_{\max}-1}{1+a_1a_2} \right] \right). \tag{32}$$

From these relations, the third order shift may be calculated as:

$$\delta^{(3)}(q_{\max}) = \left| \frac{i}{n^{(3)}(q_{\max})} - \frac{p_3}{q_3} \right|. \tag{33}$$

Since the effective resolution is given from (18) it may be more useful to consider finite length fractions with partial quotient bounded as $a_i \leq a_{\max}$. In such a case calculations are easier. At first order $i_{\max} = 1$,the number of fractions is $n^{(1)} = a_{\max}$ and such fractions $p_1/q_1 = 1/a_1$ are located at

$$i = a_{\max} - a_1 + 1, \tag{34}$$

as above in (25). At second order, $n^{(2)} = a_{\max}^2$ and fractions $p_2/q_2 = a_2/(1 + a_1 a_2)$ are located at

$$i = a_{\max}^2 - a_1^2 + a_2. \tag{35}$$

Then for $i_{\max} = 3$,$n^{(3)} = a_{\max}^3 + a_{\max}$, there are fractions $1/a_1$ located at

$$i = (a_{\max} - a_1 + 1)(a_{\max}^2 + 1), \tag{36}$$

and there are fractions $p_3/q_3 = (1 + a_2 a_3)/(a_1 + a_3 + a_1 a_2 a_3)$ that we find at

$$i = (a_{\max} - a_1)(a_{\max}^2 + 1) + a_2 a_{\max} - a_3 + 1. \tag{37}$$

These explicit calculations are of restricted interest however since they can-

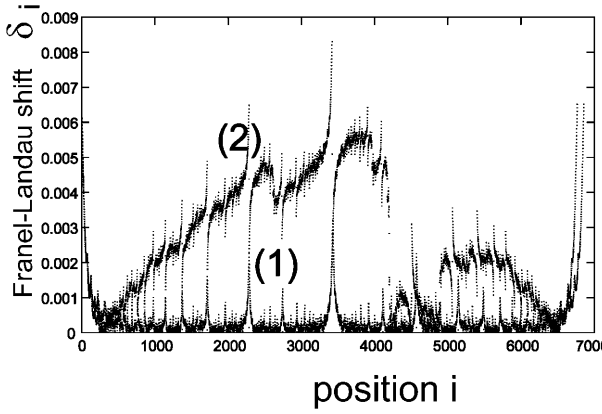

Fig. 9. The Franel–Landau shift δ_i in the well resolved case (curve 1) and for finite length fractions $i_{\max} = 7$ (curve 2)

not be generalized easily to arbitrary $i_{\max}$ or $a_{\max}$. Our MAPLE procedure does not show this drawback (see Fig. 9, curve 2) and the related comments.

2.3 Diophantine signal processing and $1/f$ frequency fluctuations

Experiments above clearly demonstrated the diophantine nature of signal processing realized in the receiver shown in Fig. 1. The spectrum of frequencies and amplitudes of beat signals was interpreted in terms of simple rules

(16) and (24) of the diophantine type. Although the amplitude-frequency relationship looks very complicated, the dynamics is predictable provided we know what limits the continued fraction expansion (17) of the frequency ratio ν.

One supplementary task is to understand the time dynamics and the origin of the transition from white frequency noise to $1/f$ frequency noise close to resonance. This may be approached from the diophantine nature of "waves"

$$\epsilon_i = \cos\{2\pi \ \ p_i/q_i\} \ \ , \text{i=1, } \ldots \text{ , n,} \tag{38}$$

which constitutes the observed signal, with $n \sim 3 \ \ q_{\max}^2/\pi^2$ for a well resolved signal, or less if the signal lacks resolution close to resonance. For a continuous (i.e. non-diophantine) signal we would expect that the resolution $d\nu = 1/n$ vanishes when n increases, so that by integrating over the frequencies $\nu = p_i/q_i$ we would get the mean value

$$\sigma = \int_{\nu=0}^{\nu=1} \epsilon(\nu) d\nu = \int_{\nu=0}^{\nu=1} \cos(2\pi\nu) d\nu = \sin(2\pi\nu)/2\pi = 0. \tag{39}$$

Such a result is not observed in a finite resolution experiment. Let us consider the discrete partial sum over the waves as:

$$M(i) = \sum_{k=1}^{i} \cos\{2\pi \ \ p_k/q_k\} \ \ , \text{i=1, } \ldots \text{ , n,} \tag{40}$$

and the discrete partial sum over the frequency shifts as:

$$S(i) = \sum_{k=1}^{i} |k/n - p_k/q_k| \quad , \text{i=1, } \ldots \text{ ,n.} \tag{41}$$

The terms ϵ_i and δ_i will be called Littlewood term and Franel–Landau shift respectively, and $M(n)$ and $S(n)$ will be called Littlewood and Franel–Landau sums. This refer to the history of Riemann zeta function that we will remind below. Let us first consider the (resonant) case when continued fraction expansions are restricted to a finite partial quotient $a_i \leq a_{\max}$. Fig. 10 illustrates the case $a_{\max} = 8$. It can be observed in (a) and (b) that the shift δ_i and the wave ϵ_i are fractal like. This is confirmed by computing numerically the fractal dimension (see the companion paper by C. Eckert). Also the mean deviation follow the laws $a_{\max}\sigma_F = S(n)a_{\max}/n \sim 0.40$ and $a_{\max}\sigma_L = M(n)a_{\max}/n \sim -1.5$ where n is the number of resolved fractions. We say that this dependence is of the $1/f$ type, since the deviation is constant at all scales n. This is similar to what was obtained in Fig. 5 in the almost resonant interaction of the two input oscillators. Going far away from the resonance we expect to reach a well resolved zone, i.e. at arbitrarily small resolution we should recover the continuous case (39). We found numerically

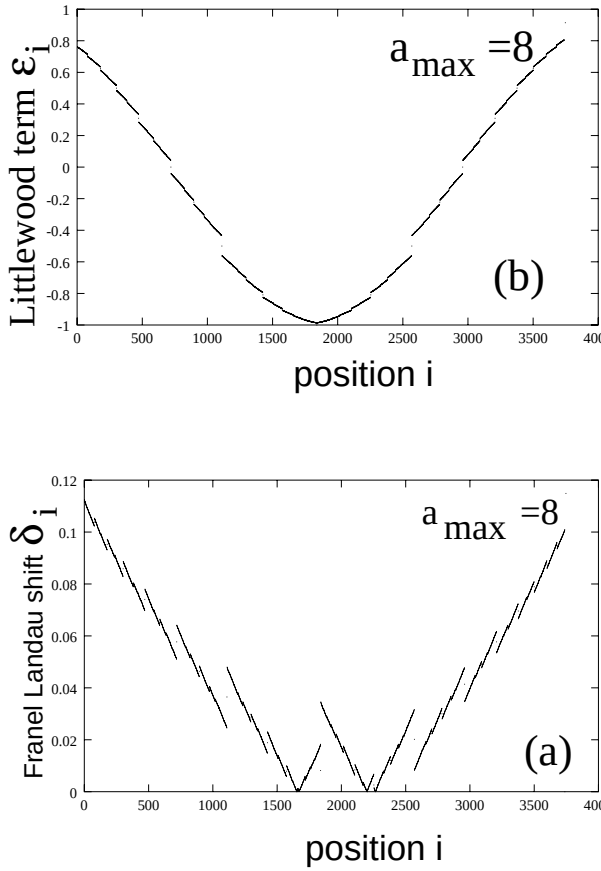

Fig. 10. The Littlewood term (b) and the Franel–Landau shift (a) for fractions with bounded partial quotient $a_{\max}$

that the deviation scales approximatively as $\sigma_F \sim n^{-3/4}$; this is what we expect also from Riemann hypothesis as will be reminded below. The experiments showed an Allan deviation depending on the integration time as $\sigma \sim \tau^{-1/2}$ (see Fig. 5).

Finally in the intermediate zone, at the outer edge of a resonant zone, there is an increase of the number of fractions and then a decrease of the resolution as shown in Fig. 8. As a result there is an increasing shift δ_i at most subharmonics (Fig. 9, curve 2 where $i_{\max} = 7$) as compared to the well resolved case (curve 1). There we calculate a drift $\sigma_F \sim n^{\alpha}$, with $\alpha > 0$ depending on the value of $i_{\max}$. For $i_{\max} \leq 3$ as calculated above, $\alpha \sim 1/4$ at large n, which reminds the experimental value $\sigma \sim \tau^{1/4}$ in Fig. 4. A similar dependance is obtained for σ_L.

We have given a diophantine approach of the low frequency noise close to a resonance in the receiver. It would be nice at this stage to compare quantitative results given from the Allan deviation $\sigma_y(\tau)$ in (15) and the mathematical result from σ_F or σ_L. To achieve such a goal we need to know precisely what determines the resolution n^{-1} and the truncation at $a_{\max}$ or $i_{\max}$ and relates these measures to physical parameters, in particular the filter bandwith and the noise of input oscillators.

2.4 The Riemann zeta function and the Riemann hypothesis and physics

In his celebrated paper on the number of prime numbers $\pi(q_{\max})$ less or equal than $q_{\max}$, Riemann studies the extension of the zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}} \text{ (with } \Re\ (s) > 1), \tag{42}$$

to the complex plane s [11]. The connection to $\pi(q_{\max})$ is through the log of this function and the Euler product formula is one way to express the decomposition of any integer as a product of powers of prime numbers. Riemann extends the formula to the whole complex plane as follows

$$\zeta(s) = \frac{\Pi(-s)}{2i\pi} \int_{+\infty}^{+\infty} \frac{(-x)^s}{e^x - 1} . \frac{dx}{x}, \tag{43}$$

where $\Pi(s) = \int_0^{\infty} e^{-x} x^s dx$ and $s > -1$ is the Euler integral, and $\Pi(s) = s!$ whenever s is a natural number. In (43) the path of integration begins at $+\infty$, moves to the left down the positive real axis, circles the origin in the counterclockwise direction and returns up the positive real axis to $+\infty$. Formula is valid for all s. It is analytic at all points of the complex plane s except for a single pole at $s = 1$, with residue 1. From calculations already performed by Euler for real x one gets a connection to Bernoulli numbers B_n. Using the expansion near $x = 0$ of $x/(e^x - 1) = \sum_{n=0}^{\infty} B_n x^n/n!$, one gets $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$, $B_4 = -1/30$,$B_6 = 1/42$... and $B_{2n+1} = 0$ if $n > 1$ and the formula $\zeta(-n) = (-1)^n B_{n+1}/(n+1)$. It follows that the zeta function has "trivial" real zeros at $s = -2, -4, -6$.... Also some real zeta functions are obtained easily as $\zeta(0) = -1/2$, $\zeta(-1) = -1/12$, $\zeta(-3) = -1/120$, $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$ ... .

Riemann hypothesis asserts that the only non trivial zeros are located on the imaginary axis $s = 1/2$. Billions of zeros are known all on the critical line but until now the hypothesis has not be proved. It was shown by Littlewood in 1912 that a statement equivalent to Riemann hypothesis may be obtained from the Dirichlet series associated to the inverse of the zeta function as

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} \quad (\text{with } \Re(s) > 1), \tag{44}$$

where the Möbius function $\mu(n)$ is zero if n is divisible by a square, is 1 if n is a product of an even number of distinct prime factors and is -1 if n is a product of an odd number of distinct prime factors. The inverse zeta function may be extended to the whole complex plane as:

$$\frac{1}{\zeta(s)} = s\int_0^\infty \frac{M(x)}{x^{s+1}}dx, \tag{45}$$

and $M(x) = \sum_{n=1}^{x} \mu(n)$ the partial sum over the Möbius function. Littlewood proved that Riemann hypothesis is equivalent to the statement

$$M(x) = 0(x^{\frac{1}{2}+\epsilon}) \quad \text{for all } \epsilon > 0. \tag{46}$$

There is an interesting history behind that early work which is reported by Edwards [11]. I refer to the conjecture by Mertens: $|M(x)| < x^{1/2}$ which was disproved only in 1984 by Odlyzko and the unproved Stieljes statement in 1885: $M(x) = 0(x^{1/2})$. A significant step was achieved in 1920 by Franel and Landau by introducing Farey fractions and the shifts δ_i as given in (24). M(x) may be alternatively written as a sum over the waves $M(n)$ as obtained from (40) so that using (22) Littlewood conjecture is $M(n) = 0(n^{\frac{1}{4}+\epsilon})$ whatever ϵ. The equivalent Franel–Landau conjecture is the sum over the shifts δ_i and may be similarly written as $S(n) = 0(n^{\frac{1}{4}+\epsilon})$ for all ϵ where $S(n)$ follows from (41).

It is useful at that point to remind that physical interpretations of (46) have been proposed. In 1931, A. Denjoy [11] suggested that since the Möbius function equals plus or minus one with apparent equal probability (which is $3/\pi^2$) it looks similar to flipping a coin with equal chance for heads and tails, so that $M(x)$ looks like a brownian motion [13]. Consequently the Riemann hypothesis would be true with probability one!

More recently it was observed [14] that the Riemann zeta function $\zeta(s)$ has a quantum statistical interpretation. The unique factorization of an integer into primes, $n = \prod_{p=1}^{p_{\max}} p^{N_p}$ translates into a unique decomposition of $\log n = \sum_{p=1}^{p_{\max}} N_p \log p$, where N_p corresponds to the number of "particles" of energy $\log p$. It follows that the zeta function is the grand canonical partition function of a "Riemann gas" of quantum bosons of energies $\log p$ (because there is no factor of indistinguishability in the expression for the whole energy $\log n$. Similarly $1/\zeta(s)$ was interpreted as a gas of free "ghost" fermions with the same energy $\log p$ (there is a fermionic aspect since $\mu(n) = 0$ at a square number n).

The interpretation of frequency variability is thus embedded into the number theoretical concepts attached to $1/\zeta(s)$. One can go a step ahead by looking at the integer n in (42) as an integration time τ and consider the variable s as a complex time $1/2 + \eta + it$, where η is a small real parameter. Using the decomposition $1/\zeta(s) = Z(t)\exp[i\theta(t)]$ one easily computes numerically the discrete Fourier transform $S(\omega) = \sum_{t=0}^{\infty} Z(t)\exp[i\omega t]$ with the result

$|S(\omega)|^2 \sim C^{te}$ on the critical line (at $\eta = 0$) and $|S(\omega)|^2 \sim 1/\omega$ close to it. This shows a transition from white to $1/\omega$ power spectrum of the modulus of $1/\zeta(s)$. Similarly the power spectral density of the phase $\theta(t)$ decreases like ω^{-2} at the critical line and like ω^{-3} close to it. This corresponds to a transition from white to $1/\omega$ frequency noise as observed in our experiments. In addition one predicts peaks in the power spectral density at $\omega_n = C\ log\ q$ with $q = p^l$, l integer and p prime. The first low frequency peaks are found at 2, 3, 2^2, 5, 7, 2^3, 3^2, 11, 13 ... . This may be explained on the basis of the derivative of log $\zeta(s)$ [11] or in relation to a quasiclassical approximation of quantum chaotic orbits [15]

Consequently one has a number theoretical model of the transition from white to $1/f$ frequency noise based on the properties of $1/\zeta(s)$ at or close to the critical line $s = 1/2$. The deviation of the beat note with respect to the perfect oscillation should be in the form $1/\zeta(1/2 + \eta + it)$ with η non-zero at resonance and η equal to zero away from it. This model has now to be checked in detail by new experiments related to receivers and particularly in the digital domain.

References

1. Klapper J., Frankle J. T. (1972) Phase-Locked and Frequency-Feedback systems. Academic Press, New-York
2. Maas S. A. (1986) Microwave Mixers. Artech House, Norwood
3. Planat M. (1993) Plurifractal Signature in the Study of Resonances of Dynamical Systems. Fractals, 1:727–734
4. Dos Santos S. , Planat M. (1998) Arithmetical Fractals in an Electronic Loop. In: Novak M. N. (Eds.) Fractals and Beyond. World Scientific, Singapore, 297–306
5. Cvitanovic P. (1992) Circle Maps: Irrationally winding. In: Waldschmidt M. , Moussa P. , Luck J. M. , Itzykson C. (Eds.) From Number Theory to Physics. Springer Verlag, Berlin, 631–658
6. Lichtenberg A. J., Lieberman M. A. (1991) Regular and Chaotic Motion. Springer Verlag, Berlin
7. Allan D. W. (1987) Time and Frequency (Time Domain) Characterization, Estimation and Prediction of Precision Clocks and Oscillators. IEEE Trans. Ultrason. Ferroelect. Freq. Contr., 34:647
8. Eckert C., Planat M., Miehé J. A. (1996) Hidden Order in the Frequency Noise of an Electronic Oscillator. Phys. Rev. E, 54:6093
9. Tagaki K., Serikawa S., Doi T., Mukumoto M. (1997) Phase Noise, Amplitude Noise and their Correlation and the Low Frequency Noise Reduction Method in Bipolar Transistors. In: Clays C., Simoen E. (Eds.) Noise in Physical Systems and $1/f$ Fluctuations. World Scientific, Singapore, 177–180
10. Perrine S. (1996) L'Arithmétique sur une Surface Percée. In: Planat M. (Eds) Oscillators. Annales des Télécommunications, Paris, 7: 407–420
11. Edwards M. H. (1974) Riemann's zeta function. Acad. Press, New-York
12. Korobov N. M. (1997) On Finite Continued Fractions. Communications of the Moscow Mathematical Society, 1302

13. MacDonald D. K. C. (1962) Noise and Fluctuations: an Introduction. John Wiley and Sons, Paris
14. Julia B. L. (1994) Thermodynamic Limit in Number Theory: Riemann-Beurling Gases. Physica A, 203: 425
15. Berry M. V. Semiclassical theory of spectral rigidity. Proc R Soc Lond A400:229-251 (1985)

Detection of Chaos in the Noise of Electronic Oscillators by Time Series Analysis Methods

C. Eckert[1⋆] and M. Planat[2⋆⋆]

[1] Laboratoire de Physique et Applications des Semiconducteurs, CNRS, 23 Rue du Lœss, BP 20, F-67037 Strasbourg Cedex 2, France
[2] Laboratoire de Physique et Métrologie des Oscillateurs, CNRS, 32 Rue de l'Observatoire, F-25044 Besançon Cedex, France

Abstract. Frequency fluctuations of an electronic oscillator are studied by time series analysis methods in order to detect an eventual underlying attractor. Measurements are performed by a period counting technique with a reference signal of 10 MHz and various frequencies of the local oscillator. Three different behaviours of the Allan variance are observed, depending on the mean frequency of the beat signal. The results of time series analyses clearly show that two of these behaviours are associated with the presence of a chaotic process, whereas the third is more intricate. When applied to data computed from continued fraction expansions of real numbers, these methods lead to the same conclusions. For two kinds of truncation, the data present a chaotic behaviour, which is not obvious for the third kind. A correspondence between the behaviour of the frequency fluctuations and the way of truncating the continued fraction expansions can thus be proposed.

1 Introduction

Since E.N. Lorenz [1] realized in 1963 that low-dimensional deterministic systems can take the appearance of noisy fluctuations, such oddnesses have been discovered in fields as various as physics [2], chemistry [3], biology [4], economy [5], and climatology [6]. Erratic time series have therefore been widely studied, and great efforts have been devoted to the distinction between chaotic and stochastic processes.

One feature common to most of the time series analysis methods is the time delay embedding procedure [7] which reconstructs a possible attractor from a scalar time series. The popularity of this procedure comes from the embedding theorem [8], [9] which states that, under suitable conditions, it unfolds the underlying attractor, so that such properties as the correlation dimension [10], the Kolmogorov entropy [11], the phase portraits and Poincaré sections [12], the Lyapunov exponents [13], etc., may be extracted.

This article describes key advances in the detection of an underlying chaotic process in the noise of an electronic oscillator, including the estimation of a low correlation dimension. Previous works [14], [15] showed that

⋆ eckert@phase.strasbourg.fr
⋆⋆ planat@lpmo.univ-fcomte.fr

M. Planat (Ed.): LNP 550, pp. 288–304, 2000.
© Springer-Verlag Berlin Heidelberg 2000

the fluctuations of an electronic oscillator can be understood in terms of a low-dimensional deterministic system, but none of its properties could be extracted. Here both experimental time series, recorded by a customary period counting technique, and data computed from continued fraction expansions (CFE) of real numbers, are examined. For low values of the mean beat frequency, the experimental fluctuations clearly have a behaviour characteristic of a chaotic system, with a correlation dimension close to one. This contrasts with the results obtained for higher beat frequencies where the ambiguity on the kind of process, deterministic or stochastic, is difficult to overcome. As concerns the computed data, the only difference between the various calculi lies in the way of truncating the CFE. Here also the outcomes can be splitted into two classes, one presenting typical features of an underlying attractor, and a second class for which the estimation of a correlation dimension is awkward. It is suggested that the different kinds of frequency fluctuations, observed by varying the frequency of the local oscillator, may be associated to different truncations of the CFE. This may help to improve the understanding of frequency fluctuations, which seems particularly important in the scope of telecommunications if one reminds that the device used to measure frequency fluctuations is the corner-stone of any communication receiver.

2 Experimental and Computed Data

Typically, a period counting experiment consists in applying the signal of the oscillator under test and a reference signal of about the same frequency to a multiplier. Both signals are mixed and result in a beat signal that passes through a low pass filter, intended to reject high frequency components, before being measured by a reciprocal counter. In the usual operating mode, the gate of the counter opens at one zero crossing of the beat signal, counts the periods of both the high frequency counter clock signal and the beat signal during a sampling duration τ, and closes. In the present experiments [16], the gate opens at one zero crossing of the beat signal and closes at the next, so that taking the reciprocal leads to the measurement of a frequency that may be considered as instantaneous for no average over a sampling duration is performed. The gate again opens at the next zero crossing, closes, and so on, resulting in the recording of a collection of x_i, with i denoting the i^{th} measurement of the instantaneous frequency x. The frequency of the reference signal was fixed to $f_0 = 10$ MHz and several time series were recorded for various frequencies f_1 of the local oscillator.

Such experiments have been shown [17] to be linked to the number theory. Indeed, the best rational approximation of any real number ν is given by its CFE

$$\nu = a_0 + 1/(a_1 + 1/(a_2 + ... + 1/(a_i + ...))). \tag{1}$$

From the experimental point of view, ν represents the ratio between the frequencies f_1 of the local oscillator and f_0 of the reference signal. The ratio

between the frequencies of the beat signal and of the reference signal can be represented formally by the relation

$$\mu = q|\nu - p/q|, \tag{2}$$

where p and q are integer numbers such that $p < q$. Considering the set of n Farey ordered fractions

$$0/1 < ... < p_i/q_i < ... < 1/1 \tag{3}$$

whose denominator not exceeds $q_{\max}$, and the set of n equidistant fractions i/n, the deviations

$$\delta_i = |i/n - p_i/q_i| \tag{4}$$

can be calculated for $i = 1, ..., n$. The requirement $q < q_{\max}$ thus gives a first way to compute a series of data δ_i. Other collections can be obtained by imposing further restrictions. The Farey fractions p_i/q_i being rational numbers, their development according to (1) involves a finite number of terms a_i. This suggests to define two sorts of truncation for further calculations. The first consists in discarding the p_i/q_i, hence the δ_i, whose CFE (1) comprises more than $i_{\max}$ terms, and the second amounts to reject the fractions involving a value a_i greater than a fixed $a_{\max}$. The three truncation criteria $q_{\max}$, $i_{\max}$, and $a_{\max}$ give rise to three kinds of computed data.

An experimental time series corresponding to a mean beat frequency of 3.3 Hz is plotted in Fig. 1. Further examples of experimental and computed time series, together with details about the experiments and the calculation of data can be found in [16].

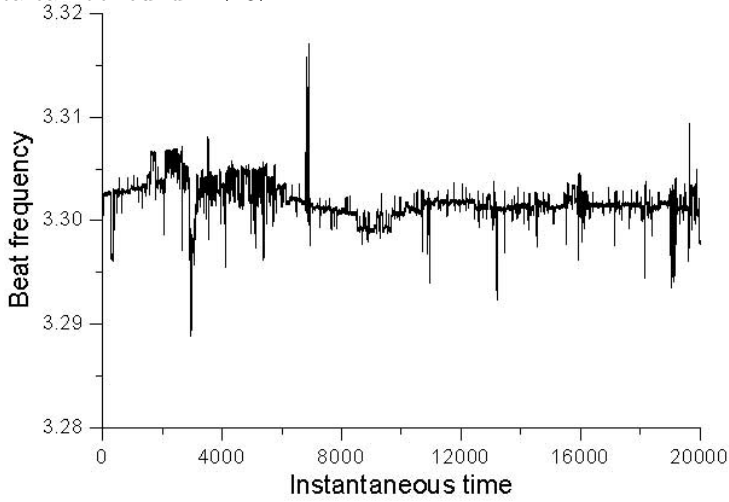

Fig. 1. Instantaneous frequency fluctuations of the beat signal recorded at 3.3 Hz.

3 Time Series Analysis Methods

During the last two decades, the distinction between stochastic and chaotic processes has been the subject of numerous works [4], [11] - [14], [18] - [22]. In the proposed methods, the time delay embedding procedure [7] has become into common use as a way to reconstruct a possible underlying attractor from a scalar time series. This procedure comes to construct a set $\{\mathbf{Y}_1, \mathbf{Y}_{1+k}, \mathbf{Y}_{1+2k}, ..., \mathbf{Y}_{1+kN_v}\}$ of m-dimensional vectors

$$\mathbf{Y}_i = \{y_i, y_{i+l}, y_{i+2l}, ..., y_{i+(m-1)l}\} \tag{5}$$

from the scalar time series $\{y_i, i = 1, ..., N_p\}$ for fixed delays k, l and embedding dimension m.

According to the embedding theorem [8], [9] the original attractor can be unfolded provided that the parameters k, l, and m are properly chosen. Denoting by d the dimension of the underlying attractor, $m > 2d$ would surely be sufficient [8], [9], but the optimal embedding dimension may be smaller. The delays k and l were introduced to prevent short-correlation effects [18] – [20], [23]. In practice k is the delay, in terms of measurements numbers, between the first components of successive vectors, while l is the delay between successive coordinates of a vector.

The method of the false nearest neighbours percentage (FNNP) [21] relies on the study of the topological properties of the reconstructed vectors. Its main advantage is to be efficient even for short and/or noisy time series. However, it is limited to the detection of an underlying attractor, and none of its features can be extracted. Regarding the correlation dimension method [10], it is the most popular method used to estimate the fractal dimension of an attractor, but it can give rise to misleading results because of its high sensitiveness to the time series length and to the inherent experimental noise.

3.1 False nearest neighbour percentage

In this method, the repartition of the vectors constructed in m dimensions is compared to that achieved in a $(m + 1)$-dimensional embedding space. The subjacent idea is that as long as the embedding dimension is too small, or when there is no attractor at all, the neighbourhood of a point is greatly affected by the increment of the dimension. So, points that are neighbours in the m-dimensional space will no longer be neighbours in $m + 1$ dimensions. This is illustrated in Fig. 3 of [24] for the Hénon map. At the opposite, in the case of an attractor embedded in a space of sufficiently high dimension, the distribution of the reconstructed vectors is essentially unaltered by the increase of the dimension. The conservation of the neighbourhood relations in passing from a m- to a $(m + 1)$-dimensional space can thus be taken as a proof of the attractor's existence and used to determine the minimal acceptable embedding dimension and delays.

Let $\mathbf{Y}_{i,n(m)}$ and $\mathbf{Y}_{i,n(m+1)}$ be the n^{th} nearest neighbour in m and $m+1$ dimensions, respectively, of any given reference point $\mathbf{Y}_i$. Their squared Euclidean distances from $\mathbf{Y}_i$ are, respectively, denoted by

$$D_m(i, n(m)) = |\mathbf{Y}_{i,n(m)} - \mathbf{Y}_i|^2 \tag{6}$$

$$D_{m+1}(i, n(m+1)) = |\mathbf{Y}_{i,n(m+1)} - \mathbf{Y}_i|^2. \tag{7}$$

In order to identify the false nearest neighbours, Kennel et al. [21] proposed to reckon the FNNP by jointly applying the two following criteria :

$$\frac{D_{m+1}(i, 1(m)) - D_m(i, 1(m))}{D_m(i, 1(m))} > R_{\text{tol}}^2 \tag{8}$$

and

$$D_{m+1}(i, 1(m)) > F_{\text{tol}}\sigma^2, \tag{9}$$

where R_{tol} stands for the relative increase of the distance that is tolerated, F_{tol} denotes some tolerance factor, and σ^2 is the variance of the data. The first inequality (8) rejects neighbours whose distance from $\mathbf{Y}_i$ increases too much when the dimension is incremented. However, due to the limited number of data in the time series, the distance may already be large in m dimensions, so that its relative increase in passing to a $(m+1)$-dimensional space is small. In this case, the neighbour successfully passes the test (8), although it should be considered as a false neighbour because of its remoteness from $\mathbf{Y}_i$. The criterion (9) is intended to properly identify such false neighbours. If one of the two inequalities (8) or (9) is verified, then the nearest neighbour of $\mathbf{Y}_i$ is declared as a false neighbour. Testing all the embedded vectors $\{\mathbf{Y}_i, i = 1, 2, ..., 1 + N_v\}$ gives rise to the number of false neighbours, which is finally divided by $1 + N_v$ in order to get the FNNP.

In the case of a time series that underlies an attractor, the FNNP is high so long that the embedding dimension is too small. It diminishes when m is increased, until becoming negligible when the accurate embedding dimension is reached. It is worth remarking that the FNNP should theoretically drop to zero. However, in passing from a m- to a $(m+1)$-dimensional space, the vector $\mathbf{Y}_i$ and its nearest neighbour $\mathbf{Y}_{i,1(m)}$ both acquire a $(m+1)^{\text{th}}$ coordinate. This leads to an increase of $D_m(i, 1(m))$ that is at least of the order of magnitude of the experimental noise, including the rounding off errors, so that some nearest neighbours may unjustly be considered as false neighbours. On the contrary, if the time series results from a stochastic process, the FNNP may decrease at the very beginning of the dimension incrementation procedure, but not so much as for chaotic data, and it can never be neglected.

Turning now to the determination of the accurate parameters of the embedding space, the embedding dimension m can be chosen as the one for which the FNNP becomes negligible. As far as the delays k and l are concerned, there are different ways to find their proper value. The first way is

to compute the percentage of false neighbours for various parameters k and l, and to examine their influence, bearing in mind that the FNNP should be as low as possible. The delays have also been shown [25] to influence critical properties such as the mean time separation (in sampling units) between a reference point and its nearest neighbour. If this time separation suddenly decreases when the embedding dimension is incremented and stays to a low value, then it can safely be inferred that the time lags k and/or l are too small. Another way of estimating the optimal parameters is to compute [20]

$$W = \ln \left\langle \prod_{n=1}^{w} \left[\frac{D_m(i, n(m))}{D_m(i, n(m+1))} \times \frac{D_{m+1}(i, n(m))}{D_{m+1}(i, n(m+1))} \right]^{1/w} \right\rangle_i \tag{10}$$

for increasing m and different delays k and l, where $\langle\ \rangle_i$ denotes an average over various reference points $\mathbf{Y}_i$. In the case of an underlying attractor, W is expected to decrease towards zero when m increases, and the proper embedding parameters are those for which W is minimal. However, this method requires more computing time than the method of the FNNP. More annoying, it is less efficient because the upper limit is known in the latter method (100 %), but not in the former, so that it is difficult to estimate what can be considered as a negligible value of W.

3.2 Correlation dimension

The starting point of the method proposed by Grassberger and Procaccia [10] is the computation of the correlation integrals

$$C(m, r) = \frac{2}{N_v(N_v + 1)} \sum_{i=1}^{N_v} \sum_{j=i+1}^{N_v+1} H(r - |\mathbf{Y}_i - \mathbf{Y}_j|) \tag{11}$$

for various embedding dimensions m, with $H(r)$ denoting the Heaviside function and —— the Euclidean distance. So, $C(m, r)$ is the fraction of pairs of vectors whose distance is lower than r. Obviously, $C(m, r)$ increases with r, for more and more pairs of vectors are taken into account in the sum of (11). At small r, the correlation integrals generally follow a power law

$$\lim_{N_v \to \infty,\, r \to 0} C(m, r) = r^{D_2(m)} \tag{12}$$

from which the correlation exponent $D_2(m)$ can be extracted.

In the presence of an underlying deterministic process, it is foreseen [10] that the correlation exponent begins by increasing with m, but less and less, and ends by reaching a constant

$$D_2 = \lim_{m \to M} D_2(m), \tag{13}$$

where M stands for the optimal embedding dimension. It has been argued [10] that D_2 is a lower bound of the fractal dimension. At the opposite, the correlation exponent of a stochastic time series is expected to continuously increase with the embedding dimension and to follow the straight line $D_2(m) = m$.

Before implementing this method, some word has to be said about its drawbacks, which arise for it clearly appears that none of the limits implicated in (12) can be achieved. The limit $r \to 0$ is unapproachable because of the experimental noise, the precision of the measurements, and/or the rounding off errors. Very small r correspond to distances whose order of magnitude is that of the noise. Consequently, taking a range of too small values of r leads to a correlation exponent characteristic of the noise [19] and not, as it should, representative of the possible underlying chaos. The local slope [19]

$$\ln[C(m,r)]/\ln(r) \tag{14}$$

may help to estimate the range of r that is suited for the estimation of $D_2(m)$. Moreover, the effect of the noise is amplified when the embedding dimension is incremented because its contribution increases, resulting in an enlargement of the range of r for which the correlation integral depicts the noise. On the other hand, the limit $N_v \to \infty$ is prevented both by experimental considerations and by the immoderate increase of the recording and computing times.

It finally must be noted that, as well as for the methods of false neighbours, the modifications of the correlation exponents induced by the variation of k and l allow to get an idea of their proper values. Yet, it should be remembered that the number of embedded vectors is related to the number of data points by $1 + N_v = 1 + [N_p - 1 - (m-1)l] / k$, so that the higher the values of k and/or l, the smaller the number of embedding vectors.

4 Detection of Chaos in Experimental and Computed Data

In this section the time series analysis methods just described are applied to the time series whose recording and computing has been reported in Sect. 2. Both kinds of data being treated in the same manner, some correspondences will become apparent.

4.1 Experimental time series

The experimental time series have been used to construct embedding vectors from their general definition (5). The embedding dimension was varied from $m = 1$ to 10 and the time lags were fixed as being equal to $k = l = 1$. The FNNP has been computed for R_{tol} ranging from 5 to 50, and $F_{\mathrm{tol}} = 4$ as suggested by Kennel et al. [21]. The percentage of false neighbours nearly equals 100 % when $m = 1$ and decreases for increasing values of R_{tol},

as expected from (8). For high embedding dimensions the FNNP becomes independent of R_{tol} (see [15], [24] for examples), due to the finite length of the time series. Indeed, the increase of the dimension leads to the lowering of the density of embedded vectors, so that false neighbours are chiefly rejected according to the criterion (9), hence independently of the tolerated threshold.

For all the considered experimental data, the percentage of false neighbours is minimal when $m = 4$ and $R_{\mathrm{tol}} = 30$, and a further increase of R_{tol} has almost no effect, indicating that none of the neighbours that have to be declared as false from the first criterion (8) are missed. This value of the tolerated relative increase of the distances has thus been retained for the plot (Fig. 2) of the FNNP. The curves corresponding to a mean beat frequency of 100, 4.43, and 0.5 Hz have 2000, 2000, and 1000 data points, respectively, whereas the time series recorded at 3.30 Hz has a length of $N_p = 20000$ points. It clearly appears from Fig. 2 that the different time series can not be considered as behaving alike. The FNNP corresponding to the data recorded at 4.43 Hz is characteristic of an underlying deterministic system, whereas it never drops to zero for the other time series. Because of the relatively high value (11.5 %) of its minimal FNNP and of the increase observed when m is further incremented, the time series recorded at 100 Hz is the one that seems the most result from a stochastic process. However, caution has to be taken for this may eventually arise because of a relatively high noise level [24]. Between these two curves there is another one (empty diamonds) that may arise as well from a stochastic process as from a chaotic system blurried by the experimental noise. The FNNP obtained for the longer time series (filled diamonds) suggests that these data have a determinist origin.

As noted before (Sect. 3.1), the alterations of the results induced by the variation of the time delays allow to fixe their proper value. The independence of the FNNP on k and l is a sign of the absence of short-time correlations. This is corroborated by the reckoning of the mean time separation, in sampling units, between a point and its first neighbour, which is near 200 for the shorter time series, close to 500 for those of intermediate length, and beyond 3000 for the longer one. The calculation of the averaged wavering product according to (10) has been performed to further test the effect of the time lags. For this computation, the geometric mean has been done over $w = 10$ nearest neighbours [20] and all the data points have been taken into account for the arithmetic average. For all the considered time series, for a given embedding dimension m, W is minimal when $l = 1$. A departure from $k = 1$ also increases W, so that the same conclusion as previously intrudes, namely, the best choice for the time delays is $k = l = 1$.

In order to evaluate the correlation dimension, correlation integrals have been computed from (11), with $k = l = 1$ and m ranging from 1 to 10, for various time series. The distances have been normalized with respect to the diameter of the embedding space in order to facilitate the comparison between the results obtained for different embedding dimensions and time

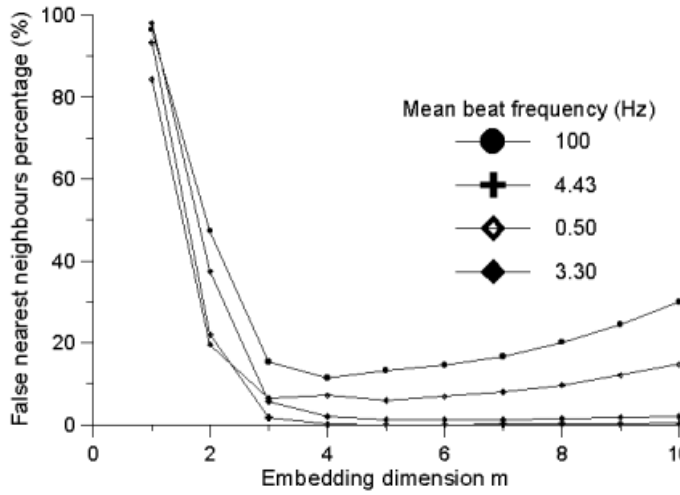

Fig. 2. False nearest neighbours percentage of time series measured at various mean beat frequencies. The embedding delays are $k = l = 1$ and the tolerated threshold is $R_{\text{tol}} = 30$.

series. It must be noted that the effect of varying the time lags was investigated, but no modification was detected. In the case of the fluctuations measured at 100 Hz, the correlation integrals lock much like those displayed in other works [24]. Their double logarithmic plot presents a linear region, but it is difficult to exclude that it depicts only the experimental noise. The

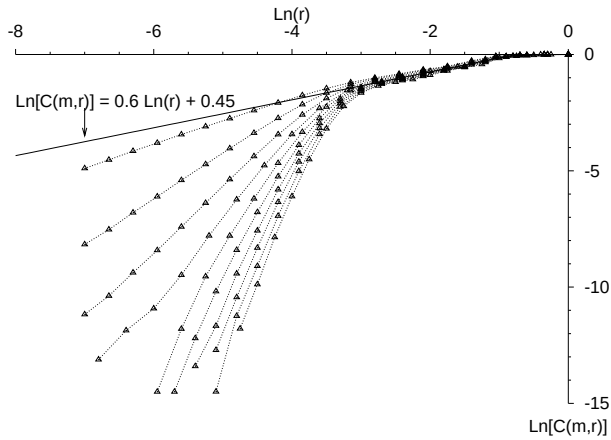

Fig. 3. Correlation integrals of the data recorded at 4.43 Hz embedded in a space with lags $k = l = 1$. The straight line results from the correlation dimension extracted in the range $0.05 < r < 0.3$.

correlation integrals of data recorded for a mean beat frequency of 4.43 Hz are shown in Fig. 3. For each embedding dimension, the curves $\ln[C(m,r)]$ versus $\ln(r)$ obviously present two linear ranges. The one corresponding to the lower values of r is characteristic of the inherent noise and leads to correlation exponents that increase continuously with m, ranging from about 1 for $m = 1$ to $D_2(m) > 8$ for $m = 10$. At the opposite, the correlation exponents extracted from the higher range of r are independent of the embedding dimension. The plots of the local slope (14) versus $\ln(r)$ reported in Fig. 4 give valuable insights on the presence of different scaling regions in the correlation integrals. Indeed, it has been argued [26] that after the region dominated by the noise, where $D_2(m)$ approaches the embedding dimension, there is a region, where the local slope presents a plateau, from which the correct correlation exponent can be estimated. So, comparison between the correlation integrals (Fig. 3) and the local slopes (Fig. 4a) of the data measured at 4.43 Hz shows evidence that $D_2(m)$ should be extracted in the range $0.05 < r < 0.3$, the linear region observed at lower values of r arising from the noise. As concerns the fluctuations recorded at 100 Hz, the local slope of the correlation integrals (Fig. 4b) doesn't exhibit any flat range, so that no correlation exponent can in principle be deduced. The double logarithmic plot of $C(m,r)$ nevertheless presents a linear range from which the values of $D_2(m)$ displayed in Fig. 5 have been extracted. The correlation exponents determined for other time series are also plotted in Fig. 5 as a function of the embedding dimension m. As can be seen on this figure, the time series can obviously be shared in two classes. In the case of the data measured at 100 Hz, the correlation exponents neither become constant, nor follow the straight line $D_2(m) = m$. This behaviour supports the conclusions deduced from the computation of the FNNP (Fig. 2) : the fluctuations may arise either from a stochastic process, or from the superposition of experimental noise to a chaotic system. The second category of data leads to correlation exponents that are nearly independent of the embedding dimension and can safely be identified with the correlation dimension D_2. It is of interest to remark that the lower correlation exponents (crosses in Fig. 5) are obtained for the data for which the percentage of false neighbours drops to zero (crosses in Fig. 2). Regarding the outcomes presented as diamonds in Figs. 2 and 5, those (empty symbols) whose correlation dimension is the higher also have the higher residual FNNP (Fig. 2), but the estimation of $D_2(m)$ (filled diamonds) is less accurate for the longer time series, though the corresponding percentage of false neighbours is lower. Nevertheless, for these three time series, the presence of an underlying low-dimensional system is manifest.

4.2 Computed time series

The outcomes related in this section concern four time series calculated from (4) by truncating the CFE (1) at $q_{\max} = 200$, $i_{\max} = 4$, $a_{\max} = 10$, and $a_{\max} = 50$. The embedded vectors have been constructed in the same

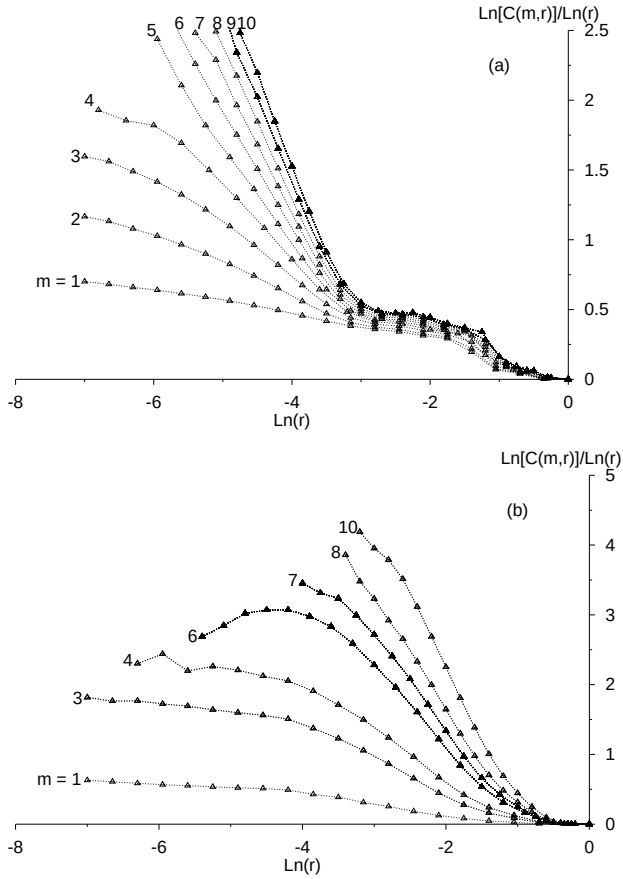

Fig. 4. Local slope of the correlation integral, for data measured at (a) 4.43 Hz and (b) 100 Hz.

way as for the experimental data, that is, with delays $k = l = 1$ for they have almost no effect, and embedding dimensions varying between 1 and 10. For all the considered time series, the FNNP that results from the application of the criteria (8) and (9), with $F_{\text{tol}} = 4$, sharply decreases in passing from $m = 1$ to $m = 2$. The data achieved by imposing the criteria $i_{\max}$ and $a_{\max}$ lead to a percentage of false neighbours that already approaches zero when $m = 3$. For this value of the embedding dimension, there still are 5 % of false neighbours left in the case of the data computed with the less severe restriction $q_{\max}$. However, this is not the minimal FNNP, it reaches zero for $m = 5$. Hence, one noticeable feature of these results, comparatively to those obtained for the experimental fluctuations, is that the percentage of false neighbours of all the computed time series rapidly tends towards zero.

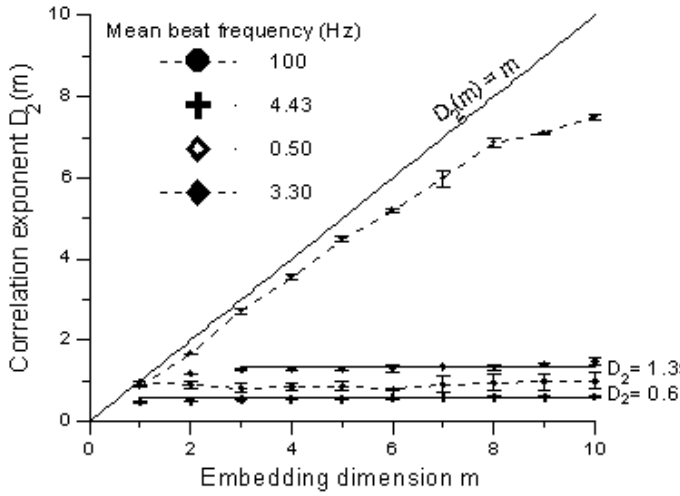

Fig. 5. Correlation exponents of time series recorded at four mean beat frequencies embedded in a space with lags $k = l = 1$. The straight line of slope 1 is also shown.

The results of applying (11) to the data calculated with the truncation requirement $q < q_{\max}$ are shown in Fig. 6 for $m = 1$ - 10. A linear range is observed at small values of r for low embedding dimensions ($m < 5$), but it shrinks when m is further incremented. This linear region can be looked on as reflecting the effect of the noise, though it is here limited to the rounding off errors, so that the extraction of a correlation exponent would not be advisable. This observation is to be compared to the features of the FNNP obtained for the same time series, which asymptotically reaches zero when m increases. Both outcomes contradict one another and therefore raise an important question about the origin, stochastic or deterministic, of the time series. Such a behaviour of the correlation integrals may be attributed to the absence of any underlying attractor. However, other chaotic time series blurried by noise have been shown [24] to entail similar apparent discrepancies.

As far as the time series get with the further truncation conditions $i_{\max}$ and $a_{\max}$ are concerned, the correlation integrals are in agreement with the FNNP. Indeed, the local slope (Fig. 7) and the correlation integrals (Fig. 8) both present a large range of values of r where no dependence upon the correlation dimension is observed. Comparison between Figs. 3 and 4 on one hand, and Figs. 7 and 8 on the other hand, evidences that in the case of the computed data the range of r from which a proper correlation exponent can be extracted is approximately five times larger than for the experimental fluctuations. The correlation exponents being independent of m, they may safely be considered as an estimation of the correlation dimension. The values derived from the different time series are $D_2 = 0.92$, 0.99, and 0.97 for

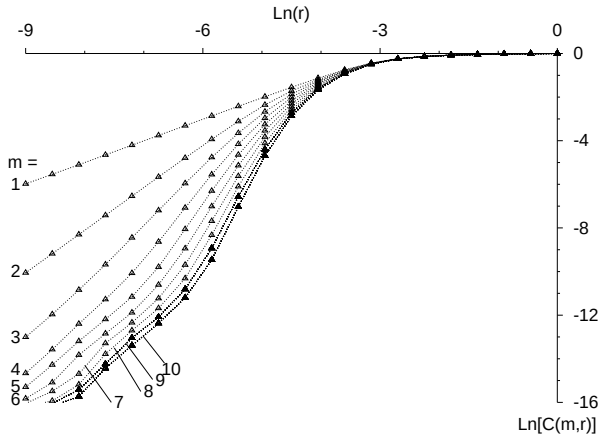

Fig. 6. Correlation integrals of data computed with the truncation condition $q_{\max} = 200$ embedded in a space with lags $k = l = 1$.

$i_{\max} = 4$, $a_{\max} = 10$, and $a_{\max} = 50$, respectively. However, they should be thought of as an order of magnitude because, from a practical point of view, a variation of about 5 % of the slope of the correlation integrals is almost imperceptible. It is therefore difficult to ascertain that the correlation dimensions of the three time series are different, nor that they differ from one.

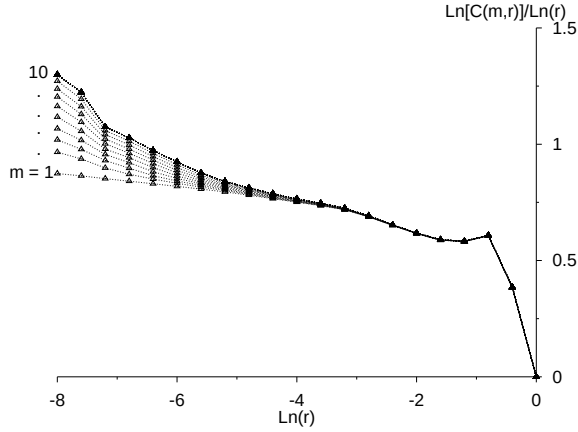

Fig. 7. Local slopes of the correlation integrals corresponding to data calculated with the restriction $i_{\max} = 4$ and embedded in a space with lags $k = l = 1$.

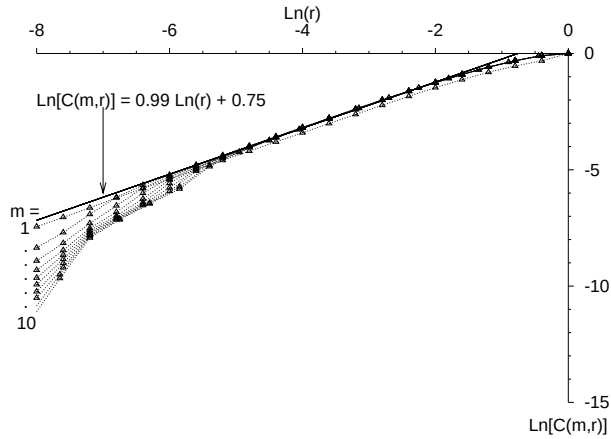

Fig. 8. Same as Fig. 6 in the case of data computed with the further restriction $a_{\max}$ =10. The straight line results from the correlation dimension extracted in the range $0.008 < r < 0.15$.

5 Discussion and Conclusion

So far, the experimental and computated time series have been handled separately, and the outcomes obtained by implementing the time series analysis methods have been discussed independently. It remains therefore to make the connexion between both kinds of data and results. As a first step, it is of importance to mention that the experimental fluctuations can be divided in three categories according to their temporal comportment and Allan variance [16], which is the common technique of characterization for the frequency noise of oscillators [27]. The time series measured at 100 Hz has the most erratic appearance and leads to an Allan variance that follows a power law $\tau^{-1/2}$. The fluctuations recorded for a mean beat frequency of 4.43 Hz are characteristic of a random telegraph signal and their Allan variance also varies as a power law, but with an exponent of 1/4. Finally, the two last time series present an irregular behaviour and their Allan variance is almost constant. Concerning the computed data, they also are of three kinds, corresponding to the three truncations $q_{\max}$, $i_{\max}$, and $a_{\max}$ (see Sect. 2).

Turning now to the examination of the outcomes of the time series analysis methods, one of the common features of the experimental and computed data is that, in each case, there is only one time series for which the presence of an underlying deterministic process doesn't assert itself as a pertinent origin. These are the data measured at 100 Hz and those calculated with the only requirement $q < q_{\max}$. In both cases, the absence of reliable correlation exponents (Figs. 4–6) is in favour of a stochastic origin of the data, which agrees with the white frequency noise indicated by the variation as $\tau^{-1/2}$ of the Allan variance [27]. However, the implementation of a noise reduction

procedure [28], especially to the experimental data, seems essential in order to make sure that the time series analyses are not warped by noise and, consequently, that no underlying attractor is present.

The comparison of the results obtained for the two other kinds of experimental and computed data enables likewise to highlight some correspondences. The fluctuations measured at 3.3 Hz appear as having the same origin as those recorded at 0.5 Hz. Indeed, they lead to a FNNP greater than that of the data whose mean beat frequency is 4.43 Hz (Fig. 2), though the former time series is much longer than the latter, thus excluding a common origin. They neither can be considered as having the same source than the data measured at 100 Hz (Fig. 2). Finally, when compared to the time series that has a mean beat frequency of 0.5 Hz, some similarities can be observed in their variation with time, and the percentage of false neighbours of the data measured at 3.3 Hz may be viewed as a curtailment, resulting from the increase of the length of the time series, of that of the data recorded at 0.5 Hz (Fig. 2). The Allan variance of these time series being constant, and taking into account that the Littlewood and Franel-Landau sums [16] of the data computed with the truncation criterion $a_{\max}$ are also constant, these experimental and calculated time series can be thought over as originating from a similar process. Besides, the time series recorded at 4.43 Hz and that computed with the truncation condition $i < i_{\max}$ are those for which the correlation dimension is the lowest (Figs. 5 and 7). Moreover, both their Allan variance and Franel-Landau sum vary as $\tau^{1/4}$ [16].

It can thus be concluded that the lowering of the mean beat frequency is accompanied by a transition from a process whose origin, stochastic or chaotic, has to be elucidated, to a behaviour characteristic of an underlying low-dimensional deterministic system. Moreover, a comparison with the results obtained for the computed data indicates that this transition can be understood as an effect of the degree of filtering, which comes out as different rational approximations of real numbers [16]. Further works have nevertheless to be undertaken in order to check if the time series measured at the higher mean beat frequency and computed with the less restrictive requirement $q < q_{\max}$ arise from a stochastic or a chaotic process. Another task consists in improving the correlation dimensions, with the aim to see if their values really differ one from another and/or from one. At last, it must be attempted to seize if the aspect of the computed time series [16] can be considered as the attractor of the corresponding chaotic systems.

References

1. Lorenz E.N. (1963) Deterministic Nonperiodic Flow. J Atmosph Sci 20:130
2. Perera A.G.U., Matsik S.G. (1995) Chaotic to Periodic Spontaneous Pulsing in Current Driven Silicon p-i-n Structures. Physica D84:615
3. Hauck T., Schneider F.W. (1994) Chaos in a Farey Sequence through Period Doubling in the Peroxidase-Oxidase Reaction. J Phys Chem 98:2072

4. Sugihara G., May R.M. (1990) Nonlinear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series. Nature 344:734
5. Scheinkman J.A., LeBaron B. (1989) Nonlinear Dynamics and Stock Returns. J Business 62:311
6. Tsonis A.A., Elsner J.B. (1992) Nonlinear Prediction as a Way of Distinguishing Chaos from Random Fractal Sequences. Nature 358:217
7. Packard N.H., Crutchfield J.P., Farmer J.D., Shaw R.S. (1980) Geometry from a Time Series. Phys Rev Lett 45:712
8. Mané R. (1981) On the Dimension of the Compact Invariant Sets of Certain Non-Linear Maps. In: Rand D.A., Young L.S. (Eds.) Dynamical Systems and Turbulence. Springer, Berlin, Lectures Notes in Mathematics 898:230
9. Takens F. (1981) Detecting Strange Attractors in Turbulence. In: Rand D.A., Young L.S. (Eds.) Dynamical Systems and Turbulence. Springer, Berlin, Lectures Notes in Mathematics 898:366
10. Grassberger P., Procaccia I. (1983) Measuring the Strangeness of Strange Attractors. Physica D9:189
11. Grassberger P., Procaccia I. (1983) Estimation of the Kolmogorov Entropy from a Chaotic Signal. Phys Rev A28:2591
12. Roux J.-C., Simoyi R.H., Swinney H.L. (1983) Observation of a Strange Attractor. Physica D8:257
13. Gao J., Zheng Z. (1994) Direct Dynamical Test for Deterministic Chaos and Optimal Embedding of a Chaotic Time Series. Phys Rev E49:3807
14. Planat M., Giorgano V., Marianneau G., Vernotte F., Mourey M., Eckert C., Miehé J.A. (1996) Is the Frequency Noise of an Oscillator of Deterministic Origin ?. IEEE Trans Ultrason Ferroelec Freq Control 43:326
15. Eckert C., Planat M., Miehé J.A. (1996) Hidden Order in the Frequency Noise of an Electronic Oscillator. Phys Rev E54:6093
16. see the companion paper by Planat M.
17. Planat M., Dos Santos S., Ratier N., Cresson J., Perrine S. (1999) Close to Resonance Interaction of Radiofrequency Waves in a Schottky Diode Mixer : 1/f Noise and Number Theory. In: Handel P., Chung A. (Eds.) Quantum 1/f Noise and other Low Frequency Fluctuations in Electronic Devices. AIP Proceedings 466:177
18. Liebert W., Schuster H.G. (1989) Proper Choice of the Time Delay for the Analysis of Chaotic Time Series. Phys Lett A142:107
19. Theiler J. (1990) Estimating Fractal Dimension. J Opt Soc Am A7:1055
20. Liebert W., Pawelzik K., Schuster H.G. (1989) Optimal Embeddings of Chaotic Attractors from Topological Considerations. Europhys Lett 14:521
21. Kennel M.B., Brown R., Abarbanel H.D.I. (1992) Determining Embedding Dimension for Phase-Space Reconstruction using a Geometrical Construction. Phys Rev A45:3403
22. Diks C. (1996) Estimating Invariants of Noisy Attractors. Phys Rev E53:R4263
23. Albano A.M., Muench J., Schwartz C., Mees A.I., Rapp P.E. (1988) Singular-Value Decomposition and the Grassberger-Procaccia Algorithm. Phys Rev A38:3017
24. Eckert C., Miehé J.A. (1996) Test of the Deterministic Conjecture for the Instantaneous Time Phase Fluctuations of an Oscillator. Ann Telecommun 51:351
25. Fredkin D.R., Rice J.A. (1995) Method of False Nearest Neighbours: a Cautionary Note. Phys Rev E51:2950

26. Smith L.A. (1988) Intrinsic Limits on Dimension Calculations. Phys Lett A133:283
27. Allan D.W. (1987) Time and Frequency (Time Domain) Characterization, Estimation and Prediction of Precision Clocks and Oscillators. IEEE Trans Ultrason Ferroelec Freq Control 34:647
28. Kostelich E.J., Yorke J.A. (1990) Noise Reduction : Finding the Simplest Dynamical System Consistent with the Data. Physica D41:183

Geometry and Dynamics of Numbers Under Finite Resolution

Jacky Cresson * and Jean-Nicolas Dénarié

Université de Franche Comté, Equipe de Mathématiques de Besançon,
CNRS-UMR 6623, France.

Abstract. We define a special set, called resolution space, which corresponds to real numbers obtained via the following resolution rule : every number greater than a given integer a is identified with ∞. This space possesses a natural scaling structure and dynamics. We introduce several notions as locking and transient resonance zones, as well as unstable irrationals numbers. This space is the natural object coming in the $1/f$ frequency noise problem. Special numbers as Markoff's irrationals are proved to play a specific role. This first criterion must be understood as a finite resolution in space for physical systems.

We then introduce an additional resolution criterion which allows only a finite construction of the previous space. A natural notion of fuzzy zone is defined. This second criterion is interpreted as a finite time experiment in physics.

1 Introduction

Many physical problems involve a careful measurement of a given quantity. For example, synchronization of oscillators is studied via a precise measurement of the frequency and amplitude of a given signal. Of course, an infinite precision measurement is impossible. This problem has little consequence in general, but as proved in recent works ([2], [11], [12]), a universal phenomenon known as $1/f$ noise comes from an explicit dependency on resolution.

The aim of this paper is to provide an understanding of the role of a *finite resolution* assumption on *real* numbers. We define a resolution space, image of an approximation function R_a which takes into account the fact that every real number greater than a given integer a is identified with ∞. In other words, we introduce a cutoff which can be understood as a finite resolution in *space*. We then construct iteratively our space via successive applications of two elementary maps on numbers : $1/x$ and $x+1$. We first remark that this construction induces a natural emergence of a *scaling* structure. We distinguish three specific zones : the first type of zone is called a *locking* resonance zone. It corresponds to an *attractive* basin of a given rational number. The second type of zone is a *transient* resonance zone. These transient zones accumulate on the boundary of locking resonance zones. The third type is the set of *unstable* points which corresponds to irrational numbers with a periodic tail (so quadratic irrational numbers).

* cresson@math.univ-fcomte.fr

M. Planat (Ed.): LNP 550, pp. 305–323, 2000.
© Springer-Verlag Berlin Heidelberg 2000

It must be pointed out that this is not the *modular group* which induces such a specific structure. Indeed, if we use the modular group in order to construct such a resolution space, then we obtain symmetric locking resonance zones with respect to rationals. But, as proved by Planat [11], such configurations are not observed. Instead, we obtain precisely the structure given by our definition of resolution space.

We discuss several particular phenomena as accumulation of locking resonance zones on irrational numbers of Markoff's type. These numbers play a special role in number theory and diophantine approximation [18].

A convenient representation of the space R_a can be obtained by the following : we represent the approximated representative of a real number. We obtain a classical object in dynamical system's theory : a *devil's staircase*. We introduce a natural hierarchy of locking resonance zones, following previous works on dynamical devil's staircase [6]. We then try to see how these locking resonance zones are classified. We then see, that when a is not too large, we obtain a classification which is approximately given by the modular group.

We then introduce a second resolution constraint. We allow only a finite number n of inversions during the recursive construction of the space R_a. It can be understood as a constraint on the time of observation. This truncation rule induces a *fuzzy* zone around irrationals in R_a, where we do not know what exactly happens.

2 Continued Fractions and Action of F_2 on Q

We recall in this section classical results dealing with continued fraction expansions and rational numbers.

To any irreducible fraction p/q of Q we associate the point (p, q) of Z^2 or, equivalently the line (additive subgroup) of Z^2 going through $(0, 0)$ and (p, q) of slope q/p of equation $py - qx = 0$. This leads to a bijection between $Q \cup \{\infty\}$ and $P^1(Z)$, the set of vector lines of Z^2, which is usually defined as the set of classes of points of $Z \times Z$ modulo the equivalence $(p, q) \sim (p', q')$ iff there exists $a \in Z$ such that $(p, q) = (ap', aq')$ or $(p', q') = (ap, aq)$.

Each line L in $P^1(Z)$ is isomorphic (as an additive discrete subgroup) to Z , and is generated by one of the two points $(p, q)(-p, -q)$ of L verifying $gcd(p, q) = 1$.

We call such points, irreducible points of Z^2. We denote by I the set of opposites irreducible points pairs $\{(p, q), (-p, -q)\}$ (Fig. 1)

We thus have the following one-to-one correspondences :

$$\begin{array}{ccccc} Q \cup \{\infty\} & \longrightarrow & P^1(Z) & \longrightarrow & P \\ \frac{p}{q} & \longmapsto & L : qy - px = 0 & \longmapsto & (p, q) \text{ with } gcd(p, q) = 1. \end{array} \tag{1}$$

The ring of 2×2 Z-matrices, $M_2(Z)$ acts naturally on Z^2. This action induces an action on Q by the so called Möbius transformations as follows :

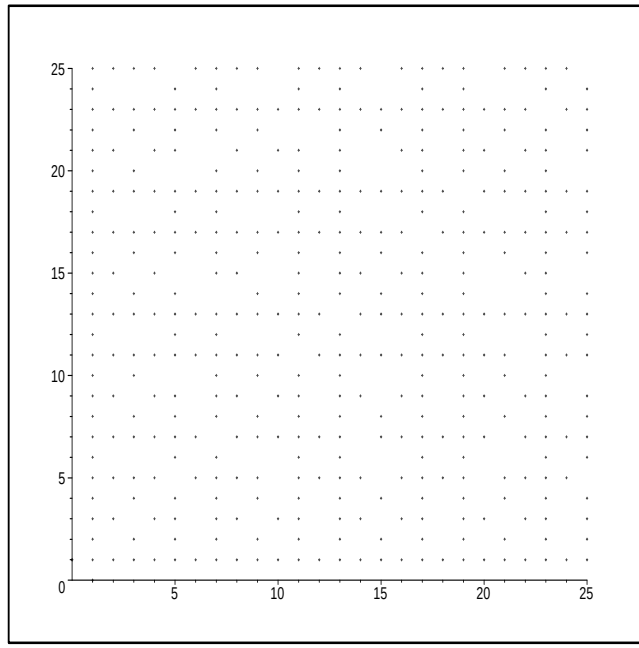

Fig. 1. Irreducible points of Z^2 $(gcd(p,q) = 1)$.

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \ (p,q) \ \longmapsto \ (ap+bq, cp+dq)$$
$$\downarrow \qquad\qquad \downarrow$$
$$z = \frac{p}{q} \overset{A}{\longmapsto} \frac{ap+bq}{cp+dq} = \frac{az+b}{cz+d}. \tag{2}$$

It can be shown that A preserves P if and only if $|\det(A)| = 1$ or equivalently, A invertible in $M_2(Z)$. Thus, we consider the action of the group $GL_2(Z)$ of invertible 2×2 Z-matrices on Q via Möbius transformations.

Notice that $GL_2(Z)$ contains $PSL_2(Z)$ (determinant one) as a subgroup of index two and also that, since $A \in GL_2(Z)$ preserves P, $gcd(p,q) = 1$ implies $gcd(ap+bq, cp+bq) = 1$ and, thus, the fraction $\dfrac{ap+bq}{cp+dq}$ is irreducible.

2.1 Continued fractions and F_2^+

We recall some basic results on continued fractions. We refer to Khintchine [8] for more details.

Let $(a_0, a_1, \dots, a_n)$ be a finite sequence of integers with $a_n \neq 0$. We denote by $[a_0, \dots, a_k]$ the finite continued fraction

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ \dots \ + \cfrac{1}{a_k}}}. \tag{3}$$

In the case of an infinite sequence, we write $[a_0, \dots, a_k, \dots]$ to denote the associated infinite continued fraction.

Assume that each $a_k > 0$, for $k > 0$, then each irrational has a unique representation as an infinite continued fraction. Nevertheless, each rational has two such representation given by:

$$[a_0, \dots, a_k] = [a_0, \dots, a_k - 1, 1] \tag{4}$$

We deduce that each rational x has a unique representation as a finite continued fraction with k odd. By the way, there also exists a unique representation as a continued fraction of even length. Notice that this (odd) continued fraction always gives the irreducible writing of the rational x, and thus the continued fraction represents an irreducible pair of I.

Those continued fractions can be built by the two Möbius transformations,

$$A : x \mapsto x + 1 \text{ and } I : x \mapsto 1/x, \tag{5}$$

applied to the irreducible fraction $1/1$.
The continued fraction $[a_0, \dots, a_n]$ is written as $A^{a_0} \circ I \circ A^{a_1} \circ \cdots \circ I \circ A^{a_n - 1}(1)$ where $(1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. To work with unimodular (determinant one) transformations, we represent each rational by its odd continued fraction. Indeed, $\det(A) = 1$ and $\det(I) = -1$, and the odd representation has an even number of inversions. Regrouping inversion two by two, and introducing the tranformation $B = I \circ A \circ I$, we obtain:

Lemma 1. *Any positive rational number admits a unique expressions:*

$$[a_0, a_1, \dots, a_{2n}] = A^{a_0} \circ B^{a_1} \circ \cdots \circ A^{a_{2n} - 1}(1). \tag{6}$$

The matrices A and B are:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{7}$$

They generate the modular group $PSL_2(Z)$ [19] of unimodular Möbius transformations which is isomorphic to the free group of rank two F_2, A and B being two free generators. Denoting by F_2^+ the semi-group of words written with positive powers of A and B, we have the following

Corollary 1. *The application of* Q^{+*} in F_2^+ *defined as:*

$$[a_0, a_1, ...a_{2n}] \longmapsto A^{a_0} \circ B^{a_1} \circ ... \circ A^{a_{2n}-1} \tag{8}$$

is a bijection.

Moreover F_2^+ *acts on* Q *by left multiplication.*

This is a simple reformulation of the previous lemma. In particular, this establishes a bijection between N^*, the set of positive numbers, and the set of words A^i with $i \geq 0$.

2.2 Geometry of continued fractions

The words B^i, with $i \geq 0$, applied on 1, generate the fractions $1/n$ with $n \in N^*$. Going back to the representation of irreducible fractions by points (p, q) of Z^2, with $gcd(p, q) = 1$, we represent these two sets as two affine half lines, L_∞ and L_0 coming out of $(1, 1)$, and of respective slope 0 and ∞.

We then let F_2^+ (it's a linear group) act on Z^2, and observe the images of those two half-lines. We obtain the following Fig. 2.

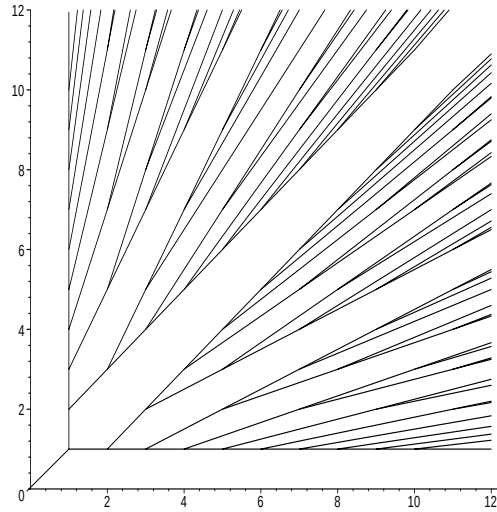

Fig. 2. A plane Farey tree

The striking facts are that it is a binary tree, $\mathcal{T}$, that it is nodes are exactly the points (p, q) with $gcd(p, q) = 1$, naturally associated with the irreducible

factions p/q, and therefore indexed by Q and, moreover, that each point (except $(1,1)$), is both living on a straight branch (called cross branch) and the root point of a new springing straight branch. Moreover, each of those branches is exactly the image, by a uniquely determined word of F_2^+, of one of the two half lines L_0 and L_∞. Leaving aside the proofs of those facts, we just give the following properties, dealing with continued fractions :

Lemma 2. *If $\frac{p}{q} \neq 1$ is a node of $\mathcal{T}$, and $[a_0, a_1, \ldots, a_{2n}]$ its odd continued fraction.*

- The root of its cross branch is: $[a_0, a_1, \ldots, a_{2n-1}, 1]$ *if* $a_{2n} > 1$,
$[a_0, a_1, \ldots, a_{2n-2} + 1]$ *if* $a_{2n} = 1$.
- The slope of this cross branch is
$[0, a_0, a_1, \ldots, a_{2n-1} - 1, 1] = [0, a_0, a_1, \ldots, a_{2n-1}]$ *if* $a_{2n} > 1$,
$[0, a_0, a_1, \ldots, a_{2n-2}]$ *if* $a_{2n} = 1$.
- As a consequence, the slope of (p, q)'s springing branch is
$[a_0, a_1, \ldots, a_{2n} - 1]$ *if* $a_{2n} > 1$.

Proof. We admitted that the cross branch of (p, q), call it C, is exactly the image of L_0 or L_∞ by a word of F_2^+.

We suppose that C is the image of L_∞, the horizontal half line, by the word M. The root of C is then $M(1)$, and p/q is $M(l) = M \circ A^{l-1}(1)$ with $l \in L_0 \simeq N$ (and $l > 1$ since (p, q) is not the root of C).

First, M doesn't end with an A. If it did, call N its subword obtained by cutting the last A. $N(1)$ would belong to C and $M(1)$ wouldn't be the root.

Second, $l = a_n$. Indeed, $p/q = M \circ A^{l-1}(1) = A^{a_0} \circ B^{a_1} \circ \cdots \circ A^{a_{2n}-1}(1)$, and, with 1 we obtain $l = a_n$.

As a consequence $M = A^{a_0} \circ B^{a_1} \circ \cdots \circ B^{a_{2n-1}}$ and $M(1) = A^{a_0} \circ B^{a_1} \circ \cdots \circ B^{a_{2n-1}} \circ A^0(1) = [a_0, a_1, \ldots, a_{2n-1}, 1]$ which establishes the first point.

The case C is the image of L_0 by M is treated the same way, replacing B by A. (and M ends with a B).

We skip to the second point which becomes clearer if we notice that L_∞ is parallel to the horizontal axe, going through $(0,0)$ and $(1,0) = B^{-1}((1,1))$. $(0,0)$ is fixed by M (it's linear), and $M((1,0)) = M \circ B^{-1}((1,1)) = A^{a_0} \circ B^{a_1} \circ \cdots \circ B^{a_{2n-1}-1}(1,1)$.

Therefore, the slope of the image of the horizontal line (and thus of B) is the rational associated to $M((1,0))$ which is $[a_0, a_1, \ldots, a_{2n-1} - 1, 1]$ as announced.

For the third point, just think the springing branch as a cross branch of an appropriated rational. Notice that this branch will then be an image of L_0, and the same proof should be done, replacing B by A and $(1,0)$ by $(0,1)$.

This tree, which is inspired by the Farey tree on the hyperbolic disk, gives us a good description of Q, for continued fractions calculus (but very complex if we wish to translate, in this language, multiplication by 2!). The first point of the last lemma, for example, shows that we can obtain the

successive truncated fractions of a given one, just by sliding down the tree from springing point to another.

The two next ones gives us another information: take a rational $p/q = [a_0, a_1, \ldots, a_{2n}]$, and his two sons, going up in the tree: $x_1 = [a_0, a_1, \ldots, a_{2n} + 1]$ and $x_2 = [a_0, a_1, \ldots, a_{2n} - 1, 1, 1]$. The two springing branches of those points are then parallels of slope q/p. Those branches (with the two segments $[p/q, x_1]$ and $[p/q, x_2]$ delimit a domain $D(p/q)$ of the plane which will never be crossed by any other branch of the tree.

Theses domains pave the plane. (look Fig. 2). The only vector line out of (0,0) which, once entered, never leaves it, is the line $L(p/q) : py - qx = 0$, associated to p/q as previously. Thus the only point of $P_1(Z) = Q$ "contained" in $D(p/q)$ is p/q, and we can say, in this sense, that it's wideness is 0. We'll see in the next section, that those domains open when we restrict the set of available words of F_2^+, truncating the tree.

3 Arithmetic Resolution Space

The upper construction gives a description of $R^+ \cup \{\infty\}$, with infinite resolution. The rational being associated to finite words of F_2^+, and irrational (and ∞) to infinite words. We now give a way to introduce finite resolution, keeping consistency with continued fractions. This question arises naturally in some physical problems.

Let a be a positive integer. We assume that the following finite resolution criterion is satisfied: every real number $x \geq a$ is identified with ∞. This means that the only words of F_2^+, of type $A^i = \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}$, we tolerate are the words A^i with $i < a$. The other words being mapped to the limit word A^∞.

Impose the same resolution condition near 0, mapping the words $B^i = \begin{pmatrix} 1 & 0 \\ i & 1 \end{pmatrix}$ to B^∞, for any $i > a$. We want those resolution conditions to be imposed near any rational of Q^+. Let $F_2^+(a)$ be the set of words of F_2^+ containing no subword A^i or B^i with $i > a$, and define the map :

$$\begin{aligned} R_a : R^+ \cup \{\infty\} &\longrightarrow R^+ \cup \{\infty\} \\ M(1) &\longmapsto M_a(1) \end{aligned} \tag{9}$$

where M_a is the word obtained from M, by replacing in M the first word A^i or B^i with $i > a$ by, respectively A^∞ and B^∞. Notice many words are fixed by R_a, as for example, the word infinite $ABABA....$ The tree corresponding to $F_2^+(a)$ is given in Fig. 3. Observe that the domains of any node of this tree (in the sense defined in previous section) are more open and the numbers associated to the vector lines "contained" in it are sent, by R_a to the rational corresponding to the node.

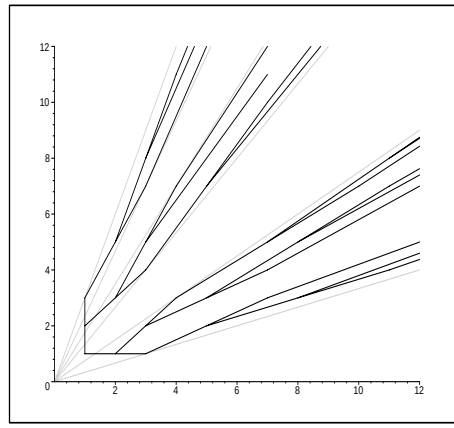

Fig. 3. The truncated tree (for $a = 2$)

The approximation function R_a defined above has a rather complicated structure with many autosimilarities. It leads, by iterations, to a dynamical system on numbers. R_a is represented below (Fig. 4).

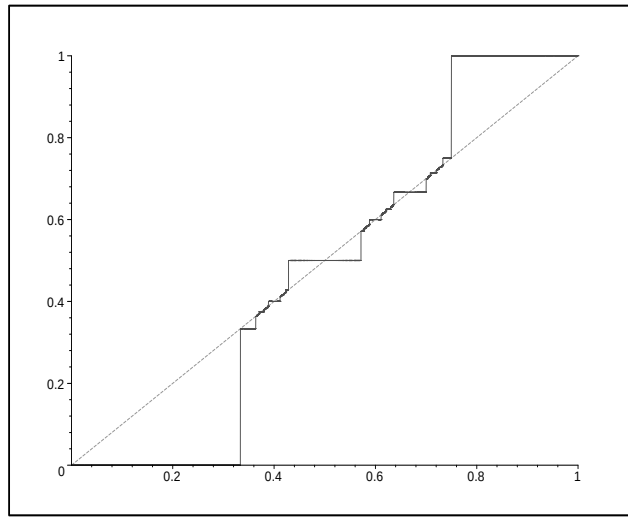

Fig. 4. Approximation function (for $a = 3$).

We can characterize the different horizontal landings of this function: *locking* resonance zones, which are basins of attraction where all real numbers are approximated by the rational defining this zone, itself being fixed (locked) by R_a;

transient resonance zones which accumulate on each side of any locking zone, but only possess a semi-basin of attraction (we say also that they are semi-stable). The numbers in these transient zones (including p/q) only reach their locking rational after several iterations of R_a.

These landings can be built on an iterative way, as images of the two fundamental attractive zones: $[a, \infty[\to \infty$ and $]0, 1/a] \to 0$, by the maps of $F_2^+(a)$, written as iterations of A and B.
Recall $A : x \mapsto x+1$ and $B : x \mapsto 1/x \mapsto 1+1/x \mapsto 1/(1+1/x) = x/(x+1)$.

For example $A(]0, 1/a] \to 0)$ gives $]1, 1+1/a] \to 1$ and $B([a, \infty[\to \infty)$ gives $[a/(a+1) = 1-1/(a+1), 1[\to 1$, which builds the locking basin of 1. The border of these zones will later be called $\nu^+(1)$ and $\nu^-(1)$. An analogous construction of the transient zones on the left the locking zone of 1 is given later in Sect. 4.2:

By iterating the previous maps (at most a consecutive A and B are tolerated), these dynamics build the resolution function R_a of Fig. 4. The two types of dynamics (transient and locking) can also be observed by a close look at the function error of resolution: $|R_a(x) - x|$ given on Fig. 5.
We define the resolution space in a, to be the image of $R^+ \cup \infty$ by R_a.

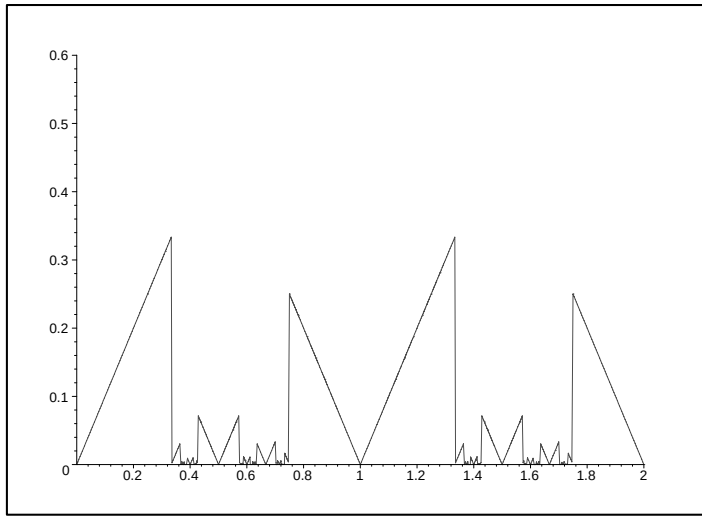

Fig. 5. Resolution error, $a = 3$.

We can now state the following theorem :

Theorem 1. *Let $a \in \mathbf{N}$ be given, the resolution set $Im(R_a)$ has the following structure :*

(i) locking rational p/q: corresponding to a locking interval, with rational boundaries $\nu^+(p/q)$ and $\nu^-(p/q)$ (with $p/q \in]\nu^-(p/q), \nu^+(p/q)[$) such that for all $x \in [\nu^-(p/q), \nu^+(p/q)]$, the resolution of x is $R_a(x) = p/q$.

(ii) transient rational p/q: corresponding to a transient interval with boundaries $]\nu(p/q), p/q[$ such that, for all $x \in]p/q, \nu(p/q)]$, the resolution of x is $R_a(x) = p/q$. (and its not the case for any $x \in]0, p/q]$

(iii) irrational obtained as accumulation of locking resonance zones.

(iv) irrational obtained as accumulation of transient and locking resonance zones.

As already pointed out in the introduction, this set occurs in $1/f$ frequency noise. Next section deals with some particular properties of this space.

3.1 An analogue of the dynamical devil's staircase in number theory

The devil's staircase is a common object in dynamical system theory. Briefly, synchronization can be studied via one parameter family of diffeomorphism of the circle [13]. One can take the Arnold's model [3] defined by $f_\omega(\theta) = \theta + \omega + \epsilon \sin(\theta)$ mod. 2π for $\theta \in S^1$. When $\epsilon = 0$, we have a *rigid* rotation of frequency ω. When $\epsilon \neq 0$, we can study the behavior of f_ω via the *rotation number* $\rho(\omega)$, which corresponds to a *mean* frequency associated to the system (see [5], [6]). Periodic behaviors of the system are given by *rational* rotation numbers. One can prove that, each rational number p/q gives rise to an interval in ω of constant rotation number, of size ϵ^q (see [4], [6]). If we picture the value of the rotation number versus ω for a fixed value of ϵ, then we obtain a devil's staircase [6].

In this case, a natural hierarchy of zones in ω of constant rational frequency numbers is given by *Farey's* numbers [6]. Indeed, as the length in ω of the p/q zone is of order ϵ^q, these zones are classified following Farey's hierarchy.

What's about our devil's staircase ? First, when a is not to large, we obtain for locking zones of resonance in the unit interval $[0, 1]$, a hierarchy (with respect to the length of these zones) which can be approximately described via the modular group. The typical value of a where we can take the modular group to classify locking resonance zones is $a = 4$. When a is greater than 4, locking resonance zones are *not* at all classified by the modular group.

We emphasize that the resolution space corresponding to physical and experimental results about $1/f$ noise is generated by our approach. This means that the *modular* group is not the natural object associated to resolution. Elementaries maps as A and I do not define a group at all. Resolution induces specific indempotent properties like $A^p = A^a$ for all $p \geq a$.

4 Behavior of Continued Fractions with Bounded Partial Quotients

In this section we characterize the various dynamical behaviors of numbers, under iterations of R_a, using their continued fractions.

4.1 On the boundary of locking resonances

Let $x = p/q$ be a locked number. Following Planat [11], we denote by $\nu^+(x)$ (resp. $\nu^-(x)$), the rational number which defines the right (resp. left) boundary of the locking zone of x. We have the following elementary properties:

Lemma 3. *For any locked number, we have*

$$\begin{array}{ll} \nu^\sigma(x+1) = 1 + \nu^\sigma(x), & \sigma = \pm, \\ \nu^\sigma(1/x) & = 1/\nu^{-\sigma}(x), \end{array} \tag{10}$$

These functional equations allow us to define ν^σ, $\sigma = \pm$ for all rational numbers.

Lemma 4. *The zone of locking resonance for a given rational*

$$\frac{p}{q} = [a_0, a_1, \ldots, a_k],$$

is determined by the following boundary numbers

$$\begin{array}{l} \nu = [a_0, \ldots, a_k, a], \\ \beta = [a_0, \ldots, a_k - 1, 1, a]. \end{array} \tag{11}$$

Proof. We have

$$\nu^+([a_0, \ldots, a_k]) = \nu^+(a_0 + \frac{1}{[a_1, \ldots, a_k]}) = a_0 + \nu^+(\frac{1}{[a_1, \ldots, a_k]}. \tag{12}$$

Then,

$$\nu^+([a_0, \ldots, a_k]) = a_0 + \frac{1}{\nu^-([a_1, \ldots, a_k])}, \tag{13}$$

and, by induction we obtain

$$\nu^+([a_0, \ldots, a_k]) = [a_0, \ldots, a_{k-1}, \nu^\sigma(a_k)], \tag{14}$$

where $\sigma = +$ if k is odd and $-$ otherwise.

As $\nu^+(a_k) = a_k + \dfrac{1}{a}$, and $\nu^-(a_k) = a_k - 1 + \dfrac{1}{1 + \dfrac{1}{a}}$, we obtain either ν or β for $\nu^\sigma([a_0, \ldots, a_k])$, $\sigma = \pm$.

This result deals with *diophantine approximation* via classical relation between continued fractions and best approximation of real numbers. We point out that from the physical view point, continued fraction and resolution arguments come from diophantine approximation problems (see [8]).

4.2 Dynamics of transient resonance zones

$BA^{a-1}([1-1/(a+1),1[\to 1) = B([a-1+a/(a+1),a[\to a) = [B(a-1+a/(a+1)),a/(a+1)[\to a/(a+1)$, which sticks on the left of $[a/(a+1) = 1-1/(a+1),1[\to 1$ attractive zone.

Transient resonance zones can be obtained via a simple recursive formula. We first construct the *primary* transient resonance zone, given by the partial locking resonance zone: take the locking interval $[a/a+1,1]$, and make $x \mapsto x+a-1$.

We obtain the interval $[a-1/a+1,a]$. By taking the image under $x \mapsto 1/x$ of this interval, we obtain the primary resonance zone.

We then produce an infinite sequence of transient resonance zones by applying the following rule on each generation of transient resonances: do $x \mapsto x+1$, $x \mapsto 1/x$, $x \mapsto x+a-1$ and $x \mapsto 1/x$ (see Fig. 6).

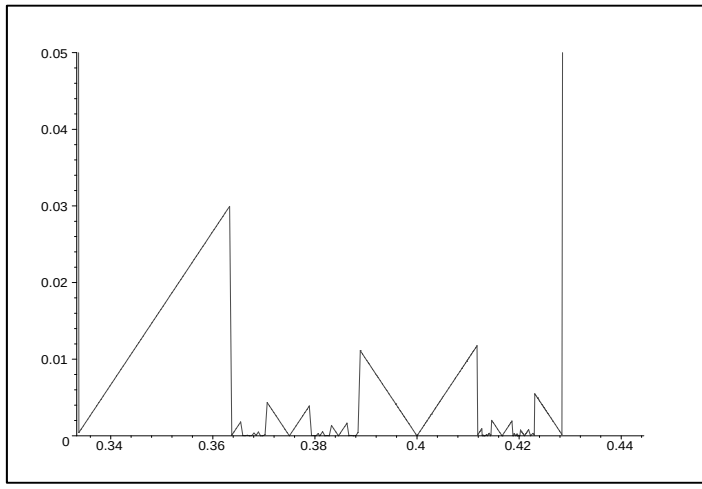

Fig. 6. Accumulation of locking and transient resonances, near $\sqrt{3}-1$ for $a=3$.

Then, we must iterate the following map

$$P = \begin{pmatrix} 1 & 1 \\ a-1 & a \end{pmatrix}, \tag{15}$$

on the primary transient resonance.

We can easily prove that this map is hyperbolic. By iteration, it converges toward a fixed point of P^n. Precisely, we have

Lemma 5. *By iteration of P we obtain an irrational number such that its tails, at a given depth, becomes periodic of period 2, given by*

$$[\dots,1,a-1,1,a-1,1,a-1,1,a-1,\dots]. \tag{16}$$

We then obtain a *quadratic* irrational number.

These points are called *unstable*. Let x be a fixed point of P (i.e. $P(x) = x$). Then, x is solution of $(a-1)x^2+(a-1)x-1=0$. For example, for $a=2$, we obtain $x = \dfrac{-1+\sqrt{5}}{2}$ which is the golden number. The stability of this fixed point is given by the sign of $P'(x) = \dfrac{1}{((a-1)x+a)^2}$. Since $P'(x) > 0$, this point is unstable from the dynamical viewpoint.

Assume that we take a small random perturbation of this unstable point. Then, we can go, with equal chance, toward two different locking zone of resonance. Then, if a system depends on such a number, we will have a high sensibility to small perturbation in this zone, which is not the case for locking or transient zone of resonances.

4.3 Accumulation of locking resonance zones

With respect to the previous lemma on the behavior of transient resonance zones under iteration of the map P, we can try to do the same for locking zone of resonances.

The basic idea is to look at a recursive construction of the rationals defining locking resonance zones. We restrict ourselves these rationals in the interval $[0,1]$.

At the first step, they are obtained by $1/x$ from the integer $2,3,\ldots,a-1$. We then obtain at step one the following set of primary resonances

$$R_1 = \{1/(a-1),\ldots,1/3,1/2\}. \tag{17}$$

To obtain the second generation of locking resonances, we must do a translation of the form $x \mapsto x+t$ where $1 \le t \le a-1$. Then, we take $1/x$. The elementary map is defined by

$$A_t = \begin{pmatrix} 0 & 1 \\ 1 & t \end{pmatrix}. \tag{18}$$

Let us first see what happens when we iterate a given number from R_1 by A_t, which of course is the most elementary kind of resonance that we can produce from R_1 (as one must, in general, allow different combination of A_t in the iterate). We then have the following accumulation lemma.

Lemma 6. (accumulation lemma) *Any rational* $1/x$, $2 \ge x \le a-1$ *of* R_1 *converges toward an irrational number of the following form*

$$\bar{t} = [t,t,\ldots,t,\ldots]. \tag{19}$$

At each generation of resonances, the action of A_t put a new t in the sequence defining the continued fraction of the previous generation step.

For example, take $1/x$, which can be written $[0, x]$. By A_1, we obtain $1/(1 + 1/x)$ which takes the form $[0, 1, x]$. By induction, the i-generation of resonances is given $[0, 1, \ldots, 1, x]$ with i times 1 in the continued fraction. Then, when $i \to \infty$, we obtain an irrational number of the form given in the lemma. This irrational number is the *golden* number. Precisely, the golden number is the best approximated irrational number by rational. It is defined by $\dfrac{-1+\sqrt{5}}{2}$ and its continued fraction expansion is given by $[1, \ldots, 1, \ldots]$. So, by induction of A_1, we obtain the set of poorly approximated irrational number by rational (in the continued fraction sense).

Let x be an irrational as in the lemma. It is a fixed point of A_t. Then, it is a solution of $x^2 + tx - 1 = 0$, which are given by $x = \dfrac{-t+\sqrt{t^2+4}}{2}$. The stability of this point is given by the sign of $A'_t(x)$. As $A'_t(x) = -x^2$, this point is stable.

The previous lemma can be generalized for other inductive scheme. We must allow combination like $A_1 \circ A_{a-1} \circ A_2$ etc., in order to obtain a complete description of the accumulation limit set. Nevertheless, no new phenomenon occurs for these kind of combination, and they are all, more or less, relevant of the "basic" accumulation lemma.

4.4 Approximation property

We shall say that ξ is approximable by rational to order n if there is a $K(\xi)$, depending only on ξ, for which

$$\left|\frac{p}{q} - \xi\right| < \frac{K(\xi)}{q^n}, \tag{20}$$

has an infinity of solutions.

A rational is approximable to order 1 and to no higher order. A quadratic irrational is approximable to order 2 and to no higher order.

Theorem 2. *An irrational real number with bounded partial quotients is approximable to order 2 and to no higher order.*

A conjecture asserts that: an irrational real number with bounded partial quotients is either quadratic or transcendental. We refer to [1] for more details and to [14] for an example of transcendental number with bounded partial quotients.

By theorem 2, two irrational real numbers with bounded partial quotients are distinguished by their constant of approximation, $K(\xi)$. It is possible to describe the spectrum of this constant. We obtain the so called Markoff's spectrum. We refer to [18] and [16] for more details.

4.5 Length of locking resonance zones

The previous section allow us to compute the size of the locking resonance zone for each rational. We do this by induction on the generation of locking resonant rationals. We will see, in the last section, that these length play an important role in physics for the understanding of $1/f$ noise fluctuation.

We define the *length* function as

$$l(x) = \nu^+(x) - \nu^-(x). \tag{21}$$

We have

Lemma 7. *The length function satisfies*

$$\begin{aligned} l(x+1) &= l(x), \\ l(1/x) &= \frac{l(x)}{\nu^+(x)\nu^-(x)}. \end{aligned} \tag{22}$$

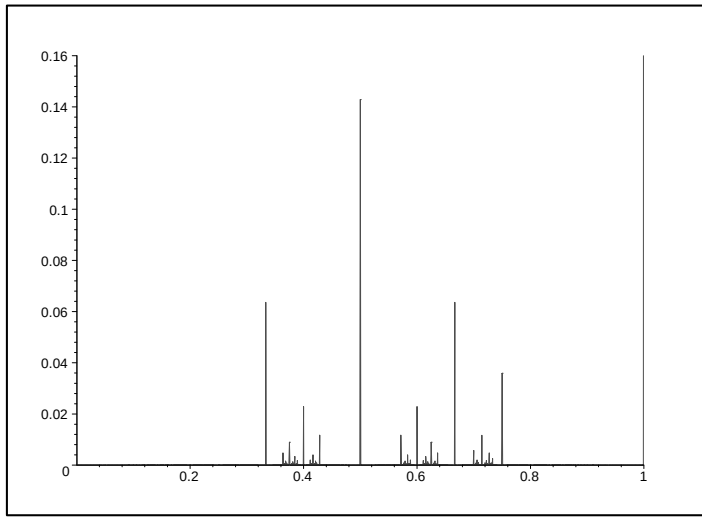

Fig. 7. Length of locking resonance zones

The preceding picture (Fig.7)leads us to the following problem: is this figure symmetric with respect to $x = 1/2$? If you look at this picture you can see on the right a little part which has no symmetric counterpart. Nevertheless, all the rest seems symmetric. The following lemma gives a complete description of this phenomenon.

First, we prove that the length function is symmetric with respect to $x = 1/2$.

Lemma 8. *For $x \in [2, \infty]$, we have*

$$l\left(\frac{x-1}{x}\right) = l\left(\frac{1}{x}\right). \tag{23}$$

Proof. For all $x \in [1, \infty[$, we have via lemmas 3, 7

$$\begin{array}{ll} \nu^\sigma(x-1) & = \nu^\sigma(x) - 1, \\ l(x-1) & = l(x). \end{array} \tag{24}$$

We note that $\dfrac{x}{x-1} = 1 + \dfrac{1}{x-1}$, then

$$l\left(\frac{x-1}{x}\right) = \frac{l(x/x-1)}{\nu^+(x/x-1)\nu^-(x/x-1)} = \frac{l(1+(1/x-1))}{\nu^+(1+(1/x-1))\nu^-(1+(1/x-1))}. \tag{25}$$

We obtain

$$\begin{aligned} l(x-1/x) &= l(1/x-1)\frac{1}{\nu^+(1/x-1)\nu^-(1/x-1)} \\ &= l(x-1)\frac{1}{\nu^+(x-1)\nu^-(x-1)}\frac{1}{\nu^+(1/x-1)\nu^-(1/x-1)}, \end{aligned} \tag{26}$$

and finally,

$$l(x-1/x) = l(x-1)\frac{1}{(\nu^+(x-1)+1)(\nu^-(x-1)+1)} = \frac{l(x)}{\nu^+(x)\nu^-(x)} = l(1/x). \tag{27}$$

So the length function is symmetric with respect to $x = 1/2$. What about the disymmetry of the picture ?

First, we state a technical lemma, which gives the continued fraction expansion of $1 - x$ from that of x.

Lemma 9. *If $x = [0, a_1, a_2, \dots]$ with $a_1 \geq 2$, then we have $1 - x = [0, 1, a_1 - 1, a_2, \dots]$.*

Proof. If $x = [0, a_1, a_2, \dots]$, then $1/x = [a_1, a_2, \dots]$. We note that $\dfrac{1}{1-x} = 1 + \dfrac{1}{\frac{1}{x} - 1}$. Then, we obtain

$$\frac{1}{1-x} = [1, a_1 - 1, a_2, \dots]. \tag{28}$$

Hence, by inversion, we have

$$1 - x = [0, 1, a_1 - 1, a_2, \dots]. \tag{29}$$

We are now able to specify points of $[1/2, 1]$ with no symmetric. We have

Lemma 10. *If $x \in [1/2, 1]$ with its continued fraction expansion of the form $x = [0, 1, a_1, a_2, \dots]$, with $a_1 = a_{\max} - 1$, then $1 - x$ belongs to the transient zone of 0.*

Proof. If $a_1 = a_{\max} - 1$, then $1 - x = [0, a_{\max}, a_2, \dots]$ and by the truncation rule, this point is identified with 0. Then, x, which possess a locking zone of resonance, has no symmetric in $R_{a_{\max}}$.

By lemma 10, all the locked rational in the interval defined by $[0, 1, a_{\max} - 1]$ and $[0, 1, a_{\max} - 1, 1, a_{\max} - 1, \dots$ possess no symmetric counterpart. This explain the existence of a single zone on the right which breaks the symmetry.

We emphasize that this property reflects the fact that the modular group is not adapted to classify locking zone of resonances.

5 Truncated Resolution Space

Despite our first resolution constraint, our space contains irrational numbers. From a physical viewpoint, this is not at all realistic, and one need to introduce a second natural constraint in order to provide a space which can really occur in physics.

The basic idea is the following. When we construct our resolution space by induction, we have a natural notion of *generation* of resonances (as used for example in the section about the accumulation of locking resonance zone). More or less, this notion of generation counts the number n of invertion that we allow for the construction of locking resonance zone. From the physical view point, this must be understood as a finite time observation. Then truncation in a is a *space* resolution and truncation in n is a *time* resolution.

Let us consider an integer n. We restrict ourselves to the structure of the resolution space on $[0, 1]$.

Definition 1. The *truncated resolution space* of order n, denoted by R_a^n, is the set of rational number in $[0, 1]$ obtained by iteration of $A : x \mapsto x + 1$ and $I : x \mapsto 1/x$ on the set $R_1 = \{1/(a-1), \dots, 1/2\}$ with only n iterations of I. Precisely, for each rational of R_1 we allow map of the form

$$I \circ A^{s_1} \circ I \circ A^{s_2} \circ I \circ A^{s_3} \circ I \circ \cdots \circ A^{s_n}, \tag{30}$$

with $1 \leq s_i \leq a - 1$.

We then obtain the following picture for $a = 3$ and $n = 2$.

So, there exists *fuzzy* zone between lockings and transient resonances zones in which we do not know what happens really, as we have not get any information on this part via our construction. Finite resolution in time induces natural fuzzy zones where one cannot say what the system really do.

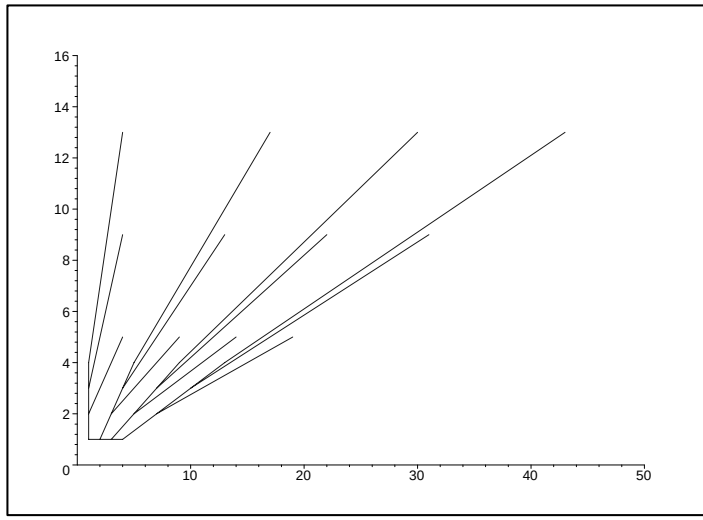

Fig. 8. Truncated resolution space.

References

1. Allouche, J-P, Algebraic and analytic randomness, this volume
2. Nottale, L, Scale-relativity and quantization of the universe I. Theoretical framework, Astron. Astrophys. 327, 867-889, (1997).
3. Arnold, V,I, Small denominators I: On the mapping of a circle onto itself, Trans. Amer. Math. Soc. Ser. 2, 46 (1965) 213-284.
4. Arnold, V,I, Remarks on the perturbation theory for problems of Mathieu type, Russian Math. Surveys 38:4 (1983), 215-233.
5. Yoccoz, J-C, An introduction to small divisors problems, in "*From number theory to physics*", Springer-Verlag, New-York, 659-679, (1992).
6. Cvitanovik, P, Circle maps : irrational winding, in "*From number theory to physics*", Springer-Verlag, New-York, (1992).
7. Hardy, G, Wright, E, An introduction to the theory of numbers, Oxford University Press, Amen House, London E.C.4, (1965).
8. Kintchine, A. Ya., Continued fractions, P. Noordhoff Ltd., Groningen 1963.
9. G.A. Jones, D. Singerman, K. Wicks, The modular group and generalized Farey graphs, in Groups St Andrews, 1989, London Math. Soc. Lectures Notes 160, 316-338, (1991).
10. G. Jones, D. Singerman, Maps, hypermaps and triangle groups, in The Grothendieck theory of Dessins d'enfants, London. Math. Soc. Lectures notes series 200, (1994).
11. Planat, M, Dos Santos, S, Ratier, N, Cresson, J, Perrine, S, Close to resonance interaction of radiofrequency wawes in a Schottky diode mixer: $1/f$ noise and number theory, in "*Quantum* $1/f$ *noise and other low frequency fluctuations in electronic devices*", edited by A. Chung and P. Handel, AIP Press (1999) 177-187.

12. Planat, M, Dos Santos, S, Cresson, J, Perrine, S, $1/f$ frequency noise in a communication receiver and the Riemann hypothesis, in "15*th International Conference on Noise in Physical Systems and* $1/f$ *Fluctuations*", edited by C. Surya, Bentham Press (1999) 409-412.
13. Planat, M, $1/f$ frequency noise in a communication receiver and the Riemann hypothesis, this volume.
14. Queffélec, Transcendance des fractions continues de Thue-Morse, Jounal of Number theory 73, 201-211, (1998).
15. Serre, J-P, A course in Arithmetic, Graduate text in Mathematics 7, Springer-Verlag, New-York-Heidelberg-Berlin, 1973.
16. Series, C, The geometry of Markoff numbers, The Mathematical Intelligencer Vol.7,no.3,20-29, (1985).
17. Apostol, T, M, Modular functions and Dirichlet series in number theory, Graduate Text in Mathematics 41, Springer-Verlag, 1976.
18. Perrine, S, Generalized Markoff theories and the associated conformal geometries, this volume.
19. Cohen, P, On the modular function and its importance in arithmetic, this volume
20. Haas, A, Diophantine approximation on hyperbolic Riemann surfaces, Acta Mathematica 156, 33-82, (1986).

Diophantine Conditions and Real or Complex Brjuno Functions

Pierre Moussa[1] and Stefano Marmi[2]

[1] Service de Physique Théorique, CEA/Saclay, F-91191 Gif sur Yvette cedex, France
[2] Dipartimento di Matematica "U.Dini" Università di Firenze, Viale Morgagni 67/A, I-50134 Firenze, Italy

Abstract. The continued fraction expansion of the real number $x = a_0 + x_0$, $a_0 \in \mathbb{Z}$, is given by $0 \leq x_n < 1$, $x_n^{-1} = a_{n+1} + x_{n+1}$, $a_{n+1} \in \mathbb{N}$, for $n \geq 0$. The Brjuno function is then $B(x) = \sum_{n=0}^{\infty} x_0 x_1 \dots x_{n-1} \ln(x_n^{-1})$, and the number x satisfies the Brjuno diophantine condition whenever $B(x)$ is bounded. Invariant circles under a complex rotation persist when the map is analytically perturbed, if and only if the rotation number satisfies the Brjuno condition, and the same holds for invariant circles in the semi-standard and standard map cases. In this lecture, we will review some properties of the Brjuno function, and give some generalisations related to familiar diophantine conditions. The Brjuno function is highly singular and takes value $+\infty$ on a dense set including rationals. We present a regularisation leading to a complex function holomorphic in the upper half plane. Its imaginary part tends to the Brjuno function on the real axis, the real part remaining bounded, and we also indicate its transformation under the modular group.

1 Hamiltonian Chaos and the Standard Map

The simplest non-trivial model for Hamiltonian chaos is a two dimensional real map, called the "Standard Map". It has been introduced more or less independently by Chirikov and Taylor [1,2]. The occurrence of chaos was discussed by Greene [3] who displayed many numerical results on this model, which describes a simplified version of the non–linear coupling of two oscillators. It occurs naturally in many domains of physics, including celestial mechanics, classical quasiperiodic systems, quantum quasicrystals, adiabatic response in non–linear mechanics, magnetic toroidal configurations in plasma physics, non–linear electronic devices, and many others.

The Standard Map is a map from the cylinder $\mathbb{T} \times \mathbb{R}$ to itself, defined as

$$\big(\theta, r\big) \mapsto \big(\theta', r'\big) = \left(\theta + r + \frac{K}{2\pi}\sin(2\pi\theta) \pmod 1 \ , \ r + \frac{K}{2\pi}\sin(2\pi\theta)\right) . \tag{1}$$

Note that the second variable can also be taken modulo 1, in which case we get a map $\mathbb{T}^2 \to \mathbb{T}^2$. The above map can be written in two equivalent forms:

M. Planat (Ed.): LNP 550, pp. 324–342, 2000.
© Springer-Verlag Berlin Heidelberg 2000

- Hamiltonian form

$$r' - r = \frac{K}{2\pi} \sin(2\pi\theta) \quad , \quad \theta' - \theta = r' \, , \tag{2}$$

- Lagrangian form, where one considers now the twice iterated map, that is $(\theta, r) \to (\theta', r') \to (\theta'', r'')$, so that

$$\theta'' - 2\theta' + \theta = \frac{K}{2\pi} \sin(2\pi\theta') \, , \tag{3}$$

and for the n-th iterated map one gets

$$\theta_{n+1} - 2\theta_n + \theta_{n-1} = \frac{K}{2\pi} \sin(2\pi\theta_n) \, . \tag{4}$$

This last equation is sometimes called the Frenkel-Kontorova model [4] which describes equilibrium positions of a chain of material points placed in a periodic potential and submitted to an harmonic elastic force between two neighbouring points.

If $K = 0$, we get the so-called "twist-map", which gives

$$r_{n+1} = r_n = r_0 = \rho = \text{constant} \quad , \quad \theta_{n+1} = \theta_0 + n\rho \pmod 1 \, , \tag{5}$$

after n iterations. This map is nothing but a rotation of angle $2\pi\rho$–we say that the rotation number is ρ. The orbits are all transverse to the axis of the cylinder. They are made of a finite number of points (and therefore discrete) if ρ is rational, and they are dense on transverse circles if ρ is irrational. The cylinder is sliced along orbits at irrational values of r, which are intertwined with discrete orbits at rational values of r. The question is: in which sense is such a pattern stable under perturbations, that is here when $K \neq 0$?

Among the orbits which are dense in a curve wrapped around the cylinder, particularly interesting are the orbits which will persist under perturbation, in particular because they separate the space into domains which do not communicate. It is known that when K is large (for example $K > 4/3$, see [5]), such orbits do not exist, and on the other hand, when K is small, some of the irrational orbits persist, depending on arithmetical properties of the rotation number. For perturbed twist maps, we define the rotation number as $\lim_{n\to\infty} n^{-1}\theta_n$, where (θ_n, r_n) is the n-th iterated map obtained from (1).

Other kinds of invariant curves may occur, attached to elliptic periodic orbits. For instance if $K > 0$ is sufficiently small, the point $(1/2, 0)$ is an elliptic fixed point. Due to KAM Theorem (see [6] for a review), there exist homotopically trivial invariant curves on the cylinder winding around this fixed point, which form the so–called elliptic islands. We shall not consider here the problem of such orbits, although there existence is very important in connection with ergodic theory. Indeed one expects chaotic behaviour for K large, but the persistence of elliptic islands could prevent the map from being ergodic.

2 The Critical Constants

For the standard map, we consider now the homotopically non–trivial invariant curves *i.e.* wrapped around the cylinder. A natural way to look for their existence is to replace the angular variable θ by the new variable $\phi \in \mathbf{T}$

$$\theta = \phi + u(\phi, K, \rho) \quad , \quad r = \rho + u(\phi, K, \rho) - u(\phi - \rho, K, \rho) \; . \tag{6}$$

With the condition $1 + \partial u/\partial\phi > 0$, it would describe a curve around the cylinder, on which the map is expressed as $\phi' = \phi + \rho$, when ϕ describes $\mathbf{T}$ for K and ρ fixed. We say that (6) expresses on the curve the conjugacy of the map to a rotation. The existence of a function $u(\phi, K, \rho)$, analytic in the variable ϕ, insures the existence of an analytic invariant curve with rotation number ρ. We are interested to determine the critical constant $K_c(\rho)$ as being the largest possible value of K for which such an analytic function u exists. Of course, one could consider regularity constraints weaker than analyticity, leading to other critical constants. We look for a perturbation expansion of the function u, and we follow the notations of [7]. From the standard map we get from (6)

$$u(\phi + \rho, K, \rho) - 2u(\phi, K, \rho) + u(\phi - \rho, K, \rho) = \frac{K}{2\pi} \sin(2\pi(\phi + u(\phi, K, \rho)) \tag{7}$$

For $k \geq 1$, we have

$$\begin{aligned} & u^{(k)}(\phi + \rho, \rho) - 2u^{(k)}(\phi, \rho) + u^{(k)}(\phi - \rho, \rho) \\ & = \frac{1}{2\pi} \sin\big(2\pi\phi + 2\pi u(\phi, K, \rho)\big)\Big|_{k-1} \; , \end{aligned} \tag{8}$$

where in the right hand side, one keeps only the terms of order $k-1$ in the expansion on powers of K. We use now the Fourier series expansion on ϕ, that is $u^{(k)}(\phi, \rho) = \sum_{\nu \in \mathbf{Z}} u_\nu^{(k)}(\rho) e^{2i\pi\nu\phi}$, and we see that the coefficient of $e^{2i\pi\nu\phi}$ in the left hand side of (8) is $2(\cos(2\pi\nu\rho) - 1)u_\nu^{(k)}(\rho)$. Therefore (8) allows a recursive computation of the Fourier coefficients $u_\nu^{(k)}(\rho)$, and we get for $u^{(k)}(\phi, \rho)$ expressions as trigonometric polynomials in ϕ. However terms of the kind $2(\cos(2\pi\nu\rho) - 1)$ occur in the denominators along the steps of the recursion. Such factors are called "small divisors", some of them vanish when ρ is rational, and may become arbitrarily small when ν becomes large, for irrational ρ. Now let $\widetilde{K}_c(\rho)$ be the minimum over ϕ of the convergence radius of the expansion

$$u(\phi, K, \rho) = \sum_{k=1}^{\infty} K^k u^{(k)}(\phi, \rho) \; . \tag{9}$$

For ρ rational, (8) cannot be solved, and we set $\widetilde{K}_c(\rho) = 0$. For irrational values of ρ, Berretti and Gentile [8] were able to control $\widetilde{K}_c(\rho)$ using the

Brjuno function $B(\rho)$ which is a number theoretic function which will define in the following Section 4. More precisely, there exists $C > 0$ such that, for any irrational ρ,

$$\left|\ln\left((\widetilde{K}_c(\rho))^{-1}\right) - 2B(\rho)\right| < C\,. \tag{10}$$

The functions $\widetilde{K}_c(\rho)$ and $e^{-2B(\rho)}$ both vanish on all rationals, but the previous equation shows that the ratio $\widetilde{K}_c(\rho)/e^{-2B(\rho)}$ remains uniformly bounded at every irrationals. The fact that this ratio remains bounded is in itself amazing, but it recalls earlier and now classical results by Yoccoz [9] on the linearisation of holomorphic maps. We shall see later that we may have even better results in the framework of holormorphic maps.

The determination of the radius of convergence in (9) is not the whole story. It is possible that the function u may be analytically continued for real values of ϕ and for $K > \widetilde{K}_c(\rho)$ real. Thus we would have another critical constant $K_c(\rho)$ such that we still have real analytic curves for $\widetilde{K}_c(\rho) < K < K_c(\rho)$ real. The numerical results [10] seem to indicate that this is indeed the case: see [11] for a detailed discussion of this issue both from the numerical and the analytical points of view, which also uses results of [12,13]. The definitive answer is not known to us today, although we are led to expect that the function $B(\rho)$ plays a central role in the determination of $K_c(\rho)$ (see also Davie [14]).

3 Complex Analytic Maps

The problems of the critical constant is better understood in the case of the complex analytic maps. We have already seen in (10) that the critical constant $\widetilde{K}_c(\rho)$ of the complexified version of the standard map is controlled by the Brjuno function. A simpler example is the "Semi-standard Map", which is a two dimensional complex map on the cylinder, closely related to the standard map (1) : to get the semi-standard map, just replace in (1) the sine function $\sin(2\pi\theta)$ by its positive frequency part $(1/2i)\exp(2i\pi\theta)$. The procedure to get analytic invariant curves proceeds in a completely similar way as Equations (6) to (9), and it was proven that in this case [15,16], the critical constant $K_{\rm ssm}(\rho)$ defined in a same way as above, fulfils as in (10)

$$\left|\ln\left((K_{\rm ssm}(\rho))^{-1}\right) - 2B(\rho)\right| < C\,. \tag{11}$$

The numerical results (especially the figure 16) in ref. [16] provide more. Not only the ratio $K_{\rm ssm}(\rho))/e^{-2B(\rho)}$ is bounded on irrationals, but it is extendable to a continuous function on $[0,1]$, bounded below and above by positive constants. This result is amazing if one remember that both $K_{\rm ssm}(\rho))$ and $e^{-2B(\rho)}$ vanish at all rationals. Therefore the Brjuno function $B(\rho)$ is a good model to represent the singular behavior of $\ln\left((K_{\rm ssm}(\rho))^{-1}\right)$.

The Brjuno function was introduced by Yoccoz [9] in the apparently simpler problem of the linearisation of complex holomorphic maps around their fixed points. This is a more than one century old problem (see [17] for a nice review), which we can state as follows. Let $f(z)$ be a holomorphic map such that $f(0) = 0$, $f'(0) = e^{2i\pi\rho}$. Is it possible to conjugate the map f to its linear part? This means that we look for a function h, holomorphic in a disk of radius R_f, such that $h(0) = 0$, $h'(0) = 1$, and $f(h(z)) = h(ze^{2i\pi\rho})$. Note that such a function h, if it exists, is unique. In this case, the function f is said to admit a Siegel disk of radius R_f. The Siegel disk is a topological disk with conformal radius R_f, since it is the image through the normalised conformal map h of the disk $|z| < R_f$.

We quote now the classical results on this question [17]. i) If ρ is rational, there is no disk, that is $R_f = 0$. ii) If ρ is irrational and satisfies a (strong) Liouville condition, we still have $R_f = 0$. iii) If ρ is a diophantine irrational (see Section 5 below), then there exists a Siegel disk and $R_f > 0$, more precisely this happens when $B(\rho)$ is finite. iv) If $B(\rho) = +\infty$, then that there exist functions f such that $R_f = 0$. Indeed Yoccoz [9] proved the following: define $R(\rho)$ as the smallest radius of the Siegel disks R_f obtained when f varies in the compact family of all univalent maps on the unit disk such that $f(0) = 0$, and $f'(0) = e^{2i\pi\rho}$. Then we have

$$\left|\ln\left((R(\rho))^{-1}\right) - B(\rho)\right| < C\ . \tag{12}$$

Now consider the family of quadratic polynomial $P_2(z) = e^{2i\pi\rho}(z - z^2)$, and call $R_2(\rho)$ the radius of the Siegel disk associated to it. Observe first that, through the rescaling $z \to e^{-2i\pi\rho}R\times z$, then P_2 is transformed in $e^{2i\pi\rho}z - Rz^2$. In the rescaled variable, we see that R_2 is the maximum value of the constant R for which a circle with conformal radius one is persistant. Therefore $R_2(\rho)$ is the critical constant adapted to the present case, and this leads to the analogies between (10), or its equivalent in the real case, and (12).

Here again, the numerical results (now the figure 6) in ref. [16] bring some continuity properties. Not only the ratio $R_2(\rho))/e^{-B(\rho)}$ is bounded on irrationals, but it is extendable to a continuous function on $[0, 1]$, bounded below and above by positive constants. Therefore the Brjuno function $B(\rho)$ is again a good model to represent the singular behavior of $\ln\left((R_2(\rho))^{-1}\right)$. In our work [18] which started from these observations, we give arguments which strongly support the conjecture that the ratio $R_2(\rho))/e^{-B(\rho)}$ is not only continuous but satifies a Hölder continuity condition with exponent $1/2$. More precisely, the Brjuno function displays the universal singular behaviour (up to some Hölder-$\frac{1}{2}$ continuous function) of the critical functions occuring in small divisors holomorphic problems in dimension one.

The use of the Brjuno function was somewhat implicit in the work of Buric et al. [19], where they attempted to find representations of the critical constants by what they called modular smoothing. Singular functions of the same type occured in MacKay [20] in relation to the Brjuno condition.

It is nevertheless useful to recall here briefly one of the steps, called renormalisation, which plays a special role in Yoccoz's argument, and will appear to be crucial in understanding the fine regularity properties of ratios of the type $R_2(\rho))/e^{-B(\rho)}$. For this purpose, we follow [17], and we we consider first a rotation of angle $2\pi\rho$, with $0 \leq \rho < 1$, that is $z \to e^{2i\pi\rho}z$, acting in an open disk of radius R_ρ centered at the origin in the complex plane. We need an arbitrary point a, such that $|a| = R_\rho$, and for simplicity we take the real point $a = R_\rho$. Let $a' = e^{2i\pi\rho}R_\rho$ its image. Now, consider the angular sector bounded by the lines $0a$ and $0a'$, namely $\Delta_\rho = \{\ z \mid 0 \leq \arg\ z < 2\pi\rho\ ,\ |z| < R_\rho\}$, the line $0a'$ being excluded. Consider the orbit made of the successive iterated points $z_n,\ n \geq 1$ starting from $z_0 \in \Delta_\rho$, and let z_q the first of these points which also belong to Δ_ρ. The map $z_0 \to z_q$ is thus the first return map in the sector. We have $(2\pi)^{-1}(\arg\ z_q) = (2\pi)^{-1}(\arg\ z_0) + q\rho - 1$. We now take in the sector the variable u such that its complex conjugate $\overline{u} = z^{\frac{1}{\rho}}$, where now u belongs to a disk of radius $R_\rho^{\frac{1}{\rho}}$. For the values u_0 and u_q, corresponding to z_0 and z_q, we have $(2\pi)^{-1}(\arg\ u_q) = (2\pi)^{-1}(\arg\ u_0) - q + \frac{1}{\rho}\ (\mathrm{mod}\ 1)\ = (2\pi)^{-1}(\arg\ u_0) + \frac{1}{\rho}$. The original map which acted in a disk of radius R_ρ, leads in the new "renormalised variable" u, to a rotation with rotation number $\frac{1}{\rho}$, acting in a disk with radius $R_{(1/\rho)}$, such that $\ln R_\rho = \rho \ln R_{(1/\rho)}$.

This construction extends to the non linear perturbed case, for example $P_2(\rho, z) = e^{2i\pi\rho}(z - z^2)$, with a lot of complications. Suppose that there is a Siegel disk for $P_2(\rho, z)$. In this disk, there are conformal coordinates on which the maps is exactly a rotation of angle $2\pi\rho$, and on these coordinates we apply the linear renormalisation. The problem is then to give an interpretation of the renormalised coordinates u which we obtain. It appears that there exist a holomorphic map in the variable u with rotation number ρ^{-1}, which admits a Siegel disk, with conformal radius $\widetilde{R}_{(1/\rho)}$ such that $\ln R_\rho - \rho \ln \widetilde{R}_{(1/\rho)} = 0$. However, this map is not a polynomial with degree 2. This led Yoccoz to extend the problem to the compact family of univalent map on the unit disk with rotation number ρ, and he has considered the minimum $R(\rho)$ of the radius of the Siegel disk taken over this family of maps. The result is two modifications to the relation $\ln R_\rho - \rho \ln R_{(1/\rho)} = 0$ obtained in the linear case. First due to the minimisation procedure, the best one could get is a positive uniform upper bound for this expression instead of zero. Second, there is a special difficulty when ρ goes to zero. In this case the Siegel disk is strongly distorted, since there is an other fixed point which tends to zero when ρ goes to zero. The comparison between the linear and the non linear case becomes unjustified in this limit. Yoccoz proved that the result is an additional logarithmic term in the estimate, so that we only get that $\ln R_\rho - \rho \ln R_{(1/\rho)} - \ln \rho$ is bounded. It is therefore natural to compare the function $-\ln R(\rho)$ (as well as $-\ln R_2(\rho)$) to the solution of the equation $B(\rho) - \rho \ln B(1/\rho) + \ln \rho = 0$ which we will see, is nothing else than the Brjuno function.

4 Continued Fractions and the Brjuno Function

We first give a somewhat unusual definition of the continued fraction expansion sometimes called "Japanese continued fractions" [21]. Let α be a fixed real number such that $\frac{1}{2} \leq \alpha \leq 1$. Then, given the starting number x, the coefficients a_n and ε_n are recursively uniquely defined by the conditions

$$x = a_0 + \varepsilon_0 x_0, \text{ and } \forall n \geq 0,\ x_n^{-1} = a_{n+1} + \varepsilon_{n+1} x_{n+1} \ , \tag{13}$$

with $\forall n \geq 0,\ \alpha - 1 \leq \varepsilon_n x_n < \alpha$. We define the modified integer part $[x]_\alpha$ and the modified fractional part $\{x\}_\alpha$ as follows,

$$[x]_\alpha = [x - \alpha + 1]_1 \quad \text{and} \quad \{x\}_\alpha = \{x - \alpha + 1\}_1 + \alpha - 1 \ , \tag{14}$$

where $[x]_1$ and $\{x\}_1$ are the usual integer and fractional parts of x (so that $0 \leq \{x\}_1 < 1$). With these notations, we can rewrite (13) as

$$\begin{aligned} &a_0 = [x]_\alpha,\ \varepsilon_0 x_0 = \{x\}_\alpha, \text{ and,} \\ &a_{n+1} = [x_n^{-1}]_\alpha,\ \varepsilon_{n+1} x_{n+1} = \{x_n^{-1}\}_\alpha, \forall n > 0 \ . \end{aligned} \tag{15}$$

Therefore the x_n are generated by iterating the function $A_\alpha(x) = \left|\{x^{-1}\}_\alpha\right|$, that is $\forall n \geq 0,\quad x_{n+1} = A_\alpha(x_n) = \left|\{x_n^{-1}\}_\alpha\right| = \left|x_n^{-1} - [x_n^{-1}]_\alpha\right|$. A more detailed description states that the map A_α is made of the following branches

$$\text{branch } k^+ : \quad A_\alpha(x) = \frac{1}{x} - k \quad \text{for} \quad \frac{1}{k+\alpha} < x \leq \frac{1}{k} \ , \tag{16a}$$

$$\text{branch } k^- : \quad A_\alpha(x) = k - \frac{1}{x} \quad \text{for} \quad \frac{1}{k} < x \leq \frac{1}{k+\alpha-1} \ . \tag{16b}$$

When $\frac{1}{2} < \alpha \leq 1$, the function A_α maps the interval $[0, \alpha)$ to itself, whereas when $\alpha = \frac{1}{2}$, it maps the interval $[0, \alpha]$ to itself. In both cases, it is convenient to set $A_\alpha(0) = 0$, and we get a map which is infinitely differentiable by pieces, and the points where it is not differentiable accumulate to 0. Now x and the reduced fraction p_n/q_n admit the following representation

$$\begin{aligned} x &= a_0 + \cfrac{\varepsilon_0}{a_1 + \cfrac{\varepsilon_1}{a_2 + \ddots + \cfrac{\varepsilon_{n-1}}{a_n + \varepsilon_n x_n}}} \quad , \\ \frac{p_n}{q_n} &= a_0 + \cfrac{\varepsilon_0}{a_1 + \cfrac{\varepsilon_1}{a_2 + \ddots + \cfrac{\varepsilon_{n-1}}{a_n}}} \ . \end{aligned} \tag{17}$$

As long as the x_n's do not vanish, we have

$$x = \frac{p_n + p_{n-1}\varepsilon_n x_n}{q_n + q_{n-1}\varepsilon_n x_n} \quad , \quad x_n = (-\varepsilon_n)\frac{p_n - x q_n}{p_{n-1} - x q_{n-1}} \ , \tag{18}$$

and the recursion relations

$$p_n = a_n p_{n-1} + \varepsilon_{n-1} q_{n-2} \ , \ p_0 = a_0 \ , \ p_{-1} = 1 \ , \tag{19a}$$

$$q_n = a_n q_{n-1} + \varepsilon_{n-1} q_{n-2} \ , \ q_0 = 1 \ , \ q_{-1} = 0 \ , \tag{19b}$$

so that we get $0 < q_0 \leq q_1 < q_2 < \ldots < q_n < q_{n+1} < \ldots$. We also define

$$\beta_n = x_0 x_1 \cdots x_n = (-1)^n \varepsilon_0 \varepsilon_1 \cdots \varepsilon_n (q_n x - p_n) \ , \tag{20}$$

and we have

$$\frac{1}{1+\alpha} \leq \beta_n q_{n+1} = \frac{q_{n+1}}{q_{n+1} + \varepsilon_{n+1} q_n x_{n+1}} \leq \frac{1}{\alpha} \ . \tag{21}$$

Now there exist $\lambda(\alpha)$, with $0 < \lambda(\alpha) < 1$, and positive constants C_1 and C_2, such that[18]

$$\beta_n < C_1 \lambda(\alpha)^n \quad , \quad q_n > C_2 \lambda(\alpha)^{-n} \ . \tag{22}$$

Indeed we have

$$\text{for} \quad \frac{\sqrt{5}-1}{2} < \alpha \leq 1 \ , \quad \lambda(\alpha) = \lambda(1) = \frac{\sqrt{5}-1}{2} = 0.618... \ , \tag{23a}$$

$$\text{and for } \frac{1}{2} \leq \alpha \leq \frac{\sqrt{5}-1}{2} \ , \ \lambda(\alpha) = \lambda\left(\frac{1}{2}\right) = \sqrt{2} - 1 = 0.414... \ . \tag{23b}$$

When x_n=0 for some n, and $x_m \neq 0$ for $m < n$, then we have $x = p_n/q_n$ which is rational, and we say that the fraction stops at order n (with our conventions, we have $x_m = 0, \forall n \geq m$). Conversely, if x is rational, the continued fraction expansion stops at some finite order n. For $\alpha = 1$, we get the classical Gauss continued fraction expansion for which all signs $\varepsilon_n = +1$, and for $\alpha = 1/2$, we have the continued fraction to the nearest integer. Note that when $\alpha \neq 1$, the results of equations (23a) and (23b) are not obvious. For details, and in particular for the extension to others values of α, with $0 \leq \alpha < \frac{1}{2}$, see [18,22].

Given a positive real function f on $(0,1)$, the Brjuno series $B_f^{(\alpha)}(x)$ is the sum (which can be infinite) of the series with positive terms

$$\begin{aligned} B_f^{(\alpha)}(x) &= \sum_{n=0}^{\infty} \beta_{n-1} f(x_n) \\ &= f(x_0) + x_0 f(x_1) + \ldots + x_0 x_1 \cdots x_{n-1} f(x_n) + \ldots \ , \end{aligned} \tag{24}$$

where $\frac{1}{2} \leq \alpha \leq 1$, and for $k \geq 0$, x_k defined in Eq. (13) or (15). As mentioned above, when x is rational, we have $x_n = 0$ for some n, and we use $x_m = 0$ for $m \geq n$, The following results are easily obtained from the definitions

$$B_f^{(\alpha)}(x) = B_f^{(\alpha)}(x+1) \, , \tag{25a}$$

$$B_f^{(\alpha)}(x) = x B_f^{(\alpha)}\left(\frac{1}{x}\right) + f(x) \ \text{ for } 0 < x < \alpha \, , \tag{25b}$$

$$B_f^{(\alpha)}(x) = B_f^{(\alpha)}(-x) \ \text{ for } 0 < x \leq 1 - \alpha \, . \tag{25c}$$

In particular, $B_f^{(\frac{1}{2})}(x)$ is an even function. More surprising is the following result [18] : in the $\alpha = 1$ case, for $B_f^{(\pm)}(x) = \frac{1}{2}(B_f^{(1)}(x) \pm B_f^{(1)}(-x))$ which are the even and odd parts of $B_f^{(1)}(x)$, we have for $0 < x \leq \frac{1}{2}$,

$$B_f^{(-)}(x) = \frac{1}{2}\left(f(x) - f(1-x) - (1-x)f\left(\frac{x}{1-x}\right)\right), \tag{26a}$$

$$B_f^{(+)}(x) = xB_f^{(+)}\left(\frac{1}{x}\right) + \frac{1}{2}G(x), \quad \text{with} \tag{26b}$$

$$G(x) = f(x) + f(1-x) + (1-x)f\left(\frac{x}{1-x}\right) + 2xB_f^{(-)}\left(\frac{1}{x}\right). \tag{26c}$$

In order to prove the previous equations, we use Equations (23a–b) and the succession of transformations

$$-x \to 1-x \to \frac{1}{1-x} \to \frac{1}{1-x} - 1 = \frac{x}{1-x} \to \frac{1-x}{x} = \frac{1}{x} - 1 \to \frac{1}{x} \to x\,,$$

which provides the requested relations between $B(x)$ and $B(-x)$.

Now, it is convenient to introduce the following specific notations:

i) In $B_f^{(\alpha)}$(x), when $\alpha = 1$, we omit the superscript (α), and when $\alpha = \frac{1}{2}$, we replace the superscript (α) by e, so that $B_f(x) = B_f^{(1)}(x)$ and $B_f^e(x) = B_f^{(1/2)}(x)$ respectively.

ii) We omit the subscript f when $f(x) = -\ln(x) = \ln(x^{-1})$, so that $B^{(\alpha)}(x) = B_{-\ln}^{(\alpha)}(x)$, $B(x) = B_{-\ln}^{(1)}(x)$ and $B^e(x) = B_{-\ln}^{(1/2)}(x)$ respectively. We will call $B(x)$ *the Brjuno function*, which has been mentioned above in Sects. 2 and 3. We have

$$\begin{aligned} B(x) &= \sum_{n=0}^{\infty} \beta_{n-1} \ln\left(x_n^{-1}\right) \\ &= -\ln(x_0) - x_0 \ln(x_1) + \ldots - x_0 x_1 \cdots x_{n-1} \ln(x_n) - \ldots\,, \end{aligned} \tag{27}$$

where the x_n are obtained from (13) using $\alpha = 1$ (Gaussian case), whereas $B^e(x)$ is given by the same equation (27) with x_n obtained from (13) using $\alpha = \frac{1}{2}$ (continued fraction to the nearest neighbour). Both functions $B(x)$ and $B^e(x)$ are 1-periodic, and take value $+\infty$ for x rational. From (25a), the odd part of $B(x)$ is given for $0 \leq x \leq \frac{1}{2}$ by $B_-(x) = \frac{1}{2}x(\ln(x^{-1} - 1))$, which is continuous (and even Hölder continuous for any exponent $\sigma < 1$). Moreover, $B^e(x)$ is even, and it has been proven [18] that the difference $B^e(x) - B^+(x)$ is not only bounded, but continuous, and even Hölder continuous for exponent $\frac{1}{2}$. This refines a more general statement [18,22] which says that the differences $B_{\ln}^{(\alpha)} - B(x)$ are bounded over the irrationals.

The numerical computation of $B(x)$ and $B^e(x)$ is delicate, due to the instabilities of the continued fraction expansion. However, it is very easy to compute their values when the continued fraction expansion is periodic, that

is when x is an irrational quadratic number. This applies to noble numbers, in which case the x_n are constant after a certain order.

5 The Brjuno Series and Diophantine conditions

A real number is said to be a *Brjuno number* if and only if $B(x)$ is finite, and we also say that x satisfies *the Brjuno diophantine condition*. Brjuno numbers are irrationals and real numbers satisfying the classical diophantine conditions (which we recall below) are Brjuno numbers. In [18], we show that for $\frac{1}{2} \leq \alpha \leq 1$, $B^{(\alpha)}(x)$ is finite if and only if x is a Brjuno number. More precisely, the proof says that for any $\alpha \in [0, \frac{1}{2}]$, the difference $|B^{(\alpha)}(x) - B(x)|$ is bounded over irrational values of x. We also show that for $\alpha = 1$, the difference $|B(x) - \sum_{n=0}^{\infty} q_n^{-1} \ln(q_{n+1})|$ is bounded over irrational values of x, so that we recover the original definition of the Brjuno numbers [23]: x is a Brjuno number if and only if $\sum_{n=0}^{\infty} q_n^{-1} \ln(q_{n+1})$ is bounded over the irrational. One can see [18,22] that such a definition of the Brjuno numbers does not depend of the particular value of α used to compute the q_n.

We now report the usual definition [6] of the diophantine conditions : we say that x is an irrational diophantine number of order $\tau \geq 0$ (and we write $x \in \mathrm{C}(\tau)$), if there exists $c > 0$ such that for any integers p and q, such that $q > 0$, we have $|x - p/q| \geq cq^{-2-\tau}$. Some classical facts need to be recalled here [24]. First, for any p and q such that $0 < q < q_{n+1}$, we have $|qx - p| \geq |q_n x - p_n|$, where p_n/q_n is the Gaussian reduced fraction to x. Therefore, in order to have $x \in \mathrm{C}(\tau)$, it is sufficient to check that for any $n > 0$, $|x - p_n/q_n| \geq cq_n^{-2-\tau}$. Second, Liouville's classical theorem asserts that algebraic numbers of degree n belong to $x \in \mathrm{C}(n-2)$. Moreover Roth's theorem shows that all algebraic numbers belong to $\mathrm{C}(\tau)$, for all $\tau > 0$. Finally, for an arbitrary irrational, and any $n > 0$, we have $(q_n q_{n+1})^{-1} \leq |x - p_n/q_n| \leq q_n^{-2}$. Using (21) for $\alpha = 1$, we get an equivalent characterisation of the diophantine conditions: $x \in \mathrm{C}(\tau)$ if and only if there exists a constant $c > 0$ such that $\beta_n \geq c\,\beta_{n-1}^{1+\tau}$ for any $n > 0$.

Now we introduce for $\nu > 0$, the Brjuno series $B_{\{\nu\}}(x) \equiv B_{x^{-\nu}}(x)$ for the fonctions $f(x) = x^{-\nu}$ (still using $\alpha = 1$),

$$B_{\{\nu\}}(x) = \sum_{n=0}^{\infty} \beta_{n-1}(x) x_n^{-\nu} = \sum_{n=0}^{\infty} \beta_{n-1} \left(\frac{\beta_n}{\beta_{n-1}} \right)^{-\nu} = \sum_{n=0}^{\infty} \beta_{n-1}^{1+\nu}(x) \beta_n^{-\nu}(x) \ . \tag{28}$$

Using (21) one gets

$$2^{-\nu} \sum_{n=0}^{\infty} q_n^{-1-\nu} |q_n x - p_n|^{-\nu} \leq B_{\{\nu\}}(x) \leq \sum_{n=0}^{\infty} q_n^{-1-\nu} |q_n x - p_n|^{-\nu} \ . \tag{29}$$

The series $B_{\{\nu\}}(x)$ converges if and only if the series $\sum_{n=0}^{\infty} q_n^{-1-\nu} |q_n x - p_n|^{-\nu}$ converges, that is if the series $\sum_{n=0}^{\infty} q_n^{-1-2\nu} |x - (p_n/q_n)|^{-\nu}$ also converges.

As a consequence, if $B_{\{\nu\}}(x) < \infty$, then $q_n^{-1-2\nu} \, |x - (p_n/q_n)|^{-\nu}$ is bounded, and $x \in \mathrm{C}(1/\nu)$. Conversely, assume $\tau \geq 0$, and $x \in \mathrm{C}(\tau)$, then we have

$$B_{\{\nu\}}(x) \leq c^{-\nu} \sum_{n=0}^{\infty} q_n^{-1+\tau\nu} \; . \tag{30}$$

Using bounds in (22), we get the following statement: If $x \in \mathrm{C}(\tau)$, then for any ν such that $\tau < \nu^{-1}$, $B_{\{\nu\}}(x) < \infty$. Therefore, there is a relation between the diophantine conditions $\mathrm{C}(\tau)$, and the convergence of the Brjuno series for $f(x) = x^{-\nu}$: the set of irrationals x such that $B_{\{\nu\}}(x) \equiv B_{x^{-\nu}}(x)$ is bounded, is contained in $\mathrm{C}(1/\nu)$, and contains $\mathrm{C}(-\varepsilon + 1/\nu)$, for any $0 < \varepsilon \leq 1/\nu$. In some sense, the Brjuno conditions is related to the limiting case $\nu = 0$, and in particular, $x \in \mathrm{C}(\tau)$ for $\tau > 0$ implies $B(x) < \infty$, that is x is a Brjuno number. Using a more general function f, positive on $(0, 1)$, and monotoneously decreasing in the vicinity of zero, we can introduce a wide family of conditions $B_f(x) < +\infty$. The diophantine conditions obtained will be mainly governed by the singular behavior of f around zero. A power law behaviour would simulate the usual conditions, whereas a logarithmic behaviour would generate a condition similar to the Brjuno condition. Other interesting examples would be obtained by taking functions f of the form $x^{-\nu} |\log(x)|^{\mu}$, $x^{-\nu} |\log(x)|^{\mu} |\log(|\log x|)|^{\sigma}$, and so one.

6 The Brjuno Operator

We will introduce now some functional analysis in order to solve Equations (25a–c). For fixed $\frac{1}{2} \leq \alpha \leq 1$, let us consider the operator $T_{(\alpha)}$, acting on locally Lebesgue integrable functions f on the real line, which verify

$$f(x) = f(x+1) \text{ for almost every } x \in \mathbb{R} \, , \tag{31a}$$

$$f(x) = f(-x) \;\; \text{for almost every } x \in (0, 1-\alpha) \, . \tag{31b}$$

The operator is defined by

$$(T_{(\alpha)} f)(x) = x f\left(\frac{1}{x}\right) \; , \;\; \text{if} \;\; x \in (0, \alpha) \; . \tag{32}$$

It is understood that the function $T_{(\alpha)} f$ is completed outside $(0, \alpha)$ by imposing on $T_{(\alpha)} f$ the same parity and periodicity conditions which are expressed for f in the above equations (31a–b). The functional equations (25a–c) can then be written in the form

$$\left(1 - T_{(\alpha)}\right) B_f^{(\alpha)} = f \; . \tag{33}$$

This suggests studying the operator $T_{(\alpha)}$ on the Banach spaces

$$X_{\alpha,p} = \{ f : \mathbb{R} \to \mathbb{R} \mid f \text{ verifies (31a–b)} \; , \;\; f \in L^p(0, \alpha) \} \tag{34}$$

endowed with the norm of $L^p(0,\alpha)$, namely

$$||f||_{\alpha,p} = \left(\int_0^\alpha |f(x)|^p \, dx \right)^{1/p} , \tag{35}$$

for $p \in [1,\infty]$. Note that one could also use $L^p(0,1)$, instead of $L^p(0,\alpha)$, and that if $p < p'$ one has the obvious inclusion $X_{\alpha,p'} \subset X_{\alpha,p}$. If $(1 - T_{(\alpha)})$ is invertible in the considered space, then (25a–c) have a unique solution for $B_f^{(\alpha)}$, provided that the f in the right hand side of (25b) also belongs to the space. The invertibility property is given by the following theorem, which states in particular that the spectral radius of $T_{(\alpha)}$ is strictly smaller than 1.

Theorem. $T_{(\alpha)}$ is a linear bounded operator from $X_{\alpha,p}$ into itself for all $\alpha \in [\frac{1}{2}, 1]$ and for all $p \in [1,\infty]$. Its spectral radius on $X_{\alpha,p}$ is bounded by the constant $\lambda(\alpha)$ of Equation (22), and therefore $1 - T_{(\alpha)}$ is invertible.

For the proof, see [18]. We will just observe here that the result is immediate in the $p = \infty$ case. Indeed,

$$(T_{(\alpha)}^n f)(x) = \beta_{n-1}(x) f(x_n) = \beta_{n-1}(x)(f \circ A_\alpha^n)(x) , \tag{36}$$

where the map A_α is defined above (see (16a–b)). Therefore

$$||T_{(\alpha)}^n f||_{\alpha,\infty} \leq \sup_x(\beta_{n-1}(x)) ||f||_{\alpha,p} \leq c\lambda(\alpha)^{n-1} ||f||_{\alpha,p} \tag{37}$$

and one gets the theorem (for p infinite) by taking the $1/n$–th root of both sides. For the other values of p, it is convenient to make use of the measure which is invariant under transformation by the map A_α, instead of the Lebesgue measure. An immediate consequence of the theorem, is that if we take $f(x) = \ln(x)$, for $0 < x < \alpha$, then $f \in X_{\alpha,p}$, for all *finite* p and therefore we also have $B^{(\alpha)} \in X_{\alpha,p}$ for all finite p.

However, we have a stronger property in the $\alpha = \frac{1}{2}$ case. Here, we set again $T_e \equiv T_{(1/2)}$. In this case, the logarithm belongs also to the set $X_* \subset X_{\alpha=1/2,p=1}$, made of even, periodic (with period 1) functions belonging to the so-called "BMO-space". In this space, f has bounded mean oscillation, more precisely the following semi-norm $||f||_*$ is bounded, with

$$||f||_* = \sup_I \frac{1}{|I|} \int_I |f - f_I| dx , \tag{38}$$

where the mean value of f over I is $f_I = |I|^{-1} \int_I |f - f_I| dx$, and the sup is taken over all possible intervals I with length $|I|$ smaller than one. The BMO space has remarkable properties. First it is contained in all L_p spaces for p finite, and it contains the L_∞ space, second it is the space adequate to describe functions having singular behaviour not worse than logarithmic, but around every point in a dense set of the real line, and third, it has remarkable properties connected to the harmonic conjugacy transformation [25,18].

In [18,26], we have shown that $1 - T_e \equiv 1 - T_{(1/2)}$ is invertible in X_*, and therefore that $B^e \equiv B^{(1/2)} \in X_*$. Since $|B^\alpha - B^e|$ is bounded for $\frac{1}{2} \leq \alpha \leq 1$,

we have the unexpected consequence that all $B^{(\alpha)}$ for $\frac{1}{2} \leq \alpha \leq 1$, also have the Bounded Mean Oscillation property, although it cannot be shown directly through the properties of $1 - T_{(\alpha)}$.

The BMO property obtained for the Brjuno function in the real case, was one of our motivations to consider the complexification procedure which we will describe in the last Section of this paper [27].

7 Application to Hölder–continuous Functions

In this Section, we will consider only the case $\alpha = \frac{1}{2}$. In this case the map $A_\alpha \equiv A_{1/2}$ is continuous on the interval $(0, \frac{1}{2}]$. The functional equation for the Brjuno function B^e for $\alpha = 1/2$ is

$$[(1 - T_e)B^e](x) = -\log x \ , \tag{39}$$

for all $x \in (0, 1/2)$, complemented with the condition that B^e is even and periodic. We suppose that the right hand side of this equation is pertubed, by an additional term f, which is less singular than the logarithmic function, and we want to study the singular properties of the perturbed solution. Since the equation is linear, we only need to consider the action on f of T_e and $(1-T_e)^{-1}$, which we will conveniently rename the Brjuno operator $\mathbf{B_e}$. We will consider even and periodic functions f which are *continuous*. It is sufficient to know the value of f on $[0, 1/2]$, so we assume $f \in C^0_{[0,1/2]}$. One can check that Tf is also continuous provided we set $Tf(0) = 0$. We need now the usual Hölder's type semi-norms for continuous functions : let $f \in C^0_{[0,1/2]}$, then we define the Hölder's γ-norm as

$$|f|_\gamma = \sup_{0 \leq x < y \leq 1/2} \frac{|f(x) - f(y)|}{|x - y|^\gamma} \ , \tag{40}$$

with $0 < \gamma \leq 1$. This is a seminorm since it vanishes on constant functions, so that we introduce the norm:

$$||f||_\gamma = A|f|_\gamma + B|f|_\infty \ , \tag{41}$$

where $|f|_\infty$ is the L_∞ norm of f, and A and B are positive constants which we can choose arbitrarily, provided that they do not vanish. We say that $f \in C^\gamma$, if $f \in C^0_{[0,1/2]}$ and $|f|_\gamma$ is finite. We now have:

Proposition. T_e is a bounded operator in C^γ, for the norm $||f||_\gamma$, when $0 < \gamma \leq 1/2$, provided B/A is large enough: if $B/A > (2^\gamma - 2^{-\gamma})^{-1}$, the norm of T_e corresponding to the norm (41) satisfies $||T_e||_\gamma \leq 2^{(2\gamma-1)} \leq 1$. Therefore for $0 < \gamma < 1/2$, T_e is a contraction, and $1 - T_e$ is invertible.

We need the following Lemma

Lemma. Let $0 < y < x \leq 1/2$, and define x_1 and y_1 by the following conditions

$$y = \frac{1}{n + y_1} \quad , \quad x = \frac{1}{m + x_1} \ , \tag{42}$$

with $n \geq 2$ and $m \geq 2$, and $-1/2 \leq x_1 < 1/2$ and $-1/2 \leq y_1 < 1/2$, then we have

$$||x_1| - |y_1|| \leq \frac{|x-y|}{|x||y|} \ . \tag{43}$$

Proof of the Lemma. Since $y < x$, we have $n - m > x_1 - y_1 > -1$. therefore $n \geq m$. Let $n - m = p \geq 0$. We have $x - y = xy(p + y_1 - x_1)$. So that we need to prove $||x_1| - |y_1|| \leq |p + y_1 - x_1|$. This is obvious when $p = 0$. We always have $||x_1| - |y_1|| \leq 1/2$, so the required inequality also holds when $p \geq 2$. In the remaining case $p = 1$, we set $\eta = \text{sign}(y_1)$ and $\epsilon = \text{sign}(x_1)$, and we need to check that $||x_1| - |y_1|| \leq |1 + \eta|y_1| - \epsilon|x_1||$. Still because the left hand side is smaller or equal to $1/2$, this last inequality is not obvious only when $\eta = -1$ and $\epsilon = +1$. It therefore remains to show that $||x_1| - |y_1|| \leq |1 - |y_1| - |x_1||$. Setting $u = 1/2 - |x_1|$ and $v = 1/2 - |y_1|$, the last inequality is equivalent to $|1 - v/u| \leq |1 + v/u|$, which is readily checked since u/v is real and non-negative.

Proof of the Proposition. Let $0 < y < x \leq 1/2$, and x_1 and y_1 as in the preceding lemma. We have

$$\begin{aligned}
|T_e f(x) - T_e f(y)| &= |xf(1/x) - yf(1/y)| = |xf(|x_1|) - yf(|y_1|)| && \text{(44a)}\\
&\leq |x-y||f(|x_1|)| + |y||f(|x_1|) - f(|y_1|)| && \text{(44b)}\\
&\leq |x-y||f|_\infty + |y||f|_\gamma ||x_1| - |y_1||^\gamma && \text{(44c)}\\
&\leq |x-y||f|_\infty + |f|_\gamma \frac{|x-y|^\gamma}{|x|^\gamma |y|^{\gamma-1}} \,, && \text{(44d)}
\end{aligned}$$

where we have used $f \in C^\gamma$, and the Lemma. Therefore

$$\begin{aligned}
\frac{|T_e f(x) - T_e f(y)|}{|x-y|^\gamma} &\leq |x-y|^{1-\gamma}|f|_\infty + \left(\frac{|y|}{|x|}\right)^\gamma |y|^{1-2\gamma}|f|_\gamma && \text{(45a)}\\
&\leq (1/2)^{1-\gamma}|f|_\infty + (1/2)^{1-2\gamma}|f|_\gamma \,, && \text{(45b)}
\end{aligned}$$

since $0 < y < x \leq 1/2$, and $\gamma \leq 1/2$. For $y = 0$, the right hand side can be replaced by its first term $(1/2)^{1-\gamma}|f|_\infty$, and the above inequality extends to the case where y vanishes, so that

$$|T_e f|_\gamma \leq K_\gamma(f) = 2^{\gamma-1}|f|_\infty + 2^{2\gamma-1}|f|_\gamma \ . \tag{46}$$

For the norm, we get

$$\begin{aligned}
||T_e f||_\gamma &= A|T_e f|_\gamma + B|T_e f|_\infty && \text{(47a)}\\
&\leq 2^{2\gamma-1}\left[A|f|_\gamma + (2^{-2\gamma}B + 2^{-\gamma}A)|f|_\infty\right] \,, && \text{(47b)}\\
&\leq 2^{(2\gamma-1)}||f||_\gamma \,, && \text{(47c)}
\end{aligned}$$

provided $2^{-2\gamma}B + 2^{-\gamma}A \leq B$, that is $A/B \leq 2^\gamma - 2^{-\gamma}$ which completes the proof. The above proposition has two obvious consequences

- Since $C^\gamma \subset C^{\gamma'}$ whenever $\gamma' \leq \gamma$, we have

$$f \in C^\gamma \quad \text{and} \quad \gamma \geq 1/2 \qquad \Longrightarrow \quad T_e f \in C^{1/2} \tag{48a}$$

$$f \in C^\gamma \quad \text{and} \quad \gamma \geq \gamma_0 \quad , \quad \gamma_0 \leq 1/2 \qquad \Longrightarrow \quad T_e f \in C^{\gamma_0} \tag{48b}$$

- When $\gamma < 1/2$, T_e is a contraction on C^γ. Therefore $1 - T_e$ is invertible and $\mathbf{B_e} = \sum_0^\infty (T_e)^n = (1 - T_e)^{-1}$ preserves C^γ, and we have

$$f \in C^\gamma \quad \text{and} \quad 0 < \gamma < 1/2 \qquad \Longrightarrow \quad \mathbf{B_e} f \in C^\gamma \tag{49a}$$

$$f \in C^{1/2} \qquad \Longrightarrow \quad \mathbf{B_e} f \in C^\gamma, \forall \gamma \text{ such that } \quad 0 < \gamma < 1/2 \; . \tag{49b}$$

We have reproduced here the proof of [26], because it is essentially elementary. In fact we have a slightly better result for $\mathbf{B}_e$ than for T_e, as shown in the next proposition, which shows that the $C^{1/2}$ property is effectively reached. Its proof [18,26] is too difficult to be reproduced here. We have

Proposition. If $f \in C^\gamma$, and $\gamma > 1/2$, then $\mathbf{B}_e f \in C^{1/2}$.

The Brjuno function B^e which we have studied in the previous section is nothing else than $\mathbf{B}_e \ell$, where ℓ is equal to minus the logarithmic fonction restricted to $[0, 1/2]$. When made even and periodic, this function is not continuous. Suppose that we perturb ℓ by a function with enough regularity properties (for example C^1 or $C^{1/2+\epsilon}$), the change in $\mathbf{B}_e \ell$ will be continuous and even in $C^{1/2}$, that is Hölder-$\frac{1}{2}$ continuous. In this sense, the 'most singular part' of the Brjuno function is stable or 'universal', roughly speaking modulo Hölder-$\frac{1}{2}$ continuous contributions. As noticed at the end of Section 5 above, we can use either $B^e \equiv B^{(1/2)}$ or $B \equiv B^{(1)}$ since a similar argument starting from (26a–c) shows that their difference is also Hölder-$\frac{1}{2}$ continuous [18].

This provides a frame to understand the properties of the critical constants K of the Sections 2 and 3 above. We assume that the singularity comes from the renormalisation equation $(1 - T^e)K = \ell + f$, and not from additional singular behaviour coming from f in the right-hand side. If it would exist, such an additional singular behaviour would require a further physical interpretation. This argument, which is usual in the renormalisation analysis of singularities, was one of the motivation for our previous work [18]. The renormalisation equation allows naturally to conjecture that the difference between the Brjuno function B^e, and the logarithm of the various critical constants (multiplied by a suitable coefficient), is continuous and even Hölder-$\frac{1}{2}$ continuous. As an example, we conjecture that the ratio of e^{-B} and the radius of the Siegel disk of the quadratic polynomial, is an Hölder-$\frac{1}{2}$ continuous function of the rotation number. These conjectures are in agreement with the numerical results [16,10].

8 The Complexification of the Brjuno Function

We consider here the case $\alpha = 1$, and we want to associate to the function f in $X_{1,2}$, a function Φ, holomorphic in the upper half plane, such that

$\mathrm{Im}\,\Phi \to B_f$ when z goes to the real axis. Since $\mathrm{Re}\,\Phi$ is associated to the harmonic conjugate of f, we expect to find better boundedness properties when f has the BMO property. We will here describe our procedure, and report the results [27].

We associate to f a function $F(z)$ holomorphic in $\mathbb{C}\backslash[0,1]$, and vanishing at infinity, as follows

$$F(z) = \frac{i}{\pi}\int_o^1 \frac{f(x)}{x-z}\,dx\ . \tag{50}$$

For x real, we have $\mathrm{Im}\,F(x \pm i\varepsilon) = \pm f(x)$ for $x \in [0,1]$, and $\mathrm{Im}\,F(x) = 0$ for $x \notin [0,1]$. We will be particularly interested in the case $f(z) = \ln(z)$, in which case we get $F(z) = -\pi^{-1}\mathrm{Li}_2(1/z)$, where Li_2 is the classical dilogaritm function [28]. Now we set

$$\Phi(z) = \lim_{N\to\infty} \sum_{-N}^{+N} F(z+n)\ , \tag{51}$$

and we get a function Φ holomorphic in the upper-half plane $\mathbb{H}_+$, periodic with a real period equal to one, and such that for x real, $\mathrm{Im}\,F(x \pm i\varepsilon) = \pm f(x)$. In fact the previous equation defines a pair on functions $\Phi_\pm$, respectively holomorphic in the upper or lower half plane $\mathbb{H}_\pm$, so that the natural frame in which our procedure takes place is the frame of complex hyperfunctions, which we will not consider here [27].

We consider now the action of T, with $(Tf)(x) = xf(1/x)$ if $0 \le x < 1$, f and Tf being complemented using periodicity. Using the above correspondence $f \mapsto Tf$, a correspondence $F \mapsto TF$ is induced on holomorphic functions in $\mathbb{C}\backslash[0,1]$, vanishing at infinity. We get

$$(TF)(z) = -z\sum_{m=1}^{\infty}\left(F\left(\frac{1}{z}-m\right) - F(-m)\right) + \sum_{m=1}^{\infty} F'(-m)\ . \tag{52}$$

In fact, TF is essentially $-z\sum_{m=1}^{\infty} F\left(z^{-1}-m\right)$, up to an affine additive correction, which could be determined by the vanishing condition at infinity.

If f is associated to F as above, the solution $B_f(x)$ (for $\alpha = 1$) of (25a–c) is associated to the series

$$\mathcal{B}_f(z) = \sum_{\mathbb{Z}}\sum_{m=0}^{\infty}(T^n F(z))\ , \tag{53}$$

where we use the notation $\sum_{\mathbb{Z}} F$ for $\Sigma_{n=-\infty}^{+\infty} F(z+n)$ understood as the symmetric summation (51) to insure convergence.

It is now interesting to display the link between (53) and the modular group $GL(2,\mathbb{Z})$. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2,\mathbb{Z})$, which mean $a,b,c,d \in \mathbb{Z}$,

$\varepsilon_g = ad - bc = \pm 1$. To g we associate the following group action on functions holomorphic on $\mathbb{C}\backslash[0,1]$, that is $g \mapsto L_g F$, with

$$\big(L_g F\big)(z) = (a - cz)\left\{F\left(\frac{dz-b}{a-cz}\right) - F\left(-\frac{d}{c}\right)\right\} - \frac{\varepsilon_g}{c} F'\left(-\frac{d}{c}\right) . \quad (54)$$

Let $\mathcal{M}_+ \subset GL(2,\mathbb{Z})$ be the multiplicative monoid generated by the unit matrix, and the set of matrices $\begin{pmatrix} 0 & 1 \\ 1 & m \end{pmatrix}$, for $m \geq 1$ integer. The monoid $\mathcal{M}_+ \subset GL(2,\mathbb{Z})$ can also be defined as the set of matrices including identity and the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(2,\mathbb{Z})$ such that first, $d \geq c \geq a \geq 0$, and second, $d \geq b \geq a \geq 0$. In $GL(2,\mathbb{Z})$ there is a unique product decomposition, namely $\forall g \in GL(2,\mathbb{Z})$ there exist a unique set of three matrices k, m and h, with $g = kmh$, and $m \in \mathcal{M}^+$, $k \in Z$, where Z is the translation subgroup of matrices $\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$, $n \in \mathbb{Z}$, and $h \in H$, where h is the order eight sugroup of $GL(2,\mathbb{Z})$ made of the matrices $\begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon' \end{pmatrix}$, and $\begin{pmatrix} 0 & \varepsilon \\ \varepsilon' & 0 \end{pmatrix}$, with $\varepsilon = \pm 1$ and $\varepsilon' = \pm 1$.

Now (53) is rewritten as

$$\mathcal{B}_f(z) = \sum_{k \in Z} \sum_{g \in \mathcal{M}^+} \big(L_{(hg)} F\big)(z) . \quad (55)$$

The double sum over g and k amounts to a sum over a part of the full modular group (here one over eight). The contribution over the seven other possible parts would be $-\mathcal{B}_f(z)$, $\pm\mathcal{B}_f(z^{-1})$, $\pm\mathcal{B}_f(-z)$, and $\mathcal{B}_f(-z^{-1})$.

We will now summarize the results:

i) The sums in (55) converge in the open upper half plane as long as f is in $L_1(0,1)$ which insures that F is holomorphic in $\mathbb{C}\backslash[0,1]$, and vanishes at infinity.

ii) When f is in $L_p(0,1)$, p finite, then $\mathcal{B}_f$ is in the Hardy $\mathbb{H}_p$ space.

iii) If f is such that F has bounded real part, then the same holds for $\mathcal{B}_f$.

iv) For $f(x) = \ln(x)$, and $F(z) = -\pi^{-1}\mathrm{Li}_2(1/z)$, we get the complexified Brjuno function, $\mathcal{B}$, holomorphic in the upper half plane, vanishing at $+i\infty$. When z goes to a real number in a non-tangential way, we have the following limits when $\varepsilon > 0$ goes to zero : the real part $\mathrm{Re}\,\mathcal{B}(x+i\varepsilon)$ has a bounded limit for any real x. This limit is continuous at all irrationals and has a decreasing jump of π/q at each rational p/q. When x is a Brjuno number, $\mathrm{Im}\,\mathcal{B}(x+i\varepsilon)$ goes to the Brjuno function $B(x)$.

The limit properties of the complex Brjuno function on the real axis are characteristic of functions f having a logarithmic singularity aroud zero. Although the boundedness of the real part reminds the BMO property of the real Brjuno function, it is in fact a stronger property. This is one more remarkable feature of this function. We are convinced that the interpretation

of the properties of the Brjuno function in terms of the modular group is promising. On the other hand, we can hope to find an interpretation of the complexified version of rotation numbers analogous with the usual interpretation of complex frequencies in terms of damped oscillations, but this is another story.

Acknowledgements. The first author thanks the CNRS and the organizors of the École Thematique held at La Chapelle des Bois, with special thanks to Michel Planat. This work begun during a visit of the second author at the S.Ph.T.–CEA Saclay and at the Dept. of Mathematics of Orsay during the academic year 1993–1994. This research has been supported by the CNR, CNRS, INFN, MURST and a EEC grant. We thank J.-C. Yoccoz for his constant help and warm encouragement, and also for allowing us to report here results obtained in collaboration with him.

References

1. Chirikov, B. (1979) A universal instability of many-dimensional oscillator systems. Phys. Reports, **52**, 263–279
2. Escande, D. (1985) Stochasticity in classical Hamiltonian systems: universal aspects. Phys. Reports, **121**, 165–261
3. Greene, J. M. (1979) A method for determining a stochastic transition. J. Math. Phys. **20**, 1183–1201
4. Aubry S. and Le Daeron, P. (1983) The discrete Frenkel–Kontorova model and its extensions. Physica **8D**, 381–422
5. Mather J. N. (1984) Non existence of invariant circles. Ergod. Theor. and Dynam. Sys. **4**, 301–309
6. Yoccoz J.-C. (1992) An introduction to small divisors problems, in: From Number Theory to Physics, Waldschmidt M., Moussa P., Luck J.-M., and Itzykson C. editors, Springer-Verlag, Berlin, pp. 659–679
7. Berretti A. and Gentile G. (1998) Scaling properties of the radius of convergence of the Lindstedt series : the standard map. University of Roma, Italy, preprint
8. Berretti A. and Gentile G. (1998) Bryuno function and the standard map. University of Roma, Italy, preprint
9. Yoccoz J.-C. (1995) Théorème de Siegel, nombres de Bruno et polynômes quadratiques. Astérisque, **231**, 3–88, (appeared first as a preprint in 1987).
10. Marmi S. and Stark J. (1992) On the standard map critical function. Nonlinearity **5**, 743–761
11. Carletti T. and Laskar J. (1999) Scaling law in the standard map critical function, interpolating hamiltonian and frequency analysis. (Preprint, Bureau des Longitudes, Paris, in preparation)
12. Treshev D. and Zubelevitch O. (1998) Invariant tori in Hamiltonian systems with two degrees of freedom in a neighborhood of a resonance. Regular and Chaotic dynamics, **3**, 73–81
13. Gelfreich G. V. (1999) A proof of exponentially small transversality of the separatrices for the standard map. Commun. Math. Phys. **201**, 155–216

14. Davie A. M. (1995) Renormalisation for analytic area preserving maps. University of Edinburgh preprint
15. Davie A. M. (1994) the critical function for the semistandard map. Nonlinearity **7**, 219–229
16. Marmi S. (1990) Critical functions for complex analytic maps function. J. Phys. A: Math.Gen. **23**, 3447–3474
17. Perez-Marco R. (1992) Solution complète du problème de Siegel de linéarisation d'une application holomorphe autour d'un point fixe. Séminaire Bourbaki nr.753, Astérisque, **206**, 273–310
18. Marmi S., Moussa P., and Yoccoz J.-C. (1997) The Brjuno function and their regularity properties. Commun. Math. Phys. **186**, 265-293
19. Buric N., Percival I. C., and Vivaldi F. (1990) Critical function and modular smoothing, Nonlinearity **3**, 21–37
20. MacKay R. S. (1988) Exact results for an approximate renormalisation scheme and some predictions for the breakup of invariant tori, Physica **33D**, 240–265, and Erratum (1989) Physica **36D**, 358–265
21. Schweiger F. (1995) Ergodic theory of fibered systems and metric number thory, Clarendon Press, Oxford,
22. Moussa P., Cassa A., and Marmi S. (1999) Continued fractions and Brjuno functions, J. Comput. Appl. Math. **105** 403–415
23. Brjuno. A. D. (1971) Analytical form of differential equations, Trans. Moscow Math. Soc. **25** 131–288, and, (1972), **26** 199–239
24. Hardy G. H., and Wright E. M. (1938) An introduction to the theory of numbers, Clarendon Press, Oxford, chapter 11, fifth edition 1979
25. Garnett J. B. (1981) Bounded Analytic functions, Academic Press, New York.
26. Marmi S., Moussa P., and Yoccoz J.-C. (1995) Développements en fractions continues, fonctions de Brjuno et espaces BMO, CEA/Saclay, Note CEA-N-2788
27. Marmi S., Moussa P., and Yoccoz J.-C. (1999) Complex Brjuno functions, Preprint SPhT/CEA Saclay T99/066, 71 p.
28. Lewin L, (1981) Polylogarithms and Associated Functions, Elsevier North Holland, New York.

Algebraic and Analytic Randomness

Jean-Paul Allouche *

CNRS, LRI, Bâtiment 490
F-91405 Orsay Cedex, France
allouche@lri.fr
http://www.lri.fr/∼allouche

Abstract. Is it possible to mathematically define words like: *randomness, chaos, disorder, irregularity, complexity*, or like: *determinism, order, periodicity, regularity, simplicity*? Are there concepts in between (*quasi-periodicity*)? How do these concepts fit objects from physics, e.g., *glasses, crystals, quasi-crystals*? We try to describe and compare various notions used in mathematics.

1 A Few "Principles"

In what follows we restrict our study to infinite sequences taking their values in a finite set (sometimes called an *alphabet*). To deal with more complicated objects (e.g., functions) is possible by "hacking" them and restricting oneself to the "caricature" of the object: if an object is random, all its caricatures should also be random.

We begin with a few principles.

- No *general* definition of randomness is given in mathematics. Only definitions suitable for a given purpose can be found.
- No *algorithm* should produce "randomness". (Algorithm here means: finite sequence of instructions taken from a finite set of possible instructions.) Algorithms could only produce quasi-randomness or pseudo-randomness.
- No "context-free" algorithm applied to a random sequence should give a deterministic sequence. (Context-free here means the algorithm does not depend on the sequence. For example taking every second term of a random sequence should give a random sequence. Of course taking each 1 out of a binary sequence will *not* give a random sequence: this algorithm is not context-free.)

We now mention the question of "representation". To take an example from number theory, let $\tau = (1 + \sqrt{5})/2 = 1.618\ldots$ be the golden ratio. Asking whether this number is random can mean various things, since this number can be represented in various ways.

* allouche@bri.fr

M. Planat (Ed.): LNP 550, pp. 345–356, 2000.
© Springer-Verlag Berlin Heidelberg 2000

- Is it true that, for any positive integer ℓ, any finite block of ℓ symbols taken from the set $\{0,1,2,3,4,5,6,7,8,9\}$ occurs in the decimal expansion of τ with frequency $1/10^\ell$? (In other words, is the real number τ *normal* in base 10?)
- What can be said about the formal power series $1+6X+X^2+8X^3+\ldots$ obtained from the decimal digits of τ (for $X = 1/10$ the numerical value of this series is τ)? For example is it transcendental over $\mathbf{Q}(X)$? Another question could be: let p be a prime number; is the series $1+6X+X^2+8X^3+\ldots$ mod p transcendental over $\mathbf{Z}/p\mathbf{Z}(X)$?
- Is the sequence of partial quotients in the continued fraction expansion of τ "random"? The reader knows that the answer is no, since $\tau = [1,1,\ldots,1,\ldots]$.
- What can be said about the expansion of τ in base b, where $b \neq 10$, or even b non-integer, for example... $b = \tau$? (The seminal papers for expansions in non-integer bases are [31] and [28].)

2 Algebraic Randomness

We first discuss algebraic notions.

2.1 Block complexity

A first intuition is that a sequence is more complicated if many different finite blocks occur in it. This leads to the following definition.

Definition 2. The (block-)complexity of a sequence with values in a finite alphabet is the function $k \longrightarrow p(k)$, where $p(k)$ is the number of (different) blocks of length k that occur in the sequence.

Remark 5. If a sequence takes its values in a d-letter alphabet, its complexity satisfies $1 \leq p(k) \leq d^k$, for all $k \geq 1$. As easily seen, if a sequence is ultimately periodic (i.e., periodic from some point on) then its complexity is ultimately constant. Moreover a "random" sequence should have maximal complexity $p(k) = d^k$, for all $k \geq 1$, since all possible blocks should occur in the sequence. The converse is not true: consider the Champernowne number obtained by concatenating in order the base 10 expansions of the natural integers:

$$0.1234567891011121314\ldots$$

The sequence of digits of this number has complexity 10^k, but it cannot be considered as random.

The following result is due to Morse and Hedlund (see [26] and [27]). It shows that a sequence with very low complexity is very far from being random.

Proposition 1. *If the complexity of a sequence satisfies*

$$\exists k \geq 1, \quad p(k) \leq k,$$

then the sequence is ultimately periodic. (Hence its complexity must be ultimately constant.)

Remark 6. The "simplest" non-periodic sequences should thus satisfy the inequality $p(k) \geq k + 1$, for all $k \geq 1$. One might even ask whether it is possible to find sequences with complexity $p(k) = k + 1$ for all $k \geq 1$. Such sequences should take their values in a two-letter alphabet (take $k = 1$). It was proved by Morse and Hedlund [27] that these sequences, called *Sturmian sequences*, are exactly the sequences obtained by cutting the infinite two-dimensional lattice $\mathbf{Z}^2$ by a straight line with irrational slope, and coding its intersections with horizontal lines and with vertical lines by two different letters. Note that Sturmian sequences are sometimes considered as a one-dimensional model for *quasi-crystals*.

To know more about complexity of sequences, the reader can look at the survey [3]. Of interest is also [20].

2.2 Some algebraic "algorithms"

Instead of trying to define randomness, we may try to define non-randomness. In particular sequences generated by "algorithms" should be considered non-random. We briefly describe below two kinds of algorithms.

Morphisms and finite automata

We first describe three morphisms.

- Define the map μ by $\mu(0) = 01$, $\mu(1) = 10$, and μ applied to a word (i.e., a string of symbols 0 and 1) by the concatenation in order of the images by μ of the letters of the word (e.g., $\mu(001) := \mu(0)\mu(0)\mu(1) = 010110$). Hence μ is a *morphism* of the monoid $\{0, 1\}^*$ that consists of all words on $\{0, 1\}$ (including the empty word) equipped with the concatenation rule. This morphism has *constant length*: the images of single letters all have same length (here length 2). Now iterating μ starting from 0 gives

$$\begin{array}{l} 0 \\ 0\ 1 \\ 0\ 1\ 1\ 0 \\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1 \\ \vdots \end{array}$$

If we iterate an infinite number of times, we obtain an infinite sequence that is a fixed point of the morphism μ (extended to infinite words)

$$0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ \ldots$$

This sequence is called the Prouhet-Thue-Morse sequence (see [29], [37], [38], and [25]). It occurs in many mathematical questions that seem unrelated, see [7].

- Define as above the morphism τ on $\{0,1\}^*$ by $\tau(0) = 01$, $\tau(1) = 0$. This morphism does not have constant length. Iterating τ yields the Fibonacci sequence

$$0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ \ldots$$

It can be proved that this sequence is Sturmian, with a slope equal to the golden ratio. This sequence is the most popular example of a one-dimensional quasi-crystal.

- Define the morphism σ on the alphabet $\{a, b, c, d\}$ by

$$\begin{aligned}\sigma(a) &= ab\\ \sigma(b) &= ac\\ \sigma(c) &= db\\ \sigma(d) &= dc\end{aligned}$$

Let φ the map from $\{a, b, c, d\}$ to $\{-1, +1\}$ defined by

$$\begin{aligned}\sigma(a) &= +1\\ \sigma(b) &= +1\\ \sigma(c) &= -1\\ \sigma(d) &= -1\end{aligned}$$

Iterating σ as above yields a fixed point on $\{a, b, c, d, \}$

$$a\ b\ a\ c\ a\ b\ d\ b\ a\ b\ a\ c\ d\ c\ a\ c\ \ldots$$

Taking the (letterwise) image of this sequence by the map φ gives the Rudin-Shapiro sequence ([36] and [32])

$$+\ +\ +\ -\ +\ +\ -\ +\ +\ +\ +\ -\ -\ -\ +\ -\ \ldots$$

Such a construction (letterwise image of an infinite fixed point of a morphism of length 2 gives what is called a 2-*automatic* sequence. We will see below that the Rudin-Shapiro sequence though clearly "deterministic" has some aspects of a random sequence.

The reader can find more details on d-automatic sequences, as well as the finite automata formalism, in [15], [17], [1], [6], and [35].

Cellular automata

We will not give here a general definition of cellular automata. The study of two-dimensional patterns generated by the time evolution of one-dimensional linear cellular automata is a very popular subject. The reader can, for example, look at [4] and the references therein. A typical example is given by the Pascal triangle reduced modulo an integer $m \geq 2$. For $d = 2$, the time evolution pattern is closely related to the Sierpinsky triangle. It begins with

$$
\begin{array}{l}
1 \\
1\ 1 \\
1\ 0\ 1 \\
1\ 1\ 1\ 1 \\
1\ 0\ 0\ 0\ 1 \\
1\ 1\ 0\ 0\ 1\ 1 \\
1\ 0\ 1\ 0\ 1\ 0\ 1 \\
1\ 1\ 1\ 1\ 1\ 1\ 1\ 1 \\
\ldots
\end{array}
$$

Some links between these algorithms

The complexity function of a sequence generated by a morphism cannot be very large. We have the following results, (see [15] and [18]).

Theorem 1 (Cobham). *The complexity of a d-automatic sequence satisfies $p(k) = O(k)$.*

Theorem 2 (Ehrenfeucht, Lee, and Rozenberg). *The complexity of a sequence that is a fixed point of any morphism (not necessarily of constant length) satisfies $p(k) = O(k^2)$.*

Some links exist between automatic sequences and sequences generated by linear cellular automata as above. We first indicate that it is possible to define morphisms of constant length d and their fixed points (hence also d-automatic sequences as letterwise images of such fixed points) in dimension 2 (see for example [33], and [34]). Define e.g., the two-dimensional morphism θ on $\{0, 1\}$ by

$$\theta(0) = \begin{array}{l} 0\ 1 \\ 1\ 0 \end{array} \qquad \theta(1) = \begin{array}{l} 1\ 0 \\ 0\ 1 \end{array}$$

The map θ can be extended to a "morphism", and iterated starting from 0, yielding

$$\theta(0) = \begin{matrix} 0\,1 \\ 1\,0 \end{matrix} \quad \theta^2(0) = \begin{matrix} 0\,1\,1\,0 \\ 1\,0\,0\,1 \\ 1\,0\,0\,1 \\ 0\,1\,1\,0 \end{matrix} \quad \theta^3(0) = \begin{matrix} 0\,1\,1\,0\,1\,0\,0\,1 \\ 1\,0\,0\,1\,0\,1\,1\,0 \\ 1\,0\,0\,1\,0\,1\,1\,0 \\ 0\,1\,1\,0\,1\,0\,0\,1 \\ 1\,0\,0\,1\,0\,1\,1\,0 \\ 0\,1\,1\,0\,1\,0\,0\,1 \\ 0\,1\,1\,0\,1\,0\,0\,1 \\ 1\,0\,0\,1\,0\,1\,1\,0 \end{matrix} \quad \dots$$

The reader may iterate, starting from 1, the map γ defined by

$$\gamma(0) = \begin{matrix} 0\,0 \\ 0\,0 \end{matrix} \quad \gamma(1) = \begin{matrix} 1\,0 \\ 1\,1 \end{matrix}$$

and guess the following result (see [19], and [5]).

Theorem 3. *Let m and d be two integers ≥ 2. The Pascal triangle reduced modulo m is a two-dimensional d-automatic sequence if and only if there exists a prime number p such that m and d are both powers of p.*

2.3 Algebraicity and transcendence

Does algebraicity or transcendence of a number tell something about the complexity of its expansion? As we said in the first section, we have to be precise on what is meant by "expansion".

The Thue-Morse sequence revisited

The Thue-Morse sequence was defined above. It begins with

$$0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\ \dots$$

We can first consider the real number whose base b expansion is

$$0.110100110010110\dots$$

This number is transcendental [23] (see also [16]). We can also look at the formal power series in $\mathbf{Q}[[X]]$ whose coefficients are given by the Thue-Morse sequence

$$X + X^2 + X^4 + X^7 + X^8 + X^{11} + X^{13} + X^{14} \dots$$

This series can be proved transcendental over $\mathbf{Q}(X)$, *but* if we consider it as a formal power series in $\mathbf{Z}/2\mathbf{Z}[[X]]$, it is algebraic over $\mathbf{Z}/2\mathbf{Z}(X)$. This can be proved directly by noting that the Thue-Morse sequence $(u_n)_{n\geq 0}$ satisfies for all $n \geq 0$ the relations $u_{2n} = u_n$ and $u_{2n+1} = 1 + u_n \bmod 2$, and by splitting the formal power series $\sum_{n=0}^{\infty} u_n X^n$ into two series (odd and even indices). This is also a consequence of Christol's theorem (see [12], see also [13]).

Theorem 4 (Christol). *Let p be a prime number and let $(u_n)_{n\geq 0}$ be a sequence with values in $\mathbf{Z}/p\mathbf{Z}$. The the formal power series $\sum_{n=0}^{\infty} u_n X^n$ is algebraic over $\mathbf{Z}/p\mathbf{Z}(X)$ if and only if the sequence $(u_n)_{n\geq 0}$ is p-automatic.*

Remark 7. Let $(u_n)_{n\geq 0}$ be again the Thue-Morse. Let p be a prime number $\neq 2$. One may ask whether the formal power series $\sum_{n=0}^{\infty} u_n X^n$ considered as an element of $\mathbf{Z}/p\mathbf{Z}[[X]]$, is algebraic over $\mathbf{Z}/p\mathbf{Z}(X)$. The answer is *no*, from a theorem of Cobham [14].

Real numbers with automatic base *b* expansion

An old standing conjecture asserts that an algebraic irrational real number is "normal" in any base (the definition of normality is given in Section 3.1 below). In particular any possible block of any length should appear in the base b expansion of such a number (in other words the complexity of the base b expansion of an algebraic irrational real number is conjecturally equal to b^k). A weak form of this conjecture consists of studying numbers that miss some blocks of digits. In particular, numbers whose base b expansion is an automatic sequence (or even the letterwise image of a fixed point of any morphism). Some results were obtained in this direction (see [22] although the paper has been reported to contain a gap, see also [21] and [8]).

Remark 8. The converse of the above conjecture is not true. The Champernowne number obtained by concatenating the decimal expansions of the integers $1, 2, 3, 4, 5, 6, 7, 8, 9, 10\ldots$ thus obtaining the real number

$$0.123456789101112\ldots$$

is normal in base 10 (see Section 3.1). It is easily proved irrational, since it contains arbitrarily long strings of 0's. It was proved transcendental by Mahler [24].

Real numbers with automatic continued fraction expansion

A conjecture that is still open asserts that the continued fraction of an irrational algebraic number has bounded partial quotients if and only if the number is quadratic (in this case the continued fraction expansion is ultimately periodic). This means that an irrational real number with bounded partial quotients is (conjecturally) either quadratic or transcendental. As above a weaker conjecture can be addressed: is it true that an irrational real number whose sequence of bounded partial quotients is automatic (or even is the letterwise image of a fixed point of any morphism) is either quadratic or transcendental? A recent paper [30] gives the following nice result (generalization to other sequences can be found in [2]).

Theorem 5 (Queffélec). *Let $a \neq b$ be two integers ≥ 2. Define a real number by the continued fraction expansion*

$$[a, b, b, a, b, a, a, b, b, a, a, b, a, b, b, a, \dots]$$

where the sequence of a's and b's is obtained from the Thue-Morse sequence by replacing 0's by a's and 1's by b's. Then this number is transcendental.

3 Analytic Randomness

3.1 Normality

If we toss a fair coin an infinite number of times, and look at the sequence of heads and tails we obtain, we expect that the frequency of heads and the frequency of tails are both equal to 1/2. We also expect that any given finite string of heads and tails of length k appears with frequency $1/2^k$. This leads to the following definition of *normality*, and to the informal claim that "randomness implies normality".

Definition 3. Let $b \geq 2$ be an integer. A real number is called *b-normal* if, for every $k \geq 1$, each block of length k occurs with frequency $1/b^k$ in the base b expansion of this number.

Note that it can be proved that almost all numbers are normal in all bases. The Champernowne number quoted above is given by its base 10 expansion

$$0.12345678910111213141 5\dots$$

This number is "easy" to compute; it was proved normal in [11].

Remark 9. To give an idea of how far we are of a proof of the conjecture quoted in the previous section that all irrational algebraic numbers are normal in any base, let us consider $\sqrt{2} = 1.4142\dots$ in base 10. It is not known whether each single digit occurs with frequency 1/10. It is not even known whether a given digit, say 4, occurs infinitely often in this decimal expansion!

3.2 Topological entropy

The notion of complexity introduced in Definition 2 is a counting property. It leads to the following definition of *entropy* for a sequence taking its values in a finite set.

Definition 4. Let $(u_n)_{n\geq 0}$ be a sequence taking its values in the finite set $\mathcal{A}$. Let $p(k)$ be its block complexity. The *topological entropy* of this sequence is defined by

$$h = \lim_{n\to\infty} \frac{\log p(n)}{n \log(\#\mathcal{A})}.$$

Remark 10. The limit above always exists (note that $p(m+n) \leq p(m)p(n)$). It always satisfies $0 \leq h \leq 1$. Of course low complexity sequences (e.g., Sturmian sequences, or automatic sequences) have zero entropy. Sequences with maximal complexity (e.g., the sequence of digits of a number normal in base b) have entropy 1.

The reader can find more information about topological entropy as well as other entropies in [9].

3.3 Fourier analysis

In this Section we survey some results on the Fourier analysis of infinite sequences. The reader can find more details in Part IV of Ref. [6].

The Fourier–Bohr spectrum

Definition 5. Let $(u_n)_{n\geq 0}$ be a sequence with values in $\mathbf{C}$ such that, for each real number λ, the quantity $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} u_n e^{-2i\pi n\lambda}$ exists. The *Fourier-Bohr spectrum* of the sequence $(u_n)_{n\geq 0}$ is the set

$$S = \left\{ \lambda;\ \widehat{u}(\lambda) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} u_n e^{-2i\pi n\lambda} \neq 0 \right\}.$$

Remark 11. The Fourier-Bohr spectrum of a sequence is countable.

The reader will easily check the following examples.

- If $u_n = e^{2i\pi \log n}$ $(n \geq 1)$, then $\widehat{u}(0)$ does not exist.
- If $u_n = (-1)^{[\alpha n]}$, where $[y]$ is the integer part of y, and α is an irrational number, then $S = \{\frac{\alpha}{2} + k\alpha \bmod 1;\ k \in \mathbf{Z}\}$.
- If $(u_n)_{n\geq 0}$ is the sequence obtained from the Thue-Morse sequence by replacing 0's by 1's and 1's by -1's, then $\widehat{u}(\lambda) = 0$ for every $\lambda \in \mathbf{R}$.
- If $(u_n)_{n\geq 0}$ is the sequence obtained from the Rudin-Shapiro sequence by replacing 0's by 1's and 1's by -1's, then $\widehat{u}(\lambda) = 0$ for every $\lambda \in \mathbf{R}$.

The Wiener spectrum

Definition 6. Let $(u_n)_{n\geq 0}$ be a sequence with values in $\mathbf{C}$ such that, for each integer k, the quantity $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \overline{u_n} u_{n+k}$ exists. The *Wiener spectrum* of the sequence $(u_n)_{n\geq 0}$ is the measure σ defined on the torus $\mathcal{R}/\mathcal{Z}$ by its Fourier transform

$$\widehat{\sigma}(k) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \overline{u_n} u_{n+k}.$$

The Wiener spectrum of the sequence $(u_n)_{n\geq 0}$ is the set

$$WS = \{\lambda;\ \sigma(\{\lambda\}) \neq 0\}.$$

Remark 12. The existence of the measure σ is given by Bochner's theorem. This measure is called the *spectral measure* of the sequence $(u_n)_{n\geq 0}$.

We have the following interesting theorem [10].

Theorem 6 (Bertrandias). *Let $(u_n)_{n\geq 0}$ be a sequence for which the hypotheses of Definitions 5 and 6 are fulfilled. Then, for each real number λ,*

$$\widehat{u}(\lambda) \leq \sqrt{\sigma(\{\lambda\})}.$$

Remark 13. This theorem has an immediate corollary: let S and WS be the Fourier-Bohr spectrum and the Wiener spectrum of a sequence satisfying the hypotheses of Definitions 5 and 6, then

$$S \subset WS.$$

A few examples

The measure σ introduced in Definition 6 above can be uniquely written (as any measure on the torus) as the sum of three measures

$$\sigma = \sigma_d + \sigma_{ac} + \sigma_{sc}$$

where σ_d is discrete, σ_{ac} absolutely continuous with respect to the Lebesgue measure, and σ_{sc} singular continuous (i.e., each point has measure zero, and the support of σ_{sc} has Lebesgue measure zero). We give below the spectral measure of some sequences.

- The spectral measure of almost all sequences on, say, $\{0, 1\}$ is equal to the Lebesgue measure (in other words the spectral measure of a "random" sequence is equal to the Lebesgue measure).
- The spectral measure of the sequence $(e^{2i\pi\alpha n^2})_{n\geq 0}$, where α is irrational, is the Lebesgue measure.
- The spectral measure of the Thue-Morse sequence with values ± 1 is singular continuous.
- The spectral measure of the Rudin-Shapiro sequence with values ± 1 is the Lebesgue measure. (Hence this sequence behaves like a random sequence.)
- The spectral measure of the Fibonacci sequence is discrete.

References

1. Allouche, J.-P. (1987) Automates finis en théorie des nombres. Exposition. Math. **5**, 239–266
2. Allouche, J.-P., Davison, L., Queffélec M. (1999) Transcendence of Sturmian or morphic continued fractions. In preparation
3. Allouche, J.-P. (1994) Sur la complexité des suites infinies. Bull. Belg. Math. Soc. **1**, 133–143

4. Allouche, J.-P., Haeseler, F. von, Peitgen, H.-O., Petersen, A., Skordev, G. (1997) Automaticity of double sequences generated by one-dimensional linear cellular automata. Theoret. Comput. Sci. **188**, 195–209
5. Allouche, J.-P., Haeseler, F. von, Peitgen, H.-O., Skordev, G. (1996) Linear cellular automata, finite automata and Pascal's triangle. Disc. Appl. Math. **66**, 1–22
6. Allouche, J.-P., Mendès France, M. (1995) Automata and automatic sequences in: Beyond quasicrystals, Éd. F. Axel, D. Gratias. Springer/Les Éditions de Physique, pp. 293–367
7. Allouche, J.-P., Shallit J. (1999) The ubiquitous Prouhet-Thue-Morse sequence. Proceedings of SETA'98, to appear
8. Allouche, J.-P., Zamboni, L. Q. (1998) Algebraic irrational binary numbers cannot be fixed points of non-trivial constant length or primitive morphisms. J. Number Theory **69**, 119–124
9. Berthé, V. (1995) Entropy in deterministic and random systems in: Beyond quasicrystals, Éd. F. Axel, D. Gratias. Springer/Les Éditions de Physique, pp. 441–463
10. Bertrandias, J.-P. (1966) Espaces de fonctions bornées et continues en moyenne asymptotique d'ordre p. Bull. Soc. Math. France, Suppl., Mém. **5**, 1–106
11. Champernowne, D. G. (1933) The construction of decimals normal in the scale of ten. J. London Math. Soc. **8**, 254–260
12. Christol G. (1979) Ensembles presque périodiques k-reconnaissables. Theoret. Comput. Sci. **9**, 141–145
13. Christol, G., Kamae, T., Mendès France, M., Rauzy, G. (1980) Suites algébriques, automates et substitutions. Bull. Soc. Math. France **108**, 401–419
14. Cobham, A. (1969) On the base-dependence of sets of numbers recognizable by finite automata. Math. Systems Theory **3**, 186–192
15. Cobham, A. (1972) Uniform tag sequences. Math. Systems Theory **6**, 164–192
16. Dekking, F. M. (1977) Transcendance du nombre de Thue-Morse. C. R. Acad. Sci. Paris Sér. I **285**, 157–160
17. Dekking, F. M., Mendès France, M., Poorten, A. J. van der (1982) Folds!. Math. Intelligencer **4**, 130–138, 173–181, 190–195
18. Ehrenfeucht, A., Lee, K. P., Rozenberg, G. (1975) Subword complexities of various classes of deterministic developmental languages without interaction. Theoret. Comput. Sci. **1**, 59–75
19. Korec, I. (1990) Pascal triangles modulo n and modular trellises. Comput. Artif. Intell. **9**, 105–113
20. Ferenczi, S., Kása, Z. (1999) Complexity for finite factors of infinite sequences. Theoret. Comput. Sci. **218**, 177–195
21. Ferenczi, S., Mauduit, C. (1997) Transcendence of numbers with a low complexity expansion. J. Number Theory **67**, 146–161
22. Loxton, J. H., Poorten, A. J. van der (1988) Arithmetic properties of automata: regular sequences. J. Reine Angew. Math. **392**, 57–69
23. Mahler, K. (1929) Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen. Math. Annalen **101**, 342–266. (1930) Corrigendum **103**, 532
24. Mahler, K. (1937) Arithmetische Eigenschaften einer Klasse von Dezimalbrüchen. Proc. Konin. Neder. Akad. Wet. **40**, 421–428
25. Morse, M. (1921) Recurrent geodesics on a surface of negative curvature. Trans. Amer. Math. Soc. **22**, 84–100

26. Morse, M., Hedlund, G. A. (1938) Symbolic Dynamics. Amer. J. Math. **60**, 815–866
27. Morse, M., Hedlund, G. A. (1940) Symbolic Dynamics II. Sturmian trajectories. Amer. J. Math. **62**, 1–42
28. Parry, W. (1960) On the β-expansions of real numbers. Acta Math. Acad. Sci. Hung. **11**, 401–416
29. Prouhet, E. (1851) Mémoire sur quelques relations entre les puissances des nombres. C. R. Acad. Sci. Paris Sér. I **33**, 225
30. Queffélec, M. (1998) Transcendance des fractions continue de Thue-Morse. J. Number Theory **73**, 201–211
31. Rényi, A. (1957) Representations for real numbers and their ergodic properties. Acta Math. Acad. Sci. Hung. **8**, 477–493
32. Rudin, W. (1959) Some theorems on Fourier coefficients. Proc. Amer. Math. Soc. **10**, 855–859
33. Salon, O. (1987) Suites automatiques à multi-indices et algébricité. C. R. Acad. Sci. Paris Sér. I **305**, 501–504
34. Salon, O. (1989) Propriétés arithmétiques des automates multidimensionnels. Thèse, Université Bordeaux I.
35. Shallit, J. (1988) A generalization of automatic sequences. Theoret. Comput. Sci. **61**, 1–16
36. Shapiro, H. S. (1951) Extremal problems for polynomials and power series. Thesis, MIT
37. Thue, A. (1906) Über unendliche Zeichenreihen, Norske vid. Selsk. Skr. Mat. Nat. Kl. **7**, 1–22. Reprinted in: (1997) Selected mathematical papers of Axel Thue, T. Nagell, ed., Universitetsforlaget, Oslo, pp. 139–158
38. Thue, A. (1912) Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norske vid. Selsk. Skr. Mat. Nat. Kl. **1**, 1–67. Reprinted in: (1997) Selected mathematical papers of Axel Thue, T. Nagell, ed., Universitetsforlaget, Oslo, pp. 413–478

From Symbolic Dynamics to a Digital Approach: Chaos and Transcendence

K. Karamanos *

Centre for Nonlinear Phenomena and Complex Systems C.P. 231, Campus Plaine ULB, Boulevard du Triomphe, B-1050, Brussels, Belgium

Abstract. We review recent progress on Feigenbaum attractors and their interconnection with Number Theory. We further enlight the relation between Chaos and Transcendence.

1 Introduction

What has a deterministic dynamical system to do with noise? In the beginning of this century, H. Poincaré and J. Hadamard have shown that under some not very restrictive conditions a purely deterministic system can present an extremely irregular, unpredictable and erratic behavior. Moreover, this may depend very much on the variations of the initial conditions and/or control parameters. This phenomenon whereby the behavior of a dynamical system is indeed reminiscent of noise has been dubbed "deterministic chaos", although the term "chaos" has not yet been accepted by many mathematicians.

We shall be interested in the connection of the behavior of a "mechanical" view of a system (represented at this level by an endomorphism, that is a function $f : [0,1] \to [0,1]$) with the corresponding "thermal" view (represented by the long-time statistical properties of the iterations $x_{n+1} = f(x_n)$ of the endomorphism).

To be more concrete, the minimal non-linearity one can introduce is a quadratic one, leading to the *logistic map*

$$x_{n+1} = r x_n (1 - x_n) \tag{1}$$

where r is the control parameter, $r \in [0,4]$. The continuous time form of this map, the *logistic equation*, was introduced in 1845 by the Belgian sociologist and mathematician P.-F. Verhulst to model the growth of populations limited by finite resources. The logistic map has been essentially introduced in the literature of mathematical physics by Metropolis, Stein and Stein (MSS) in a paper considered classical by now [14], but became really famous only after Feigenbaum's work [17,18].

* kkaramanos@ulb.ac.be

M. Planat (Ed.): LNP 550, pp. 357-371, 2000.
© Springer-Verlag Berlin Heidelberg 2000

2 Symbolic Dynamics at the Feigenbaum Points

MSS actually considered the logistic map as a member of a more general class of functions, called unimodal maps. Let us summarize some key properties of the familly of unimodal maps defined as follows:

(i) They are functions $f_\lambda : [0,1] \to [0,1]$, depending on a real parameter λ (control parameter),

(ii) f_λ is continuous and (at least) piecewise differentiable in $[0,1]$,

(iii) f_λ is convex and has a unique maximum but is otherwise arbitrary. Let c the position of the maximum, that is if $x \neq c : f(x) < f(c)$, $f'(x) > 0$ if $x < c$ and $f'(x) < 0$ if $x > c$. A particular iterate x_n will be said to be of "0" or of L type, and of "1" or of R type, according as $x_n < c$ or $x_n > c$, respectively. Given an initial condition x_o, the "minimum distinguishing information" about the sequence of iterates x_n will consist in a *pattern* of 0's (L's) and 1's (R's). Note that there are many possible ways to partition the phase space and many candidates as initial points. However, according to the Julia theorem [12], the partition $\big([0,c] \cup (c,1]\big)$ and the initial point $x_o = c$ turn out both to be the most fruitful from the point of view of the extraction of information available by the system. More specifically, the resulting coarse-grained path (symbolic dynamics [1–5,8,9]) is in a one-to-one correspondence with the actual trajectory and hence there is no loss of information in the topological sense. We are thus in position to completely characterize the dynamical system as an information generator by examining its symbolic dynamics only. This is why this partition is called a "generating" one.

For polynomial maps a stronger result, called Fatou and Julia theorem [12,13], holds: for appropriate λ's one can further localize the stable periodic orbits starting from the "critical" point c (where $f'(c) = 0$) and iterating it.

Based on this theoretical framework one can develop a whole machinery for the construction and the classification in an ordered way of the "patterns" of the (super)stable periodic orbits of any period. This task has been accomplished first in [14] for finite limit sets and later completed in [15] in order to include infinite limit sets.

Let P be a pattern associated with the m-period. By definition the (first) harmonic of P is the pattern $\hat{H}(P) = P\mu P$, where $\mu = L$ if P contains an odd number of R's, and $\mu = R$ otherwise. The procedure can be iterated, so that one may speak of the second, third,..., j-th harmonic, etc, hereafter denoted as $\hat{H}^j(P)$. MSS also introduce the $\hat{H}$-extension of a pattern P as the pattern generated by iterating the harmonic construction applied j times to P, when j increases indefinitely. In the sequel we shall rather adopt the notation $\hat{H}^\infty(P)$ for this asymptotic pattern.

MSS [14] prove that, if P is allowed, $\hat{H}(P)$ is allowed too (their theorem 1). Furthermore in their universal ordering we have $P < \hat{H}(P)$ and the harmonics are adjacent, that is, no allowed sequence exists between P and

$\hat{H}(P)$. Upon iterating the process

$$\forall j, \quad P < \hat{H}(P) < \hat{H}^2(P) < ... < \hat{H}^j(P). \tag{2}$$

This is the famous period-doubling scenario or period-doubling route to chaos. For the shake of illustration we list hereafter the first few harmonics associated with the 2^k and $3 \cdot 2^k$ cycles.

- The 2^∞ cycle.
$P \equiv R$ (2 period) $\rightarrow \hat{H}(R) = RLR$ (4 period) $\rightarrow$
$\hat{H}^2(R) = RLRRRLR = RLR^3LR$ (8 period) $\rightarrow$
$\hat{H}^3(R) = RLRRRLRLRLRRRLR = RLR^3LRLRLR^3LR$ (16 period) $\rightarrow$
$\hat{H}^4(R) = RLRRRLRLRLRRRLRRRLRRRLRLRLRRRLR =$
$RLR^3LRLRLR^3LR^3LR^3LRLRLR^3LR$ (32 period)

- The $3 \cdot 2^\infty$ cycle.
$P \equiv RL$ (3 period) $\rightarrow H(RL) = RLLRL$ (6 period) $\rightarrow$
$\hat{H}^2(RL) = RLLRLRRLLRL = RL^2RLR^2L^2RL$ (12 period) $\rightarrow$
$\hat{H}^3(RL) = RL^2RLR^2L^2RLLRL^2RLR^2L^2RL$ (24 period)

Yet, the question about the coarse-grained statistical properties of the resulting Feigenbaum attractors (attractors of the accumulation points of the $m \cdot 2^k$ cycles) remained open. In view of the particular importance of the period-doubling route to chaos [1–4], several authors tried to examine the symbolic dynamics of the corresponding attractors and some fragmentary, semi-empirical results have been reported [24–27]. In our paper [16], we have derived the statistical properties of the Feigenbaum attractors on a systematic basis. Note that for these attractors the so-called zeta-function formalism does not apply, because the system is non-hyperbolic and its Lyapounov exponent strictly vanishes [10,11].

One can do some more precise statements on that. It has been shown in [11] that at the accumulation point of the 2^k cycles the reciprocal (topological) zeta function equals

$$\zeta^{-1}(r_\infty, t) = (1-t)^2(1-t^2)(1-t^4)(1-t^8)... \tag{3}$$

where each new cycle contributes a cyclotomic polynomial of the form $(1-t^{2^n})$ as a factor in the infinite product. The accumulation points predicted by the MSS algorithm are conjectured to be unique [11,15].

For the 2^∞ accumulation (Feigenbaum) point of the logistic map one has $r_\infty \simeq 3.56994567$ (with 8 digits) [25]. The arithmetic nature of r_∞ is not known for the moment. We do not know whether the control parameter values corresponding to the accumulation points are rational or irrational numbers, and if this is the case if they are algebraic irrational or transcendental.

3 Self–Similarity

In [16] we established certain invariance and self-similarity properties of the symbolic sequences associated to the $m \cdot 2^{\infty}$ attractors. To do this we introduced a set of operators acting on these symbolic sequences. Taking advantage of these properties we have then introduced a scheme mapping the original sequence into a new sequence of hypersymbols with well-defined statistical properties, obtained by the lumping of groups of original symbols.

More specifically, we have shown that the 2^{∞} sequence remains invariant under the replacements (or *lumping* or *hypersymbol formation*)

$$RR \to L, \quad RL \to R \tag{4}$$

starting from the beginning of the sequence. Actually this self-similarity property can be extended straightforwardly to guarantee invariance under a repetitive use of this replacement rule.

We turn next to the invariance properties of the $m \cdot 2^{\infty}$ sequences. Let P_m be a specific (fixed) pattern of the m-cycle (many different patterns may correspond to an m-cycle). We first perform the lumpings

$$P_m R \to R, \quad P_m L \to L \tag{5}$$

if P_m contains an even number of R's, or

$$P_m R \to L, \quad P_m L \to R \tag{6}$$

if P_m contains an odd number of R's, starting from the first P_m. One can then show that this action maps the $m \cdot 2^{\infty}$ sequence into the 2^{∞} sequence. The result can obviously be extendend to guarantee invariance under a first lumping (5)-(6) and then repetitive use of the lumpigs (4).

Note that the symbolic dynamics at the accumulation points turns out to be *exactly* self-similar, and not in a statistical sense. (Actually a similar (in view of [36]) property, *automaticity* has been already reported by mathematicians [21,22].) By a simple number-theoretic argument on the transfer of information from the trajectory to the topological level, this implies exact self-similarity (which means *fractality* [29]) of the actual path. Self-similarity in an approximate sense has been reported in the past by many authors, see for instance [19,20]. Again, we know almost nothing about the arithmetical properties of the fractal dimensionality D as a number.

4 Entropy Analysis and Complexity at a Feigenbaum Point

A very useful tool for analyzing symbolic sequences are entropy-like quantities. Of special interest are the *block entropies*, extending Shannon's classical

definition to the entropy of a sucession of states rather than of a single state [5,6].

We are interested in the properties of the block entropies at a Feigenbaum point $m \cdot 2^{\infty}$, in the light of the analysis of the preceding section.

So far there was no specification on the way the sequences are to be analyzed or "read". One could thus glide, take portions (lumping), jump, form hypersymbols or even make dynamical renormalization (applying majority rules, for instance) on the symbolic sequence. If the sequence is generated by a random processor, a well-known result of probability theory establishes that the nature of the process is crucially dependent on lumping [7]. One of our main objectives is to explore the relation between these various ways of reading and the statistical properties of Feigenbaum attractors.

Based on the Cantor-like structure of the attractor of the 2^{∞} cycle for the logistic map, and measuring the words by gliding along the sequence, Grassberger [19] proposed the following decimation scheme

$$H(n) = H(n/2) + 1, \tag{7}$$

for n even and larger than 2, giving

$$H(n) = \ln\left(\frac{3n}{2}\right), \tag{8}$$

for $n = 4, 8, 16, \ldots$. Note that the penalization of long words is in this case logarithmic, leading to a probability of occurrence

$$p_n(A_1, ..., A_n) \sim \frac{1}{n} \tag{9}$$

for a word of length n, which is far less restrictive than for the case of developed chaos, where it is exponential.

Based on simple empirical rules, Ebeling and Nicolis achieved a full calculation of the word frequencies (incuding Grassberger's result as a special case) for this way of reading the sequence [25]. This work has been completed in [26,27], after the numerical observation of some apparent symmetries of the sequence and the formulation of an empirical rule for its generation. Moreover, the effects of different types of noise have been numericaly investigated [28].

In the light of this work Ebeling and Nicolis [25] propose a scaling of the type

$$H(n) = e + g(\ln n)^{\mu_1} \tag{10}$$

with $g \neq 0$ and $\mu_1 > 0$, for Feigenbaum attractors in general.

In [16], we have presented a new approach to this subject based on alternative decimation schemes.

Consider indeed a subsequence of length n selected out of a very long (theoretically infinite) symbolic sequence obtained by iterating the trajectory

of the underlying dynamical system. We stipulate that this subsequence is to be read in terms of distinct "words" of length ℓ,

$$\ldots \underbrace{A_1 \ldots A_\ell}_{B_1} \underbrace{A_{\ell+1} \ldots A_{2\ell}}_{B_2} \ldots \underbrace{A_{j\ell+1} \ldots A_{(j+1)\ell}}_{B_j} \ldots \tag{11}$$

Clearly, if the words just defined happen to be identical to the patterns appearing in the replacement rules (4) or (5)-(6), the original subsequence of A's will be reformulated entirely in terms of a limited number of hypersymbols B. Upon substituting the latter from the replacement rules one will then obtain a subsequence identical to the original one if $m = 2$, and identical to a subsequence of length n/l of the 2^∞ sequence if $m \geq 3$. This implies the following invariance property of the block entropies,

$$H_2(2^k) = H_2(2) = H_2(1) \tag{12}$$

$$H_m(rm) = H_2(r), \quad rm \neq 2^k \tag{13}$$

where the subscript m refers to the block entropy of a subsequence of the $m \cdot 2^\infty$ sequence.

As a corollary, the entropy per letter $h^{(n)}$ will diminish exponentially on the successive words of length 2^k of the 2^∞ sequence,

$$h_2^{(2^k)} = \frac{H_2(1)}{2^k} \to 0, \quad k \to \infty \tag{14}$$

whereas for the $m \cdot 2^\infty$ sequence one will have

$$h_m(mr) = \frac{H_2(r)}{m} \tag{15}$$

Until now we have been referring to the relations between block entropies of *different* symbolic sequences. One can further relate block entropies of *the same* symbolic sequence. Indeed, our operator formalism enables us to introduce the following entropic decimation scheme for subsequences of the full (infinite) sequence

$$H_m(m \cdot 2r) = H_m(m \cdot r) \quad (m \cdot 2^\infty \textit{sequence}) \tag{16}$$

which is to be compared with the decimation scheme for gliding.

In particular, for $m = 2$ one has

$$H_2(2 \cdot 2r) = H_2(2 \cdot r) \tag{17}$$

and for $m \cdot 2r = 2^k$ one recovers the first part of eq.(12). Decimating further once again, we recover the right hand side of eq.(12).

For the less trivial case $m = 3$, applying eq.(16) one finds

$$H_3(3 \cdot 2r) = H_3(3 \cdot r). \tag{18}$$

As an example, consider the subsequence

$$RLRRRLRL \tag{19}$$

which is part of the 2^∞ symbolic trajectory. Applying twice the rule (4) one obtains successively,

$$RLRR \tag{20}$$

$$RL \tag{21}$$

as stipulated, precisely, in (12). Notice that this reduction would be inapplicable, had the sequence been read by gliding one symbol each time as in [19,24,25].

We thus conclude that the very notion of the block entropy as a measure of *complexity* is subjective, as it depends on the way of reading, that is on the *observer* (see also [23]). At this moment there exist many open questions and conjectures related to these problems.

5 Other Coarse–Grained Statistical Properties at the Feigenbaum Points

We next proceed to the explicit evaluation of the probability mass of the symbols in a sequence, from which explicit values of the block entropies can be deduced.

Consider first the 2^∞ sequence. According to the MSS algorithm the percentage of R's is

$$\%\quad of\quad R's = \lim_{\nu\to\infty}\left(\frac{\left[\frac{2^{\nu+1}-1}{3}\right]}{2^\nu - 1}\right) = \frac{2}{3} \tag{22}$$

In other words, the a priori probabilities of R and L in a long sequence are

$$p_R = \frac{2}{3}, \quad p_L = \frac{1}{3} \quad (2^\infty \quad sequence) \tag{23}$$

These results are in agreement with the real space renormalization technique of Fraser and Kapral [20] but are derived in [16] in a much more straightforward and compact manner. These equations provide precious information for such a complicated mathematical object as the invariant density at the Feigenbaum point.

The procedure can be easily adapted to the $m \cdot 2^{\infty}$ cycle. Here we present the result for the case of the $3 \cdot 2^{\infty}$ cycle. Utilizing (23) one has

$$\%\quad of\quad R's = \frac{2\frac{1}{3} + \frac{2}{3}}{3\frac{1}{3} + 3\frac{2}{3}} = \frac{4}{9} \tag{24}$$

in other words

$$p_R = \frac{4}{9}, \quad p_L = \frac{5}{9} \quad (3 \cdot 2^{\infty} \quad sequence) \tag{25}$$

As a corollary, one is in position to evaluate explicitly the block entropies of any subsequence of length $m \cdot 2^k$. In particular, for the 2^{∞} sequence,

$$H(2^k) = H(2) = H(1) = -\left(\frac{1}{3}\ln\frac{1}{3} + \frac{2}{3}\ln\frac{2}{3}\right) \simeq 0.63651417 \tag{26}$$

(a transcendental constant).

For words of an odd legth, as explained in Sec 7 of [16], the entropies $H_2(r)$ coincide with the corresponding values calculated when gliding and can also be calculated in a constructive manner. For a general $m \cdot 2^{\infty}$ sequence the situation turns out to be much more complicated.

Actually, one may push the analysis much further and derive, using symbolic dynamics, the statistics of recurrence and escape times from the left and from the right halfs of the interval $(0, 1)$. This provides an extension of the results of Balakrishnan, Nicolis and Nicolis [30], limited to the regime of developed chaos. On the same grounds one can also show that the symbolic dynamics at the Feigenbaum points does not correspond to a Markovian procces of any order.

Note that all the numbers involved in the expressions of the long-time coarse-grained statistical properties of the system (except from those defined by transcendental functions, as the block entropies) turn out to be simple fractions, that is rational numbers. This is a very interesting fact by itself.

6 Replacements and Morphisms

In [16] we have introduced a suggestive visualization of the self-similar symbolic dynamics at the Feigenbaum points, thus giving an interpretation of our result in terms of geometric arguments.

Still, a similar description in terms of arithmetical constants (such as the scaling exponents or the Feigenbaum constants in the renormalization group approach) was lacking.

The question so remains: Is it possible to express these results somehow in terms of numbers? The answer, as we shall see in this and the next Sections will be in the affirmative.

These last years there has been a renewal of interest on the connection between Number Theory and Finite Automata, mainly following A. Cobham's pioneering work [34–39]. In addition to theorems and important conjectures that relate number theory with the theory of automatic sequences, many results have been reported concerning automatic sequences which enjoy some invariance properties (very reminiscent indeed of these exposed in Sec 3 for Feigenbaum sequences), as for instance being the fixed points of non-trivial morphisms of constant length or of primitive *morphisms*. (A morphism is a set of replacement rules for each letter, see [39].)

In particular a very recent theorem by J.-P. Allouche and L. Q. Zamboni asserts either the rationality or the transcendence of numbers whose binary expansion is a fixed point of the above kind of morphisms [39]. The possibility to take advantage of these properties gives us an additional motivation to somehow invert our viewpoint and to represent the $m \cdot 2^{\infty}$ sequences as "tag" (that is automatic in A. Cobham's terminology) sequences generated by morphisms.

Note that in Sec 3 we had been given the MSS algorithm and we exploited its symmetries. The idea now is to generate the infinite symbolic sequences from an initial pattern and their symmetries.

The generation of the 2^{∞} sequence through the inverse of the replacements (4) is formally a quite non-trivial task. But, in [31] we have shown that if we denote as $Re^{-1}()$ the inverse operation, that is the replacements $R \rightarrow RL, L \rightarrow RR$ we have

$$\hat{H}^{\infty}(R) = (Re^{-1})^{\infty}(R) \tag{27}$$

Moreover, if we denote as $RE^{-1}()$ the replacements

$$R \rightarrow P_m L, \quad L \rightarrow P_m R \tag{28}$$

if P_m contains an odd nomber of R's, and

$$R \rightarrow P_m R, \quad L \rightarrow P_m L \tag{29}$$

if P_m contains an even number of R's, then we have

$$\hat{H}^{\infty}(P_m) = RE^{-1}[(Re^{-1})^{\infty}(R)] \tag{30}$$

This means that the symbolic sequences at the Feigenbaum points can be generated by the above *morphisms*. According now to a result of Cobham, one can explicitly construct a deterministic finite automaton which generates these sequences [36]. In other words these sequences are also *automatic*, a fact already known in the mathematical literature [21,22].

7 Digital Approach, Transcendence and Non-Normality

Let us consider now the number generated by the standard arithmotheoretic procedure when identifying R as the digit 1 and L as the digit 0 (the passage

from *symbolic dynamics* to the *digital approach*), and write down formally

$$x = 0.\hat{H}^{\infty}(1) \quad (base \quad 2) \tag{31}$$

or, more explicitly

$$x = 0.1011101... \quad (base \quad 2) \tag{32}$$

and more generally for an arbitrary $m \cdot 2^{\infty}$ sequence

$$x_{P_m} = 0.\hat{H}^{\infty}(P_m) \quad (base \quad 2) \tag{33}$$

where P_m is again expressed in terms of 0's and 1's instead of L's and R's.

We further need the following theorem [39]:

Theorem: (J.-P. Allouche and L.Q. Zamboni, 1998)
Let x a positive real number whose binary expansion is a fixed point of a morphism on the alphabet {0,1}. If the morphism is either of constant length ≥ 2 or primitive, then the number x is either rational or transcendental.

Utilizing this theorem and the self-similarity, we have shown in [31] that the universal (for all unimodal maps) numbers introduced in (32), (33) are transcendental. They cannot result as roots of a polynomial with rational coefficients. These sequences form an infinite countable set. Moreover one can show relatively easy that these numbers are all non-normal. This is in agreement with a widely accepted conjecture in Number Theory according to which algebraic irrational numbers are believed to be normal (see also [33,34,39–43]).

These results can also be rigorously extended in some other cases, as for the pure m^{∞} sequences [31,32].

8 On the Feigenbaum Constants δ and α

After all what we have presented in the three previous Sections, one could start to wonder about the arithmetic nature of the Feigenbaum constants δ and α.

Both constants are defined by an approximate real space renormalization procedure. The constant δ has to do with the spacing in the control parameter space of the successive values of occurrence of the superstable periodic orbits and can be roughly estimated by the bifurcation diagram. If we denote as $\{r_n\}$ this set of values, δ is defined as

$$\delta = \lim_{n \to \infty} \frac{r_n - r_{n-1}}{r_{n+1} - r_n} \tag{34}$$

and for the logistic map

$$\delta \simeq 4.669201609102990... \tag{35}$$

The constant α is related to the rescaling of the period doubling functional composition law and its value for the logistic map is

$$\alpha \simeq -2.502907875095892... \tag{36}$$

The two constants are not unrelated between them. A crude approximation gives

$$\delta \simeq \alpha^2 + \alpha + 1 \tag{37}$$

More accurate relations can be found by more refined renormalization group arguments. The values of the two constants depend only on the order of the maximum and have long been studied [20]. They are thus, for instance, *universal* for quadratic maps irrespectively of the exact way one writes down the map.

The constants δ and α for quadratic maps are actually known up to about 1000 digits. Despite this, due to the indirect way they are defined we ignore almost everything for their precise arithmetical nature.

9 The "Standard" Conjecture of Chaos

The general conjecture stated at the end of Sec 7, that algebraic irrational numbers are supposed to be normal, is not as innocent as it seems. It has some extremely profound consequences.

In fact, let us focus our attention to the case of a unimodal map, for instance of the logistic map in three important regimes [1–4]:

(i) In the chaotic region extended to the right of the Feigenbaum point,

(ii) In the intermittent region and

(iii) In the region of developed chaos with zero memory, just before the fully developed chaos.

If we still consider the numbers generated as in eq. (32), but this time by the symbolic dynamics of the iterates of the maximum

$$x(\lambda) = 0. \quad J(f(c)) \quad J(f(f(c)))... \quad J(f^{(n)}(c))... \quad (base \quad 2) \tag{38}$$

where J denotes the symbolic dynamics and c the critical point of the map (the point where $f'(c) = 0$), we observe that in general the attractors are asymmetric around the maximum $x_o = c$ and the coarse-grained invariant measure is in the general case not equidistributed on the two partitions R and L of the phase space, that is $p_R \neq p_L$. This implies that if we define as in eq. (38) the associated numbers for these three regions, these numbers cannot be normal. And as they are not normal they cannot be algebraic irrational. But they cannot be rational either, as a positive Lyapounov exponent means chaoticity - certainly an aperiodic occurrence of the digits 0 and 1 in the binary expansion (38).

Adding now to the previous cases the systems at the onset of chaos (at the Feigenbaum and other accumulation points) we arrive to a result of a great generality: the occurrence of Chaos is ultimately connected with transcendence, as far as numbers defined as in eq. (38) cannot be normal. Because as they are not normal they cannot be algebraic irrational.

Unfortunately this very simple argument does not cover the case of fully developed chaos (with Lyapounov exponent equal to $\ln 2$). In fact in the case of the logistic map in fully developed chaos, as well as in the case of the tent map, the iterates of the maximum are trapped by the unstable fixed point. To avoid this difficulty, we could thus consider in that case the symbolic dynamics of another, arbitrary point, not belonging to the exceptional set of points with poor ergodic properties (which is of zero measure). As now in fully developed chaos successive steps are completely uncorrelated [1–3], any outcome compatible with normality (that is $p_R = p_L = \frac{1}{2}$) is possible, and it is plausible to accept that any normal number, algebraic irrational or transcendental is a possible outcome.

This "typical" (or "standard" in view of its profound character and of the difficulty of its possible proof) conjecture of chaos, that chaos is ultimately related with transcendence, is a very important step towards the classification of the numbers defined by this digital approach, for 1d chaotic systems in general.

More specifically from a dynamical viewpoint transcendence manifests itself also through the existence of (weakly) chaotic systems, irrationality through the existence of fully developed chaotic systems and rationality through the existence of periodic orbits.

10 Conclusion

We have shown that the symbolic sequences generated by Feigenbaum attractors can be reformulated entirely in terms of hypersymbols. Under well-defined replacement rules the hypersymbol sequences could then be mapped into the original ones. We have analyzed the block entropies of the sequences and found that, under the convention that the sequence is to be read in terms of such hypersymbols, the entropies satisfy some well-defined invariance properties. Thus, the blurred image (due to the inherent spatial inhomogeneity of the corresponding attractor) observed at the trajectory level of description, is replaced by a set of clearcut rules at the level of symbolic dynamics.

The realization that the kind and the amount of information of a given symbolic sequence may depend on the way of reading brings symbolic dynamics closer to natural languages, in which the existence of distinct privileged words conveying a precise "meaning" is crucial. It would be interesting to reconsider the analysis of physiological time series including DNA or RNA sequences in the light of our results.

Moreover, based on very recent mathematical and computational arguments, we have observed an interesting but apparently unexpected connection between the chaoticity and the transcendental character of the corresponding symbolic sequence considered as a binary number. Transcendence has to do with rational polynomials and chaos with exponential sensitivity to initial conditions and erratic behavior. As this result is really new, we do not know how to interpret it.

It is conjectured that this is a quite general result and it also holds for continuous time flows and in higher dimensionalities. We thus conclude that *transcendence* is an appropriate measure of *complexity.*

This kind of results are of obvious importance in information theory, in theoretical computer science and in metrology.

Acknowledgements

It is a pleasure to thank G. Nicolis, J.-P. Allouche and M. Waldschmidt for illuminating discussions and suggestions. I would also like to thank the organizer of this École Thématique for the kind invitation. K.K. has been a NATO Fellow for the last three academic years and is currently a Van Buuren Foundation Fellow and a Petsalys-Lepage Foundation Fellow. This work has been supported in part by the Pôles d'Attraction Interuniversitaires program of the Belgian Federal Office of Scientific, Technical and Cultural Affairs.

References

1. Nicolis G., (1995) Introduction to Nonlinear Science. Cambridge University Press, Cambridge
2. Bai-Lin H., (1994) Chaos. World Scientific, Singapore
3. Schuster H.G., (1984) Deterministic Chaos. Physik-Verlag
4. Schröder M., (1991) Fractals, Chaos, Power Laws. Freeman, New York
5. Nicolis G. and Gaspard P., (1994) Toward a Probabilistic approach to Complex Systems. Chaos, Solitons & Fractals **4**, No. 1, 41–57
6. Khinchin A.I., (1957) Mathematical Foundations of Information Theory. Dover, New York
7. Kemeny J. and Snell L., (1976) Finite Markov Chains. Springer-Verlag, Berlin
8. Thom R., (1983) Paraboles et Catastrophes. Flammarion, Paris
9. Nicolis J.S., (1991) Chaos and Information Processing. World Scientific, Singapore
10. Eckmann J.P. and Ruelle D., (1985) Ergodic Theory of chaos and strange attractors. Rev. Mod. Phys. **57**, 617
11. Milnor J. and Thurston W., (1986) On iterated maps of the interval, in **1342** Dynamical Systems, Proceedings, University of Maryland, J.C. Alexander (Ed.), 465–563, Lecture Notes in Mathematics, Springer-Verlag; and references therein

12. Julia G., (1918) Mémoire sur l'itération des fonctions rationelles. J. de Math. (Liouville) Ser. 7, **4**, 47–245. The relevant theorem appears on p. 129ff
13. Fatou M.P., (1919) Sur les équations fonctionnelles. Bull. Soc. Math. France **47**, 161–271; (1920) **48**, 33–94, 208–314
14. Metropolis N., Stein M.L. and Stein P.R., (1973) On Finite Limit Sets for Transformations on the Unit Interval. Journal of Combinatorial Theory, Vol. A **15**, No. 1, 25–44
15. Derrida B., Gervois A. and Pomeau Y., (1978) Iteration of endomorphisms on the real axis and representation of numbers. Ann. Inst. Henri Poincaré, Section A: Physique Théorique, Vol. XXIX, n^o 3, 305–356
16. Karamanos K. and Nicolis G., (1999) Symbolic Dynamics and Entropy Analysis of Feigenbaum Limit Sets. Chaos, Solitons & Fractals, **10**, No 7, 1135–1150
17. Feigenbaum M., (1978) Quantitative Universality for a Class of Nonlinear Transformations. J. Stat. Phys. **19**, 25
18. Feigenbaum M., (1979) The Universal Metric Properties of Nonlinear Transformations. J. Stat. Phys. **21**, 669
19. Grassberger P., (1981) On the Hausdorff Dimension of Fractal Attractors. J. Stat. Phys. **26**, No. 1, 173
20. Fraser S. and Kapral R., (1985) Mass and dimension of Feigenbaum attractors. Phys. Rev. **A31**, 3, 1687
21. Christol G., Kamae T., Mendes-France M. and Rauzy G., (1980) Suites algébriques, automates et substitutions. Bull. Soc. Math. France **108**, 401–419
22. Allouche J.-P. and Cosnard M., (1983) Itérations de fonctions unimodales et suites engendrées par automates. C. R. Acad. Sc. Paris (Série I), **296**, 159–162
23. Grassberger P., (1986) Toward a Quantitative Theory of Self-Generated Complexity. Int. J. Theor. Phys. **25**, No. 9, 907
24. Ebeling W. and Nicolis G., (1991) Entropy of Symbolic Sequences: the Role of Correlations. Europhys. Lett. **14** (3), 191–196
25. Ebeling W. and Nicolis G., (1992) Word Frequency and Entropy of Symbolic Sequences: a Dynamical Perspective. Chaos, Solitons & Fractals **2**, 635
26. Freund J., Ebeling W. and Rateitschak K., (1996) Self-similar sequences and universal scaling of dynamical entropies. Phys. Rev. **E54**, 5, 5561–5566
27. Rateitschak K., Freund J. and Ebeling W., (1996) Entropy of sequences Generated by Nonlinear Processes: The Logistic Map, in Entropy and Entropy Generation. Shiner J.S. (Ed.), Kluwer Acad. Publ., 11–26
28. Freund J. and Rateitschak K., (1998) Entropy Analysis of Noise Contaminated Sequences. Int. J. Bif. Chaos **8(5)**, 933
29. Mandelbrot B.B., (1982) The Fractal Geometry of Nature. Freeman, San Franscisco
30. Balakrishnan V., Nicolis G. and Nicolis C., (1997) Recurrence Time Statistics in Chaotic Dynamics. I. Discrete time Maps. J. Stat. Phys., **86**, 191
31. Karamanos K., (1999) On the connection between 1D Chaotic Dynamics and Number Theory. From Symbolic Dynamics to a Digital Approach. Submitted for publication to *Nonlinearity*
32. Karamanos K., unpublished (1998). In fact the pure m^∞ sequences turn out to be self-similar under the repeated use of the replacements (21) and (22) of [16]. This means that $\hat{L}^n((P_m*)^\infty) = (P_m*)^\infty$ for any finite n, where $\hat{L}$ is the operator defined in Sec 4 of [16] and * is the composition rule defined in [15]. This can be shown by a method similar to that used in [16]

33. Chaitin G.J., (1994) Randomness and Complexity in Pure Mathematics. Int. J. Bif. Chaos **4**, No 1, 3–15
34. Allouche J.-P., (April 1987) Automates finis en Mathématiques et en Physique. Pour la Science **94**
35. Waldschmidt M., (1993) Introduction to recent results in Transcendental Number Theory. Lectures given at the Workshop and Conference in number theory held in Hong-Kong, June 29 - July 3 1993, preprint 074-93, M.S.R.I., Berkeley
36. Cobham A., (1972) Uniform tag sequences. Math. Systems Theory **6**, 164–192
37. Loxton J.H. and van der Poorten A.J., (1988) Arithmetic properties of automata: regular sequences. J. Reine Angew. Math. **392**, 57–69; Allouche J.-P., personal communication
38. Ferenczi S. and Maduit C., (1997) Transcendence of Numbers with a Low Complexity Expansion. J. of Number Theory **67**, 146–161
39. Allouche J.-P. and Zamboni L.Q., (1998) Algebraic Irrational Binary Numbers Cannot Be Fixed Points of Non-trivial Constant Length or Primitive Morphisms. J. of Number Theory **69**, 119–124
40. Wagon S., (1985) Is π normal? Math. Intelligencer **7**, 65–67
41. Klee V. and Wagon S., (1989) New and Old Unsolved Problems in Plane Geometry and Number Theory. Mathematical Association of America, Washington D.C.
42. Borwein J.M., Borwein P.B. and Bailey D.H., (March 1989) Ramanujan, Modular Equations, and Approximations to Pi, or How to Compute One Billion Digits of Pi. Am. Math. Monthly **96**, 201–219
43. Bailey D.H., Borwein J.M., Borwein P.B. and Plouffe S., (1997) The Quest for Pi. Math. Intelligencer **19**(1), 50–57

Algebraic Dynamics and Transcendental Numbers

Michel Waldschmidt *

Institut de Mathématiques de Jussieu, Université P. et M. Curie, Théorie des Nombres, Case 247, 4, Place Jussieu, F-75252 PARIS Cedex 05 France

Abstract. A first example of a connection between transcendental numbers and complex dynamics is the following. Let p and q be polynomials with complex coefficients of the same degree. A classical result of Böttcher states that p and q are locally conjugates in a neighborhood of ∞: there exists a function f, conformal in a neighborhood of infinity, such that $f(p(z)) = q(f(z))$. Under suitable assumptions, f is a transcendental function which takes transcendental values at algebraic points. A consequence is that the conformal map (Douady-Hubbard) from the exterior of the Mandelbrot set onto the exterior of the unit disk takes transcendental values at algebraic points. The underlying transcendence method deals with the values of solutions of certain functional equations.

A quite different interplay between diophantine approximation and algebraic dynamics arises from the interpretation of the height of algebraic numbers in terms of the entropy of algebraic dynamical systems.

Finally we say a few words on the work of J.H. Silverman on diophantine geometry and canonical heights including arithmetic properties of the Hénon map.

1 Transcendental Values of Böttcher Functions

For any complex number $c \in \mathbf{C}$, define the polynomial $p_c \in \mathbf{C}[z]$ by $p_c(z) = z^2 + c$. For $n \geq 1$, let p_c^n be the n-th iterate of p_c:

$$p_c^1(z) = p_c(z) = z^2 + c, \qquad p_c^2(z) = p_c(z^2 + c) = (z^2 + c)^2 + c,$$

$$p_c^n(z) = p_c^{n-1}(z^2 + c) \qquad (n \geq 2).$$

The *Mandelbrot set* M can be defined as

$$M = \big\{c \in \mathbf{C} \mid p_c^n(0) \text{ does not tend to } \infty \text{ as } n \to \infty\big\}.$$

In 1982, A. Douady and J. Hubbard showed that M is connected. They constructed a conformal map

$$\Phi : \mathbf{C} \setminus M \longrightarrow \{z \in \mathbf{C}\,;\, |z| > 1\}$$

from the complement of M onto the exterior of the unit disk, which is defined as follows.

* miw@math.jussieu.fr

M. Planat (Ed.): LNP 550, pp. 372–378, 2000.
© Springer-Verlag Berlin Heidelberg 2000

For each $c \in \mathbf{C}$, there is a unique power series φ_c with coefficients in $\mathbf{Q}(c)$,

$$\varphi_c(z) = z + c_0 + \frac{c_1}{z} + \frac{c_2}{z^2} + \cdots \in \mathbf{Q}(c)((1/z)),$$

such that

$$\varphi_c(z^2 + c) = \varphi_c(z)^2.$$

For $c \notin M$, φ_c defines an analytic function near c. Then the above mentioned map Φ is defined by $\Phi(c) = \varphi_c(c)$.

According to P.G. Becker, W. Bergweiler and K. Nishioka [2], [3], [11], *for any algebraic $\alpha \in \mathbf{C} \setminus M$, the number $\Phi(\alpha)$ is transcendental.* .

The function φ_c is the unique *Böttcher function* with respect to $p_c = z^2 + c$ and z^2. More generally, let

$$p = az^d + \cdots \quad \text{and} \quad q = bz^d + \cdots$$

be two polynomials in $\mathbf{C}[z]$ of degree $d \geq 2$ and let $\lambda \in \mathbf{C}$ satisfy $\lambda^{d-1} = a/b$. There exists a unique function f, which is defined and meromorphic in a neighborhood of ∞, such that

$$\lim_{z \to \infty} \frac{f(z)}{\lambda z} = 1 \quad \text{and} \quad f\big(p(z)\big) = q\big(f(z)\big)$$

for sufficiently large $|z|$. Such a conjugating function f is called a *Böttcher function with respect to p and q.*

Assume p and q have algebraic coefficients and are not linearly conjugate to monomials or Chebychev polynomials. Then *f is a transcendental function, which takes transcendental values at algebraic points.*

This result holds more generally for classes of analytic functions which satisfy certain functional equations. There are two methods to study the transcendence of values of such functions.

The first one originates in the solution, by Th. Schneider, of Hilbert's seventh problem on the value of the exponential function, which satisfies the functional equation $f(z_1 + z_2) = f(z_1)f(z_2)$. This method can be used to consider other functional equations, like

$$f(z^d) = af(z)^d + bz^h.$$

The second method has been introduced by K. Mahler (see [11]) and enables one to prove transcendence results for the values of analytic functions f which are solutions of more general functional equations, like

$$P\big(z, f(z), f(z^d)\big) = 0.$$

In the present situation, both methods provide the desired result.

2 Lehmer's Problem and the Entropy of Algebraic Dynamical Systems

Let $F \in \mathbf{Z}[X]$ be a monic polynomial of degree $d \geq 1$ with complex roots $\alpha_1, \ldots, \alpha_d$. Define, for any positive integer n,

$$\Delta_n(F) = \prod_{i=1}^{d} (\alpha_i^n - 1) \in \mathbf{Z}.$$

In case $F = X - 2$ we have $\Delta_n(F) = 2^n - 1$, and the prime values of the sequence $2^n - 1$ are the so-called Mersenne primes. In 1933 [7], D.H. Lehmer suggested that the sequence $\Delta_n(F)$ is likely to produce prime numbers, provided that it grows slowly. If no $|\alpha_i|$ is 1, then

$$\lim_{n\to\infty} \frac{\Delta_{n+1}(F)}{\Delta_n(F)} = \prod_{\substack{1\leq i\leq n\\ |\alpha_i|>1}} |\alpha_i|.$$

More generally, for a polynomial

$$F = a_0 X^d + \cdots + a_d = a_0 \prod_{i=1}^{d} (X - \alpha_i) \in \mathbf{C}[X],$$

define, with K. Mahler,

$$M(F) = |a_0| \prod_{\substack{1\leq i\leq n\\ |\alpha_i|>1}} |\alpha_i| = \exp \int_0^1 \log |F(e^{2i\pi t})| dt.$$

For any polynomial $F \in \mathbf{C}[X]$, we have $M(F) \geq 1$. When $F \in \mathbf{Z}[X]$, we have $M(F) = 1$ if and only all its roots α_i are either zero or roots of unity. For his calculations, Lehmer used the polynomial $F(X) = X^3 - X - 1$. It turns out that this actually is the polynomial having smaller measure > 1 among the non reciprocal polynomials (its root > 1 is the smallest Pisot-Vijiyaraghavan number). For reciprocal polynomials F, that is for F satisfying $F(X^d) = X^d F(1/X)$, Lehmer said he could not find a polynomial having smaller measure than $M(F_0) = 1.1762808183\ldots$, with

$$F_0(X) = X^{10} + X^9 - X^7 - X^6 - X^5 - X^4 - X^3 + X + 1 = X^5 Q\big(X + (1/X)\big)$$

and

$$Q(T) = (T+1)^2(T-1)(T+2)(T-2) - 1.$$

He asked whether for each $c > 1$ there is a polynomial $F \in \mathbf{Z}[X]$ for which $1 < M(F) \leq c$, and this open question is known as *Lehmer's problem.*

The number $M(F)$ has a dynamical interpretation, which is a bridge between the notion of height of a polynomial and ergodic theory [6], [12], [5].

For our purposes, an algebraic dynamical system is a continuous endomorphism $T : X \to X$ of a metrizable compact topological group. The easiest case, which will be sufficient for our purpose, is the torus $X = \mathbf{T}^d = \mathbf{R}^d/\mathbf{Z}^d$. Each continuous endomorphism of $\mathbf{T}^d$ is given by a $d \times d$ matrix A_T with integer coefficients, and

$$T\mathbf{x} \equiv A_T\mathbf{x} \pmod 1.$$

Automorphisms are given by a matrix A_T with determinant 1 of -1.

An endomorphism T is "ergodic" if, whenever a measurable subset B of X (for a Haar measure μ) satisfies $T^{-1}B = B$, we have $\mu(B) = 0$ or 1. In the torus case $\mathbf{T}^d$, this condition amounts to say that for every square integrable function f, the condition $f(Tx) = f(x)$ almost everywhere implies that f is constant almost everywhere.

It follows that an endomorphism T is ergodic is and only if no eigenvalue of A_T is a root of unity. So we shall be interested in polynomials (namely the characteristic polynomial $\chi(A_T)$ of A_T) with no root a root of unity.

The set of periodic points of period n under T is

$$\mathrm{Per}_n(T) = \{x \in \mathbf{T}^d \mid T^n(x) = x\}.$$

If T is ergodic, then the number of periodic points of period n is

$$|\mathrm{Per}_n(T)| = |\det(A_T^n - I)| = |\Delta_n(\chi(A_T))|.$$

The *topological entropy* of T can be defined in terms of the metric: for $\epsilon > 0$ denote by B_ϵ the ball around the origin with radius ϵ. Then

$$h(T) = \lim_{\epsilon \to 0} \lim_{n \to \infty} -\frac{1}{n} \log \mu \left(\cap_{j=0}^{n-1} T^{-j}(B_\epsilon) \right).$$

Denote by $\lambda_1, \ldots, \lambda_d$ the eigenvalues of A_T (counting multiplicities). Then Yuzvinskii's formula reads

$$h(T) = \sum_{i=1}^{d} \log \max\{1, |\lambda_i|\}.$$

Hence the entropy of T is nothing else than the logarithm of Mahler's measure of the characteristic polynomial $\chi(A_T)$ of A_T.

Since any monic polynomial $X^d + a_1 X^{d-1} + \cdots + a_d$ is the characteristic polynomial of a matrix, namely

$$A = \begin{pmatrix} 0 & & & \\ \vdots & & I_{d-1} & \\ 0 & & & \\ -a_d & -a_{d-1} & \cdots & -a_1 \end{pmatrix},$$

it follows that Lehmer's problem is equivalent to asking: *which values in* $[0, \infty]$ *can occur as an entropy?*

According to D.A. Lind, a positive answer to Lehmer's problem (i.e. the existence of polynomials in $\mathbf{Z}[X]$ with arbitrarily small $M(F) > 1$) is equivalent to the existence of a continuous endomorphism of the infinite torus $\mathbf{T}^{\mathbf{Z}}$ with finite entropy.

Interesting problems occur when one tries to replace the torus $\mathbf{R}/\mathbf{Z}$ by an elliptic curve [1].

This section has been prepared with the help of Paola D'Ambros.

3 Canonical Heights and Dynamical Systems

Define the *absolute multiplicative height* of a polynomial $F \in \mathbf{Z}[X]$ of degree $d > 0$ by

$$H(F) = M(F)^{1/d}$$

and the *absolute multiplicative height* of an algebraic number α by

$$H(\alpha) = H(F)$$

where $F \in \mathbf{Z}[X]$ is the minimal polynomial of α over $\mathbf{Z}$. The name is motivated by the property

$$H(\alpha^n) = H(\alpha)^n$$

for any algebraic number α and any positive integer n. Hence this height H behaves nicely with respect to the polynomials $\phi(X) = X^n$.

J.H. Silverman [18] introduced a height function which behaves nicely for an arbitrary rational function ϕ with algebraic coefficients, viewed as a map $\mathbf{P}_1(\overline{\mathbf{Q}}) \to \mathbf{P}_1(\overline{\mathbf{Q}})$.

Definition. Let $\phi \in \overline{\mathbf{Q}}(X)$ be a rational function of degree $n \geq 2$. The *ϕ-canonical height* of an algebraic number is

$$\hat{H}_\phi(\alpha) = \lim_{r \to \infty} \left(\phi^r(\alpha)\right)^{1/n^r}, \qquad (*)$$

where $\phi^r = \phi \circ \phi^{r-1}$ and ϕ^0 is the identity.

This construction has been introduced by Tate in his work on Abelian varieties, and has been extended to this general context by Silverman in a series of papers. He proved:

The limit $(*)$ *defining* $\hat{H}_\phi(\alpha)$ *exists, and*

$$\hat{H}_\phi(\phi\alpha) = \hat{H}_\phi(\alpha)^n.$$

Moreover

$$\hat{H}_\phi(\alpha) \geq 1$$

for any $\alpha \in \overline{\mathbf{Q}}$*, with equality if and only if* α *is* pre-periodic for ϕ.

Recall that α is a *pre-periodic point for* ϕ if the orbit $\{\alpha, \phi\alpha, \phi^2\alpha, \dots\}$ contains only finitely many points.

The natural generalization of Lehmer's conjecture to this more general setting had been raised by P. Moussa, J-S. Geronimo and D. Bessis in 1984 [10].

The rational map

$$\phi(X) = \frac{(X^2-1)^2}{4X^3+4X}$$

corresponds to the duplication map on the elliptic curve $Y^2 = X^3 + X$. For this map, and more generally for the rational maps corresponding to multiplication by an integer on an elliptic curve, partial results towards this Lehmer-type conjecture are known (M. Laurent, D.W. Masser and S.W. Zhang, M. Hindry and J. Silverman).

Variants of this construction have been proposed, mainly by J.H. Silverman. In [13], he defined heights on $K3$ surfaces using two involutions which generate an infinite group of automorphisms. With G.S. Call in [4], he did the same on general varieties V by using a morphism $\phi : V \to V$ and a divisor D for which $\phi^* D$ is linearly equivalent to αD with $\alpha > 1$. In [14], he considered a variety V related with the Hénon map

$$\phi : \mathbf{A}^2 \to \mathbf{A}^2, \qquad \phi(X,Y) = (Y, Y^2 + aX + b)$$

as follows: blowing up each of the points $(1:0:0)$ and $(0:1:0)$ three times, one obtains a variety V so that both ϕ and ϕ^{-1} extend to morphisms $V \to \mathbf{P}^2$. For $P \in \mathbf{A}^2(\overline{\mathbf{Q}})$, the relation

$$h(\phi^n P) + h(\phi^{-n} P) \geq (2^n + 2^{-n})\big(h(P) - c\big) + 2c$$

holds with some constant $c = c(\phi)$. Silverman used this inequality to prove that ϕ has only finitely many periodic points with rational coordinates and to count the growth rate of points in an infinite orbit.

The author wishes to thank Michel Planat for the excellent organization of this conference.

References

1. D'Ambros, P., Everest, G., Miles, R., Ward, T. (1999) Dynamical systems arising from elliptic curves. Manuscript, 9 pp.
2. Becker, P-G., Bergweiler, W. (1993) Transcendency of local conjugacies in complex dynamics and transcendency of their values. Manuscripta Math. **81** no. 3-4, 329–337.
3. Becker, P-G., Bergweiler, W. (1995) Hypertranscendency of conjugacies in complex dynamics. Math. Ann. **301** no. 3, 463–468.

4. Call, G. S., Silverman, J. H. (1993) Canonical heights on varieties with morphisms. Compositio Math. **89** no. 2, 163–205.
5. Everest, G., Ward, T. (1999) Heights of Polynomials and Entropy in Algebraic Dynamics. Springer, London.
6. Halmos, P. (1956) Lectures on ergodic theory. Publications of the Mathematical Society of Japan, no. 3, The Mathematical Society of Japan. Chelsea Publishing Co., New York 1960.
7. Lehmer, D.H. (1933) Factorization of certain cyclotomic functions. Ann. of Math. **34** 461–479.
8. Morton, P., Silverman, J. H. (1994) Rational periodic points of rational functions. Internat. Math. Res. Notices no. 2, 97–110.
9. Morton, P., Silverman, J. H. (1995) Periodic points, multiplicities, and dynamical units. J. Reine Angew. Math. **461**, 81–122.
10. Moussa, P., Geronimo, J. S., Bessis, D. (1984) Ensembles de Julia et propriétés de localisation des familles itérées d'entiers algébriques. C. R. Acad. Sci. Paris Sér. I Math. **299** no. 8, 281–284.
11. Nishioka, Kumiko (1996) Mahler functions and transcendence. Lecture Notes in Mathematics **1631**, Springer-Verlag, Berlin.
12. Schmidt, K. (1995) Dynamical systems of algebraic origin. Birkhäuser, Progress in Math. **128**.
13. Silverman, J. H. (1991) Rational points on $K3$ surfaces: a new canonical height. Invent. Math. **105**, no. 2, 347–373.
14. Silverman, J. H. (1993) Integer points, Diophantine approximation, and iteration of rational maps. Duke Math. J. **71** no. 3, 793–829.
15. Silverman, J. H. (1994) Geometric and arithmetic properties of the Hénon map. Math. Z. **215** no. 2, 237–250.
16. Silverman, J. H. (1995) The field of definition for dynamical systems on $\mathbf{P}^1$. Compositio Math. **98** no. 3, 269–304.
17. Silverman, J. H. (1996) Rational functions with a polynomial iterate. J. Algebra **180** no. 1, 102–110.
18. Silverman, J. H. (1996) Small Salem numbers, exceptional units, and Lehmer's conjecture. Symposium on Diophantine Problems (Boulder, CO, 1994). Rocky Mountain J. Math. **26** no. 3, 1099–1114.

Dynamics of Some Contracting Linear Functions Modulo 1

Yann Bugeaud[1⋆] and Jean-Pierre Conze[2]

[1] Université Louis Pasteur et C.N.R.S., 7 rue René Descartes, 67084 Strasbourg Cedex, France
[2] Institut Mathématique de Rennes, 263, Avenue du Général Leclerc, CS 74205, 35042 Rennes Cedex

Abstract. In various problems in signal theory, the following family of functions of $[0,1[$ into itself arises naturally :

$$T_{\gamma,\alpha} : x \mapsto \gamma x + \alpha \text{ modulo } 1, \text{ with } 0 < \gamma < 1, 0 \leq \alpha < 1.$$

The purpose of the present work is to describe very precisely the asymptotic behaviour of the iterates of the functions $T_{\gamma,\alpha}$. Further, we give a continued fraction type algorithm which allows us to obtain, for a given pair (γ, α), the numerical information on the dynamic of $T_{\gamma,\alpha}$.

1 Introduction

The problem considered here is motivated by questions arising naturally in signal theory (see for example [5]). Indeed, in order to describe certain observations, one is led to define the following sequences (s_n) and (x_n) :

$$s_n = A \sin\bigl(2\pi(nf + f_0)\bigr),$$

$$x_{n+1} = \{\gamma\, x_n + s_n\},$$

where A, γ, f_0 and f are real numbers, with f irrational and $0 < \gamma \leq 1$. Here, and in the rest of the paper, $\{x\}$ denotes the fractional part of the real number x. We would like to study the existence of the limit of the mean sequence $m_n = \frac{1}{n}\sum_{k=1}^{n} x_k$ and of the correlation sequence $c_n = \frac{1}{n}\sum_{k=1}^{n} x_k\, x_{k+1} - m_n^2$. Both questions are not yet solved and they look very difficult.

A preliminary step is the study of the dynamic associated to the recurrence $x_{n+1} = \{\gamma\, x_n + s_n\}$ when $f = 0$, *i.e.* when $s_n =: \alpha$ is a constant input. The case $\gamma = 1$ corresponds to the rotations of the circle, and hence is well known. However, for $0 < \gamma < 1$, the associated map, say $T_{\gamma,\alpha} : x \mapsto \{\gamma\, x + \alpha\}$, contracts the torus $[0,1[$, and it seems that the asymptotic behaviour of the iterates of $T_{\gamma,\alpha}$ has not been studied previously. In a joint work with J.-P. Conze [3], we completely solve this question and, for any fixed $\gamma \in]0,1[$ and any given rational $p/q \in]0,1[$, we determine precisely the set of values

⋆ bugeaud@math.u-strasbg.fr

M. Planat (Ed.): LNP 550, pp. 379–387, 2000.
© Springer-Verlag Berlin Heidelberg 2000

of the parameter α for which $T_{\gamma,\alpha}$ behaves asymptotically like the rotation $x \mapsto x + p/q$. Some of the results of [3] were announced in [2].

This expository paper is organized as follows. We first recall the basic results about the rotations of the torus, and introduce some (classical) notations used in the statement of the main theorem. Then, we state our main result, as well as a slight generalization, and we briefly discuss two different ways to prove them. Finally, we describe an algorithm which, for any given pair (γ, α) of real numbers with $0 < \gamma, \alpha \leq 1$, computes the asymptotic behaviour of the iterates of $T_{\gamma,\alpha}$, and we conclude by studying an example.

2 Notation

We denote by X the unit interval $[0, 1[$ and by $[x]$ the integer part of the real number x. For any $\alpha \in [0, 1[$, we denote by $T_\alpha : X \to X$ the rotation of the torus given by $T_\alpha x = \{x+\alpha\}$ for $x \in X$, and by $\mathcal{R}$ the set of all the rotations.

One can code the rotation T_α using the partition of the interval $[0, 1[$ in the two sub-intervals $[0, 1-\alpha[$ and $[1-\alpha, 1[$. Namely, let $\varepsilon : X \to X$ be the function defined by

$$\varepsilon(x) = 0, \text{ if } x \in [0, 1-\alpha[, \quad \text{and } \varepsilon(x) = 1, \text{ if } x \in [1-\alpha, 1[.$$

For any $x \in X$, the sequence $(\varepsilon(T_\alpha^{k-1}x))_{k\in\mathbf{Z}}$ is a sequence of "0" and "1" coding the point x. Applying the relation $x + \alpha = T_\alpha x + \varepsilon(x)$ at the point $T_\alpha^{n-1}x$, we get by induction that

$$x + n\alpha = T_\alpha^n x + \varepsilon(T_\alpha^{n-1}x) + \varepsilon(T_\alpha^{n-2}x) + \ldots + \varepsilon(x).$$

We will simply denote by $(\varepsilon_k)_{k\in\mathbf{Z}}$ the sequence corresponding to $x = 0$. Then, by definition, we have $\varepsilon_k = \varepsilon(T_\alpha^{k-1}0)$, for all $k \in \mathbf{Z}$. In order to point out that this sequence depends on α, we also denote it by (ε_k^α). We point out that for any α, we have $\varepsilon_1^\alpha = 0$ and, for $k \geq 1$, we get

$$\varepsilon_k^\alpha = [k\alpha] - [(k-1)\alpha].$$

The sequence (ε_k^α) gives the dynamic of the map T_α on the orbit of the point 0. If we denote by A_0 (*resp.* A_1) the rotation $x \mapsto x + \alpha$ (*resp.* $x \mapsto x + \alpha - 1$), we have

$$T_\alpha^n(0) = A_{\varepsilon_n^\alpha} \ldots A_{\varepsilon_1^\alpha}(0). \tag{1}$$

In the sequel, we call the sequence (ε_k^α) the coding sequence of the rotation T_α. This is also called the Sturmian sequence of α. For a rational $\alpha = p/q$, with p/q irreducible, the sequence $(\varepsilon_k^{p/q})$ is q-periodic. Moreover, the set $\{T_{p/q}^n 0\}_{n\in\mathbf{Z}}$ is finite and has exactly q distinct elements. If α is irrational, then the set $\{T_\alpha^n 0\}_{n\in\mathbf{Z}}$ is dense in $[0, 1[$, and even equidistributed. The following observation appears to be the key of the proofs of our main results.

Lemma 11. *Let p/q and p'/q' be rational numbers with $p'q - pq' = 1$ and set $r = (p+p')/(q+q')$. Let $(\varepsilon_k^{p/q}$ (resp. $(\varepsilon_k^{p'/q'})$) be the coding sequence of the rotation $T_{p/q}$ (resp. $T_{p'/q'}$). Then the coding sequence (ε_k^r) of the rotation T_r is given by*

$$\varepsilon_\ell^r = \varepsilon_\ell^{p/q} \quad \text{for all } 1 \le \ell \le \mathrm{q},$$

and

$$\varepsilon_{q+\ell}^r = \varepsilon_\ell^{p'/q'} \quad \text{for all } 1 \le \ell \le \mathrm{q}'.$$

We observe that, for $n \ge 1$, we have $[n\alpha] = \sum_1^n \varepsilon_k$ and $n\alpha = \{n\alpha\} + \sum_1^n \varepsilon_k$. This furnishes a method of binary quantification for α, since α can be written, if $\alpha = p/q \in [0,1[$ is rational :

$$\alpha = \frac{1}{q}[\varepsilon(T_\alpha^{q-1}0)+\varepsilon(T_\alpha^{q-2}0)+...+\varepsilon(0)] = \lim_n \frac{1}{n}[\varepsilon(T_\alpha^{n-1}0)+\varepsilon(T_\alpha^{n-2}0)+...+\varepsilon(0)],$$

and if α is irrational :

$$\alpha = \lim_n \frac{1}{n}[\varepsilon(T_\alpha^{n-1}0) + \varepsilon(T_\alpha^{n-2}0) + ... + \varepsilon(0)].$$

3 Results

The problem of the asymptotic behaviour of iterates of smooth perturbations of rotation maps has been extensively studied. The more famous example is given by the family of applications over X

$$A_{\alpha,\varepsilon} : \quad x \mapsto \{x + \alpha + \varepsilon \sin 2\pi x\}, \tag{2}$$

for $0 \le \alpha < 1$ and $0 \le \varepsilon < 1/(2\pi)$, which leads to the so-called "Arnold's tongues" (see [1], [4], [7]).

The question considered here is on an essentially different nature, since the perturbated maps are no more onto. Let $0 < \gamma, \alpha \le 1$ be two real numbers and set

$$T_{\gamma,\alpha} : \ x \longmapsto \{\gamma x + \alpha\},$$

where γ and α are two real parameters with $\gamma \in]0,1]$ and $\alpha \in [0,1[$. We observe that the maps $T_{\gamma,\alpha}$ are injective and that, for $\gamma = 1$, we obtain the rotation T_α. We denote by $\mathcal{C}$ the set of these maps $T_{\gamma,\alpha}$. As we are now interested in the contracting case, we assume in the sequel that $0 < \gamma < 1$.

Similarly as for diffeomorphisms of the torus (see [4], Theorem 14.6), one can define a *rotation number* for the maps $T_{\gamma,\alpha}$.

Proposition 2. *Let $f : X \to X$ be an injective map and assume that there exists a continuous, increasing map $F : [0,1[\to \mathbf{R}$ such that for all $x \in [0,1[$ we have $f(x) = \{F(x)\}$. We still denote by F the extension of F to $\mathbf{R}$ given*

by $F(x) = [x] + F(\{x\})$, for all $x \in \mathbf{R}$. Then, F is called a lift of f and the limit

$$\lim_{n\to\infty} \frac{F^n(x)}{n} =: \rho$$

exists and does not depend on $x \in \mathbf{R}$: the real $\rho \in [0,1]$ is called the rotation number of f.

The dynamic of the iterates of the map $T_{\gamma,\alpha}$ is given in terms of the parameters γ and α by the following result of [2].

Theorem 1. *Let q and p be two relatively prime integers with $1 \le p \le q$, and let define the interval $I_q^p(\gamma)$ by :*

$$I_q^p(\gamma) = \left[\frac{P_q^p(\gamma)}{1+\gamma+...+\gamma^{q-1}}, \frac{P_q^p(\gamma)+\gamma^{q-1}-\gamma^q}{1+\gamma+...+\gamma^{q-1}} \right],$$

where P_q^p is the polynomial

$$P_q^p(X) = \sum_{k=0}^{q-1} \varepsilon_{-k}^{p/q} X^k. \tag{3}$$

Then, the application $T_{\gamma,\alpha}$ has a periodic attracting orbit with the same dynamic as the rotation $T_{1,p/q}$ if, and only if, $\alpha \in I_q^p(\gamma)$. Say differently, the rotation number of $T_{\gamma,\alpha}$ equals p/q if, and only if, $\alpha \in I_q^p(\gamma)$.

We point out that unlike the case of the rotations, the applications $T_{\gamma,\alpha}$ have almost surely a rational rotation number.

Corollary 2. *Let $\gamma \in]0,1[$. The Lebesgue measure of the set of parameters $\alpha \in]0,1[$ for which the rotation number of $T_{\gamma,\alpha}$ is rational is equal to one.*

Since the proof of Corollary 1 is short and easy, we put it in here. The sum of the lengths of the intervals $I_q^p(\gamma)$ defined in Theorem 1 is given by

$$\mu(\gamma) = \sum_{q=1}^{\infty} \frac{\Phi(q)(\gamma^{q-1}-\gamma^q)}{1+\gamma+\ldots+\gamma^{q-1}},$$

where μ is the Lebesgue measure on $[0,1]$ and $\Phi(q)$, the Euler totient function, counts the number of positive integers p less than q and coprime with q. Then, Theorem 308 of [6] yields $\mu(\gamma) = 1$.

A comparison with the asymptotic behaviour of the applications $A_{\alpha,\varepsilon}$ (see (2)) may be in order here. Indeed, let $0 < \varepsilon < 1/(2\pi)$ and set

$$\mu_\varepsilon = \sum_{1\le p\le q \text{ and } (p,q)=1} \mu\big\{\alpha \in [0,1[\,|\, \rho(A_{\alpha,\varepsilon}) = p/q\big\}.$$

Then we have (see [1]) $0 < \mu_\varepsilon < 1$ and $\mu_\varepsilon \to 0$ when $\epsilon \to 0$.

One can also describe very precisely the set of values of α such that the rotation number of $T_{\gamma,\alpha}$ is irrational.

Theorem 2. *Let $\gamma \in]0,1[$ and let $\alpha \in]0,1[$ be an irrational number. Then there exists a unique real number $\tau := \tau_\gamma(\alpha)$ such that the rotation number of the function $T_{\gamma,\tau}$ is equal to α. The map $\gamma \mapsto \tau_\gamma(\alpha)$ is analytic and is given by*

$$\tau_\gamma(\alpha) = (1-\gamma)\sum_{k=0}^{\infty} \varepsilon_{-k}^{\alpha}\gamma^k.$$

Further, we have $\lim_{\gamma\to 1} \tau_\gamma(\alpha) = \alpha$.

It is easy to check that for $0 < \gamma < 1$, the rotation number of $T_{\gamma,\alpha}$ is strictly less than this of the rotation T_α.

It turns out that Theorem 1 can in a natural way be generalized to a wider class of contracting maps on the torus, namely to maps with two distinct slopes.

Let $\theta \in [0,1]$, u, v be three parameters with $0 < u < 1-\theta < v < 1$ and denote by $\gamma_0, \alpha_0, \gamma_1, \alpha_1$ the four reals defined by

$$\gamma_0 = (1-v)/\theta, \alpha_0 = v, \gamma_1 = u/(1-\theta), \alpha_1 = -u\theta/(1-\theta).$$

They satisfy

$$-1 < \alpha_1 < 0 < \alpha_0, \gamma_0, \gamma_1 < 1, \quad \alpha_0 > \alpha_1 + \gamma_1$$

and also the relation

$$\alpha_1\gamma_0 = (\alpha_0 - 1)\gamma_1.$$

We consider the applications

$$[0,\theta] \overset{A_{\gamma_0,\alpha_0}}{\longrightarrow} [v,1], \text{ where } A_{\gamma_0,\alpha_0}(x) = \gamma_0 x + \alpha_0,$$

$$[\theta,1] \overset{A_{\gamma_1,\alpha_1}}{\longrightarrow} [0,u], \text{ where } A_{\gamma_1,\alpha_1}(x) = \gamma_1 x + \alpha_1.$$

In the sequel, we will write A_0 and A_1 these affine maps (compare with (1)).

The map $T_{\gamma_0,\gamma_1,\alpha_0} : X \longrightarrow X$, which we denote by T when there is no source of confusion, is defined as the unique application which coincides with A_0 on $[0,\theta[$ and with A_1 on $[\theta,1[$.

We denote by $\mathcal{S}$ the set of these transformations $T_{\gamma_0,\gamma_1,\alpha_0}$ contracting, piecewise linear over $[0,1[$, and we observe that $\mathcal{R} \subset \mathcal{C} \subset \mathcal{S}$. The following theorem provides a generalization of Theorem 1 to the set $\mathcal{S}$.

Theorem 3. *Let γ_0 and γ_1 be such that $0 < \gamma_0, \gamma_1 \leq 1$. To any rational p/q irreducible, $1 \leq p \leq q$, there is an associated compact interval $I_q^p(\gamma_0, \gamma_1)$ such that, if $\alpha_0 \in I_q^p(\gamma_0, \gamma_1)$, then the rotation number of the map $T_{\gamma_0,\gamma_1,\alpha_0}$ is equal to p/q. Further, setting $\varepsilon_\ell = \varepsilon_\ell^{p/q}$, we have*

$$I_q^p(\gamma_0, \gamma_1) = \left[\frac{A_q^p(\gamma_0, \gamma_1)}{B_q^p(\gamma_0, \gamma_1)}, \frac{C_q^p(\gamma_0, \gamma_1)}{D_q^p(\gamma_0, \gamma_1)} \right],$$

where

$$\begin{aligned}
A_q^p(\gamma_0, \gamma_1) &= 1 + \sum_{k=1}^{q-1} \varepsilon_k \, \gamma_{\varepsilon_{q-1}} \cdots \gamma_{\varepsilon_k} \,, \\
B_q^p(\gamma_0, \gamma_1) &= 1 + \sum_{k=1}^{q-1} \gamma_{\varepsilon_{q-1}} \cdots \gamma_{\varepsilon_k} \,, \\
C_q^p(\gamma_0, \gamma_1) &= 1 + \sum_{k=2}^{q} \varepsilon_k \gamma_{\varepsilon_2} \cdots \gamma_{\varepsilon_k} - \gamma_{\varepsilon_1} \cdots \gamma_{\varepsilon_q} \,, \\
D_q^p(\gamma_0, \gamma_1) &= 1 + \sum_{k=2}^{q} \gamma_{\varepsilon_2} \cdots \gamma_{\varepsilon_k} \,.
\end{aligned} \tag{1}$$

It is now time to give some details of the proofs. For the sake of simplicity, we will focus on Theorem 1, since the proofs of Theorem 3 rest on the same ideas, but are a little more technical. The key ingredient for both methods is Lemma 1.

A first way for proving Theorem 1 is given in details in [3]. Roughly speaking, for a fixed $\gamma \in]0, 1[$ and for a polynomial P_q^p as given in (3), we compute the orbits of the point 0, hence

$$T^n_{\gamma, P_q^p(\gamma)/(1+\ldots+\gamma^{q-1})}(0) \text{ and } T^n_{\gamma, (P_q^p(\gamma)+\gamma^{q-1}-\gamma^q)/(1+\ldots+\gamma^{q-1})}(0),$$

for all integers $n \geq 0$. This method allows us to determine explicitely the points of the limit cycle of $T_{\gamma,\alpha}$, when α belongs to the interval $I_q^p(\gamma)$. Indeed, for $0 \leq u \leq \gamma^{q-1} - \gamma^q$, the asymptotic orbit of $T_{\gamma,(P_q^p(\gamma)+u)/(1+\ldots+\gamma^{q-1})}$ is composed by the points

$$\frac{R_\ell(\gamma) + u/(1-\gamma)}{1 + \ldots + \gamma^{q-1}},$$

where, for $1 \leq \ell \leq q$, the polynomial R_ℓ is the rest in the Euclidean division of $(1 + X + \ldots + X^{\ell-1}) \times P_q^p(X)$ by the polynomial $1 + X + \ldots + X^{q-1}$.

The second approach for proving Theorem 1 does not yield such an information. This is briefly described in [2], and more details are available from

the author. It depends on the observation that the asymptotic behaviour of the iterates of $T = T_{\gamma,\alpha}$ is deeply related to the position of the critical point $\theta = (1-\alpha)/\gamma$, which is the preimage of 0 by T. More precisely, one has the following result.

Lemma 12. *Let $n_0 = \inf\{n \geq 0 \mid \theta \notin T^n(X)\} + 1$. For all integers $n < n_0$, $T^n(X)$ is a disjoined union of $n+1$ intervals. If n_0 is finite, then, for all $n \geq n_0$, $T^n(X)$ is the union of exactly n_0 disjoined intervals.*

Proof : Let $n \geq 0$ be given and assume that $T^n(X)$ is the disjoined union of ℓ intervals $I_1, ..., I_\ell$. Then, if $\theta \in I_\ell$, we have $T^{n+1}(X) = \bigcup_{i=1}^{\ell-1} T(I_i) \cup J_1 \cup J_2$, where the $T(I_i)$ are intervals and where J_1 and J_2 are respectively of the shape $[0, .[$ and $[., 1[$. These $\ell + 1$ intervals are disjoined, since the map T is injective; in this case, an additional iteration yielded an additional interval. If $\theta \notin T^n(X)$, then $T^{n+1}(X)$ is the distinct union of the ℓ intervals $T(I_i)$. We conclude by noticing that $T^0(X) = X$ is composed by a unique interval. □

The idea is then to write the condition $\theta \notin T^n(X)$ in terms of inequalities involving γ and α and to conclude by an induction based on Lemma 1.

4 An Algorithm

Finally, we describe a continued fraction algorithm which computes the rotation number of any given function $T_{\gamma,\alpha}$. The existence of such an algorithm is not surprising at all, since it is well known that, for $0 < \alpha < 1$, induction of the rotation $x \mapsto x + \alpha$ on the interval $[1-\alpha, 1]$ allows us to recover the continued fraction expansion of α. Here, and in the sequel, we mean by continued fraction expansion a (signed) continued fraction expansion, *i.e.* an expansion of the shape

$$a = a_0 \pm \cfrac{1}{a_1 \pm \cfrac{1}{a_2 \pm \cfrac{1}{a_3 \pm \cfrac{1}{\cdots}}}}.$$

The sign essentially depends on the choice of the interval on which we induce and we merely describe here an algorithm leading to an expansion with the minus sign.

Unlike the class $\mathcal{C}$, the class $\mathcal{S}$ is stable by induction, and this is the main reason why we have introduced it. Let γ_0, γ_1, α_0 and α_1 be parameters satisfying (4) and (5). As above, we denote by θ the critical point of the transformation $T = T_{\gamma_0,\gamma_1,\alpha_0}$, and we will induce on the interval $[\theta, 1]$. As in the case of the rotations, one can define an algorithm which allows use to know precisely the asymptotic behaviour of the iterates of T. The critical

point θ is in $[0,1]$ if and only if $\alpha_0 + \gamma_0 > 1$, and if θ does not belong to the interval $[0,1]$, then T contracts the torus X, and the iterates of T converge to its unique fixed point. We then repeat the induction process while the induced transformation has a critical point. The parameters of the induced map of T on $[\theta,1]$ are computed in [3]; they yield the following algorithm.

`Input` : Two real numbers $0 < \alpha, \gamma \leq 1$.
`Output` : The continued fraction expansion with minus sign of the rotation number of $T_{\gamma,\alpha}$.

$\gamma_1 := \gamma$;
while $\alpha + \gamma > 1$ do
begin
  $u := 1$;
  $z := \gamma$;
  $b := 1$;
  while $u \leq 1/\alpha$ do
  begin
    $b := b + 1$;
    $u := 1 + \gamma * u$;
  end ;
  $\alpha := (\alpha * u - 1)/(\gamma + \alpha - 1)$;
  $\gamma := \gamma^b * \gamma_1$;
  $\gamma_1 := \gamma/z$;
  print (b) ;
end.

By Corollary 1, this algorithm stops for almost all choice of the parameters. If $b_1, b_2, \ldots, b_n$ denote the integers successively obtained through the procedure, then the rotation number of T is the rational ρ given by

$$\rho = \cfrac{1}{b_1 - \cfrac{1}{b_2 - \cfrac{1}{\ddots - \cfrac{1}{b_n}}}}.$$

5 An Example

We illustrate the preceding sections by the study of a numerical example. We consider the map $T_{\gamma,\alpha}$ given by $\gamma = 0.8$ and $\alpha = 0.5$, hence the map

$$T : x \mapsto \{0.8x + 0.5\}.$$

By an easy calculation, we get

$$T\big([0,1[\big) = [0; 0.3[\cup [0.5; 1[,$$

$$T^2\big([0,1[\big) = [0;0.3[\cup [0.5;0.74[\cup [0.9;1[,$$

$$T^3\big([0,1[\big) = [0;0.092[\cup [0.22;0.3[\cup [0.5;0.74[\cup [0.9;1[,$$

and

$$T^4\big([0,1[\big) = [0;0.092[\cup [0.22;0.3[\cup [0.5;0.5736[\cup [0.676;0.74[\cup [0.9;1[.$$

We observe that the critical point $\theta = (1-\alpha)/\gamma = 0.625$ does not belong to $T^4\big([0,1[\big)$, hence the limit cycle has 5 elements. Further, by Theorem 1, $P_5^2(\gamma) = 1+\gamma^2$, and

$$0.5 \in I_5^2(\gamma) = \left[\frac{1.64}{3.3616}, \frac{1.72192}{3.3616}\right],$$

hence $\rho(T) = 2/5$. As for the algorithm of Sect. 4, we check that it gives

$$\rho(T) = \frac{1}{3-\dfrac{1}{2}} = \frac{2}{5}.$$

References

1. Arnold, V. I. (1970) Chapitres supplémentaires de la théorie des équations différentielles ordinaires, Editions de Moscou.
2. Bugeaud, Y. (1993) Dynamique de certaines applications contractantes linéaires par morceaux sur $[0,1[$. C. R. Acad. Sci. Paris Sér. I **317**, 575–578.
3. Bugeaud, Y., Conze J.-P. (1999) Calcul de la dynamique de transformations linéaires contractantes mod 1 et arbre de Farey. Acta Arithmetica. To appear.
4. Devaney, R. L. (1989) An Introduction to Chaotic Dynamical Systems. Addison-Wesley : Menlo Park, second edition.
5. Feely, O., Chua L. O. (1991) The effect of integrator leak in $\Sigma - \Delta$ modulation. IEEE Transactions on Circuits and Systems **38**, 1293-1305.
6. Hardy, G. H., Wright, E. M. (1979) An introduction to the theory of numbers. Oxford : Clarendon Press, fifth edition.
7. de Melo, W., van Strien, S. (1993) One-dimensional dynamics. Springer-Verlag, 3. Folge, Band 25.

On the Modular Function and Its Importance for Arithmetic

Paula B. Cohen *

CNRS, UMR 8524, UFR de Mathématiques, Université des Sciences et Technologies de Lille, 59655, Villeneuve d'Ascq, France

Abstract. The modular function

$$j(\tau) = \exp(-2i\pi\tau) + 744 + \sum_{n=1}^{\infty} a_n \exp(2i\pi n\tau), \quad a_n \in \mathbf{Z},$$

automorphic with respect to the action of $\mathrm{SL}(2, \mathbf{Z})$ on the Poincaré upper half plane of those $\tau \in \mathbf{C}$ with positive imaginary part, is very important for the theory of elliptic curves and of modular forms. Indeed, the values of j parametrise the isomorphism classes over $\mathbf{C}$ of elliptic curves. In this lecture, we give an introduction to the modular function, and explain in particular a celebrated result of Th. Schneider (1937) which says that the j function takes an algebraic value at an algebraic point τ if and only if τ is imaginary quadratic, that is the associated class of elliptic curves has complex multiplication. We also discuss some more recent results.

1 The Modular Function and Elliptic Curves

Among the objects of central importance in number theory are the elliptic curves. The richness of their structure stems from its multiple faces: topological, algebraic and analytic. Elliptic curves are complex curves, or Riemann surfaces, of genus 1. The complex points of an elliptic curve can be represented as a torus $\mathbf{C}/\Lambda$ where Λ is a lattice in $\mathbf{C}$, that is a $\mathbf{Z}$-module of rank 2: $\Lambda = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$, such that $\mathbf{R}\omega_1 + \mathbf{R}\omega_2 = \mathbf{C}$. By definition, two complex elliptic curves corresponding to lattices Λ and Λ' are isomorphic exactly when there is a $\lambda \in \mathbf{C}, \lambda \neq 0$ with $\lambda\Lambda = \Lambda'$. Therefore, there is a representative of each isomorphism class whose lattice is of the form $\Lambda_\tau = \mathbf{Z} + \tau\mathbf{Z}$ with $\tau \in \mathcal{H}$, where $\mathcal{H}$ is the upper half plane of complex numbers with positive imaginary part. We denote by E_τ the corresponding elliptic curve $\mathbf{C}/\Lambda_\tau$. Let $\mathrm{SL}(2, \mathbf{Z})$ be the group of 2 by 2 matrices with integer coefficients and determinant 1. Then the elements $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $a, b, c, d \in \mathbf{Z}, ad - bc = 1$, act on the elements $\tau \in \mathcal{H}$ by fractional linear transformations:

$$\tau \mapsto \tau' = \frac{a\tau + b}{c\tau + d}.$$

* Paula.Cohen@univ-lille1.fr

M. Planat (Ed.): LNP 550, pp. 388–397, 2000.
© Springer-Verlag Berlin Heidelberg 2000

The lattice $\mathbf{Z}+\tau\mathbf{Z}$ is clearly isomorphic to the lattice $\mathbf{Z}+\tau'\mathbf{Z}$, the action of $\mathrm{SL}(2,\mathbf{Z})$ amounting to a base change over $\mathbf{Z}$. Let $\mathrm{PSL}(2,\mathbf{Z})$ be the quotient of $\mathrm{SL}(2,\mathbf{Z})$ by the matrices ± 1. Then this projective group has a well defined action by fractional linear transformations induced by that of $\mathrm{SL}(2,\mathbf{Z})$. The above remarks show that we have a bijection,

$$\mathrm{PSL}(2,\mathbf{Z})\backslash\mathcal{H} \simeq \{\text{Complex isomorphism classes of elliptic curves}\}. \qquad (1)$$

This bijection is fundamental to all that follows. We shall mainly be concerned with the analytic nature of elliptic curves, but let us recall briefly their status as algebro-geometric objects

One of the advantages of algebro-geometric constructions is that they allow us to work over essentially any field, something of crucial importance for arithmetic considerations. Let k be a field. Then an elliptic curve over k is by definition a non-singular projective cubic curve together with a point with coordinates in k. Up to a suitable change of coordinates, such a curve can be given in affine form by an equation of the shape

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

The point defined over k is then the projective point at infinity $(0:1:0)$.

When working over $k=\mathbf{C}$, we can further simplify the above equation to one of the form

$$y^2 = 4x^3 - g_2x - g_3, \qquad g_2^3 - 27g_3^2 \neq 0.$$

We can relate this to the original picture of an elliptic curve as a complex torus $\mathbf{C}/\Lambda$.

Proposition 1: *Let $\Lambda = Z\omega_1 + \mathbf{Z}\omega_2$ be a lattice in $\mathbf{C}$ with $\mathrm{Im}(\omega_1/\omega_2) > 0$ and define*

$$g_2 = g_2(\Lambda) = 60\sum_{(m,n)\neq(0,0)}(m\omega_1+n\omega_2)^{-4},$$
$$g_3 = g_3(\Lambda) = 140\sum_{(m,n)\neq(0,0)}(m\omega_1+n\omega_2)^{-6}.$$

Then $g_2^3 - 27g_3^2 \neq 0$ and there is an isomorphism of abelian groups between the complex points $E(\mathbf{C})$ of the curve

$$E: \qquad y^2 = 4x^3 - g_2x - g_3,$$

together with the point at infinity, and the complex torus $\mathbf{C}/\Lambda$.

In order to understand the statement of Proposition 1, we must understand the group law. On $\mathbf{C}/\Lambda$ it is just the addition of complex numbers modulo Λ induced by the usual addition of complex numbers. Hence all the numbers of the form $m\omega_1+n\omega_2$, with $m,n\in\mathbf{Z}$, are the zero element in $\mathbf{C}/\Lambda$. In the algebraic description as the complex points on the curve E, the zero element is the point at infinity $(0:1:0)$. The inverse of a point (x,y) is the

point $(x, -y)$. If $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ on E are not inverses of each other, then their sum $P_3 = (x_3, y_3)$ is given by the formulae,

$$\begin{aligned} x_3 &= -x_1 - x_2 + \tfrac{1}{4}G^2, \\ y_3 &= -y_1 - G(x_3 - x_1) \end{aligned}$$

where

$$\begin{aligned} G &= (y_1 - y_2)/(x_1 - x_2), \qquad P_1 \neq P_2, \\ &= (12x_1^2 - g_2)/(2y_1), \qquad P_1 = P_2. \end{aligned}$$

Geometrically speaking, given two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ on E, with neither being the point at infinity, there is a line $\overline{P_1P_2}$ joining them. If $P_1 = P_2$ we understand this as the tangent to P_1. If $\overline{P_1P_2}$ is a vertical line we have $P_1 + P_2 = 0$. If $\overline{P_1P_2}$ is not a vertical line, then $-P_1 - P_2$ is the third point of intersection of E with $\overline{P_1P_2}$.

The two descriptions of an elliptic curve in Proposition 1 are linked by the Weierstrass parametrisation. Given a lattice Λ in $\mathbf{C}$, we associate to Λ the Weierstrass elliptic function $\mathcal{P}(z) = \mathcal{P}(z, \Lambda)$ defined by

$$\mathcal{P}(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right). \tag{2}$$

By construction, the function $\mathcal{P}(z)$ is periodic with respect to Λ, that is $\mathcal{P}(z) = \mathcal{P}(z + \omega)$ for all $\omega \in \Lambda$, the extra terms $\frac{1}{\omega^2}$ being necessary to ensure convergence. Differentiating the above expression term by term yields

$$\mathcal{P}'(z) = -2 \sum_{\omega \in \Lambda} \frac{1}{(z-\omega)^3}. \tag{3}$$

The function $\mathcal{P}(z)$ has a double pole at each element of Λ, its derivative $\mathcal{P}'(z)$ has triple poles at these same points, and it satisfies the differential equation

$$(\mathcal{P}'(z))^2 = 4(\mathcal{P}(z))^3 - g_2\mathcal{P}(z) - g_3.$$

In fact, we have a parametrisation of E given by the Weierstrass map

$$\begin{aligned} &\mathbf{C} \to \mathbf{C}/\Lambda \simeq E(\mathbf{C}) \\ &z \mapsto (\mathcal{P}(z) : \mathcal{P}'(z) : 1), \quad z \neq 0 \\ &0 \mapsto (0 : 1 : 0). \end{aligned}$$

The algebraic addition law for the complex points of E can be translated into an addition law for the Weierstrass functions.

The modular function arises when one wishes to look at all isomorphism classes of elliptic curves at once. As we saw above, the isomorphism classes of elliptic curves are parametrised by the points of the moduli space $\mathrm{PSL}(2, \mathbf{Z})\backslash\mathcal{H}$. Therefore our modular function should be "periodic" or automorphic with respect to the action of $\mathrm{PSL}(2, \mathbf{Z})$ on $\mathcal{H}$. It turns out that all

reasonable such functions form a field generated by a single modular function which one can normalise in a suitable way so that it is uniquely determined.
Definition 1: *A function $h = h(\tau)$ on $\mathcal{H}$ is said to be modular (of weight zero) if it is meromorphic on $\mathcal{H}$ and satisfies*

$$h(\frac{a\tau + b}{c\tau + d}) = h(\tau), \quad \tau \in \mathcal{H}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbf{Z}).$$

We require in addition that h be meromorphic at infinity, namely that its Fourier series

$$h(\tau) = \sum_{n \in \mathbf{Z}} a_n q^n, \quad q = e^{2\pi i \tau},$$

have at most finitely many nonzero a_n with $n < 0$.

Such functions form a field. Moreover, as the transformation $\tau \mapsto \tau + 1$ is in $\mathrm{PSL}(2, \mathbf{Z})$, each modular function has a Fourier expansion in powers of $q = \exp(2\pi i \tau)$. The field of modular functions is generated by one element which we can choose by stipulating that $a_n = 0$ for all $n < 1$ and by requiring that $a_{-1} = 1$ and $a_0 = 744$. We recover then *the* modular function

$$j(\tau) = q^{-1} + 744 + \sum_{n=1}^{\infty} a_n q^n, \tag{4}$$

also referred to as the j-function. All of the coefficients $a_n, n \geq -1$ are positive integers, for example $a_1 = 196884$ and $a_2 = 21493760$. We can express the j-function in terms of the g_2 and g_3 of Proposition 1. From their definition we see at once that

$$g_2(\lambda \Lambda) = \lambda^{-4} g_2(\Lambda), \quad g_3(\lambda \Lambda) = \lambda^{-6} g_3(\Lambda)$$

for all $\lambda \in \mathbf{C} \setminus \{0\}$ and all lattices Λ in $\mathbf{C}$. Let $\Delta = \Delta(\Lambda) = g_2^3 - 27 g_3^2$. Then the ratio $j(\Lambda) = g_2^3(\Lambda)/\Delta(\Lambda)$ clearly satisfies $j(\lambda\Lambda) = j(\Lambda), \lambda \in \mathbf{C} \setminus \{0\}$ and so depends only on the isomorphism class of Λ.
Proposition 2: *The function*

$$j : \mathcal{H} \to \mathbf{C}$$

induces a bijection between the moduli space $\mathrm{PSL}(2, \mathbf{Z}) \backslash \mathcal{H}$ *and the complex numbers* $\mathbf{C}$. *Therefore, to each isomorphism class of elliptic curves corresponds one and only one value of the function j. Moreover, we have*

$$j(\tau) = j(\mathbf{Z} + \tau \mathbf{Z}), \quad \tau \in \mathcal{H}.$$

Proof: For a full proof of the Proposition see [4], Proposition 11, p.119. By the defining properties of modular functions, we know that the function j is well defined modulo $\mathrm{PSL}(2, \mathbf{Z})$. One can verify that $j(\tau) = j(\mathbf{Z} + \tau \mathbf{Z})$ for

$\tau \in \mathcal{H}$ by using its expression in terms of g_2 and g_3 and computing the first few coefficients of the resulting Fourier expansion of $j(\tau)$. The function Δ has a simple zero at infinity and no other zero. Since g_2 does not vanish at infinity, the j-function has a simple pole at infinity and is holomorphic on $\mathcal{H}$. By properties of modular forms, we know that for any $c \in \mathbf{C}$ the modular form $1728g_2^3 - c\Delta$ must vanish at exactly one point $P \in \mathrm{PSL}(2, \mathbf{Z})\backslash\mathcal{H}$. Dividing by Δ, we see that $j(\tau) - c = 0$ has exactly one zero for $\tau \in \mathrm{PSL}(2, \mathbf{Z})\backslash\mathcal{H}$. Thus j takes ∞ to ∞ and induces a bijection of $\mathrm{PSL}(2, \mathbf{Z})\backslash\mathcal{H}$ with $\mathbf{C}$.

2 Complex Multiplication

The group $\mathrm{PSL}(2, \mathbf{Z})$ has a presentation, which determines it up to conjugacy in $\mathrm{PSL}(2, \mathbf{R})$, in terms of generators and relations by

$$\langle M_1, M_2, M_3 \mid M_1^2 = M_2^3 = M_1 M_2 M_3 = 1 \rangle.$$

We can take as generators M_1 and M_3, the transformations given respectively by the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

We then can take $M_2 = ST^{-1}$. Notice that M_3 is the transformation $z \mapsto z+1$ and so has order infinity. The fundamental domain for the action of $\mathrm{PSL}(2, \mathbf{Z})$ on $\mathcal{H}$ is given by two adjacent copies of a hyperbolic triangle with angles $\pi/2, \pi/3$ and 0 and vertices at fixed points of elements of $\mathrm{PSL}(2, \mathbf{Z})$ of order $2, 3$ and ∞ respectively. We can take this fundamental domain to have closure the region

$$F = \{z \in \mathcal{H} \mid -1/2 \leq \mathrm{Re}(\tau) \leq 1/2,\ |\tau| \geq 1\}.$$

The matrix S has fixed point $\tau = i = \sqrt{-1}$ whereas T has fixed point at infinity and ST^{-1} has fixed point $\tau = (1 + \sqrt{-3})/2$. The points $\tau = i$ and $\tau = (1 + \sqrt{-3})/2$ correspond to elliptic curves with special properties.

Definition 2: *A non-zero complex number α is called an endomorphism of a lattice Λ in $\mathbf{C}$ if $\alpha\Lambda \subset \Lambda$. Such an endomorphism induces an endomorphism of the elliptic curve $E = \mathbf{C}/\Lambda$ by multiplication in $\mathbf{C}$ and reduction modulo Λ. These endomorphisms form a ring denoted* $\mathrm{End}(E)$. *Let* $\mathrm{End}_o(E)$ *denote the algebra of endomorphisms of E defined by* $\mathrm{End}_o(E) = \mathrm{End}(E) \otimes_{\mathbf{Z}} \mathbf{Q}$.

It is clear that any lattice Λ of $\mathbf{C}$ is invariant under multiplication by the elements of $\mathbf{Z}$. Hence we always have $\mathbf{Q} \subset \mathrm{End}_o(E)$. Consider the lattice $\Lambda_i = \mathbf{Z} + i\mathbf{Z}$ corresponding to the fixed point i of S. We clearly have $i\Lambda_i = \Lambda_i$ since $i^2 = -1$. Therefore $\mathbf{Q}(i) \subset \mathrm{End}_o(E_i)$ and as $\Lambda_i \otimes_{\mathbf{Z}} \mathbf{Q} = \mathbf{Q}(i)$ we in fact have $\mathbf{Q}(i) = \mathrm{End}_o(E_i)$. We can argue in a similar way for the point $\tau = (-1 + \sqrt{-3})/2$.

Definition 3: *An elliptic curve E is said to have complex multiplication if its endomorphism ring* $\mathrm{End}(E)$ *strictly contains* $\mathbf{Z}$.

We have the following observation.

Proposition 3: *Let E be an elliptic curve and assume that E has complex multiplication. Then* $\mathrm{End}(E)$ *is isomorphic as a ring to* $\mathbf{Z}+\tau\mathbf{Z}$ *for some* $\tau \in \mathcal{H}$ *which is an imaginary quadratic integer, that is it satisfies an equation of the form*

$$\tau^2 - T\tau + N = 0$$

where T and N are integers with $D = T^2 - 4N < 0$. Conversely, if $\tau \in \mathcal{H}$ is quadratic imaginary, the elliptic curve E_τ has complex multiplication and $\mathrm{End}_o(E) = \mathbf{Q}(\tau)$.

Proof: We write $E = \mathbf{C}/\Lambda$ for a lattice Λ in $\mathbf{C}$. Then $\mathrm{End}(E)$ is given by the $\alpha \in \mathbf{C} \setminus \{0\}$ with $\alpha\Lambda \subset \Lambda$. As we are just interested in $\mathrm{End}(E)$, we can suppose $\Lambda = \Lambda_\tau = \mathbf{Z} + \tau\mathbf{Z}$ with $\tau \in \mathcal{H}$. As α leaves Λ_τ fixed, there must be integers a,b,c,d with

$$\alpha = a + b\tau, \qquad \alpha\tau = c + d\tau.$$

Therefore α is an eigenvalue of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and hence is an algebraic integer of degree 2 and satisfies an equation as in the statement of the Proposition with $T = a + d$ and $N = ad - bc$. Since $\alpha = a + b\tau$ we have $\mathbf{Q}(\alpha) = \mathbf{Q}(\tau)$ which is an imaginary quadratic extension of $\mathbf{Q}$ containing $\mathrm{End}(E)$ as a subring of its ring of integers.

For example, thinking algebraically, the curves $y^2 = x^3 - ax$ all have complex multiplication by the Gaussian integers $\mathbf{Z}[i]$. Indeed, the action of i itself is given by the map of (x, y) to $(-x, iy)$. Similarly, the curves $y^2 = x^3 + b$ all have complex multiplication by $\mathbf{Z}[(-1 + \sqrt{-3})/2]$ with map of (x, y) to $(((-1 + \sqrt{-3})/2)x, y)$. For another example, one can check that the curve

$$y^2 = x^3 - (3/4)x^2 - 2x - 1$$

has complex multiplication by $\mathbf{Z}[\kappa]$ where $\kappa = (1 + \sqrt{-7})/2$, multiplication by κ sending (x, y) to (u, v), where

$$u = \kappa^{-2}(x + a + \tfrac{b}{x-a})$$
$$v = \kappa^{-3}y(1 - \tfrac{b}{(x-a)^2}),$$

with $a = (\kappa - 3)/4$ and $b = -\frac{7}{16}(3\kappa - 1)$. These computations are due to D. Bernardi (see [1], p.224).

A far more difficult result than Proposition 3 concerns the values of the j-function at complex multiplication moduli τ (that is moduli for which E_τ has complex multiplication).

Theorem 1: *Let $\tau \in \mathcal{H}$ be quadratic imaginary. Then $j(\tau)$ is an algebraic integer.*

Some examples, again taken from the reference [1], p225, of such special values of the j-function are,

$j((1+i\sqrt{3})/2) = 0$
$j(i) = 1728 = 12^3$
$j((1+i\sqrt{7})/2) = -3375 = (-15)^3$
$j(i\sqrt{2}) = 8000 = 20^3$
$j((1+i\sqrt{11})/2) = -32768 = (-32)^3$
$j((1+i\sqrt{19})/2) = -884736 = (-96)^3$
$j((1+i\sqrt{43})/2) = -884736000 = (-960)^3$
$j((1+i\sqrt{67})/2) = -147197952000 = (-5280)^3$
$j((1+i\sqrt{163})/2) = -262537412640768000 = (-640320)^3$
$j(i\sqrt{3}) = 54000 = 2(30)^3$
$j(2i) = 287496 = (66)^3$
$j((1+3i\sqrt{3})/2) = -12288000 = -3(160)^3$
$j(i\sqrt{7}) = 16581375 = (255)^3$
$j((1+i\sqrt{15})/2) = (-191025+85995\sqrt{5})/2(1+\sqrt{5})/2((75-27\sqrt{5})/2)^3$
$j((1+i\sqrt{23})/2) = \theta,$

where θ is the largest negative root of the cubic equation

$$X^3 + 3491750X^2 - 5151296975X + 12771880859375 = 0.$$

As also noted in [1], we know that $j(\tau)$ is rather well approximated for $\mathrm{Im}(\tau)$ large by $1/q+744$, just by looking at its Fourier expansion. Hence, for example $e^{\pi\sqrt{163}}$ should be close to an integer and $(e^{\pi\sqrt{163}} - 744)^{1/3}$ should be even closer. In fact, one finds

$$e^{\pi\sqrt{163}} = 262537412640768743.99999999999925007259...,$$

and

$$(e^{\pi\sqrt{163}} - 744)^{1/3} = 640319.99999999999999999999999939031735....$$

By transcendence theory, the above quantities cannot be integers and are even transcendental numbers.

3 Schneider's Theorem

There is a converse to Theorem 1 proven in 1937 by Th. Schneider [5]. It cannot expect to say anything about the converse to being an "algebraic integer", but it says something about transcendence, that is the converse to being "algebraic".

Theorem 2: *Let τ be an algebraic point of $\mathcal{H}$, that is $\tau \in \mathcal{H} \cap \overline{\mathbf{Q}}$, then if τ is not quadratic imaginary then $j(\tau)$ is transcendental.*

We can put this together with Theorem 1.

Corollary *The numbers $\tau \in \mathcal{H}$ and $j(\tau) \in \mathbf{C}$ are both algebraic if and only if τ is imaginary quadratic.*

This Corollary is really a remarkable statement. It says that for an algebraic number $\tau \in \mathcal{H}$, the algebraicity of the value of a transcendental function at τ will determine whether or not the degree of τ is 2! Generally speaking, we expect that suitably normalised transcendental functions will take transcendental values at algebraic points. In our case the suitable normalisation corresponds to the fact that the Fourier coefficients a_n of $j(\tau)$ are algebraic. When algebraic values are taken by a transcendental function at algebraic points, it means something very special happens. Indeed, a more modern name for complex multiplication points in $\mathcal{H}$ is "special points". The special points in the elliptic situation are being given by fixed points of elements of $\mathrm{PSL}(2, \mathbf{Q})$ and the value of the j-function at a special point τ generates a particular type of class field, namely the Hilbert class field H of $K = \mathbf{Q}(\tau)$. The extension H/K is an abelian Galois extension with abelian Galois group isomorphic to the ideal class group of K. The generalisation of such constructions of class fields, arising from moduli spaces of abelian varieties, has been done by Shimura. The generalisation of Schneider's theorem to this higher dimensional situation is due to myself, Shiga and Wolfart [2].

The proof of Schneider's theorem uses heavily, through the use of Weierstrass functions, the fact that the orbits of the action of $\mathrm{PSL}(2, \mathbf{Z})$ correspond to isomorphism classes of elliptic curves. An important open problem originally posed by Th. Schneider is the following.

Open Problem: *Prove Theorem 2 using only the intrinsic properties of the modular function $j(\tau)$, in particular by bypassing the use of elliptic curves.*

It would go beyond the scope of this lecture to enter into the proof of Theorem 2 using Weierstrass functions. It relies on the fact that if τ is not imaginary quadratic then the two functions $\mathcal{P}(z)$ and $\mathcal{P}(\tau z)$ are algebraically independent. This allows to make an auxiliary construction using a polynomial $P(X, Y)$ with $X = \mathcal{P}(z)$ and $Y = \mathcal{P}(\tau z)$. One shows that if $\Lambda = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ is a lattice in $\mathbf{C}$ with generators ω_1 and ω_2, where $\tau = \omega_1/\omega_2$, and with "invariants" $g_2(\Lambda)$ and $g_3(\Lambda)$ algebraic numbers, then the $\overline{\mathbf{Q}}$-vector space $\Lambda \otimes_{\mathbf{Z}} \overline{\mathbf{Q}}$ has $\overline{\mathbf{Q}}$-dimension equal to $2/[\mathrm{End}_o(E) : \mathbf{Q}]$. If g_2 and g_3 are in $\overline{\mathbf{Q}}$, then the algebraic equation defining E has coefficients in $\overline{\mathbf{Q}}$, and E is said to be defined over $\overline{\mathbf{Q}}$. We have $j(\tau) \in \overline{\mathbf{Q}}$ by the expression of j in terms of the invariants. Conversely if $j(\tau) \in \overline{\mathbf{Q}}$, there is a lattice Λ isomorphic to Λ_τ with algebraic invariants. Hence if τ and $j(\tau)$ are in $\overline{\mathbf{Q}}$, then $2/[\mathrm{End}_o(E) : \mathbf{Q}] = 1$ so that $\mathrm{End}_o(E)$ has to be an extension of $\mathbf{Q}$ of degree 2 and this means it is quadratic imaginary and so E has complex multiplication.

4 Generalisations

It is natural to ask about the transcendence properties of the j-function as a function of $q = \exp(2\pi i\tau)$: we shall write $J = J(q) = j(\tau)$. This has turned

out to be a motor for very important recent developments in the theory of transcendence and algebraic independence of quantities derived from modular functions. In 1995, K. Barré-Sirieix, G. Diaz, F. Gramain, and G. Philibert showed that $J(q)$ is transcendental for all $q \in \overline{\mathbf{Q}}$ with $0 < |q| < 1$, so solving a problem posed by Mahler in 1969. There is also a p-adic version of this problem solved by the same four authors. In 1996, Y. Nesterenko developed techniques inspired by this progress to deduce a powerful result about the algebraic independence of the values of certain Eisenstein series. One can deduce many results from this one, for example the algebraic independance of the three numbers $\pi, e^{\pi}, \Gamma(1/4)$. The algebraic independence of $\pi, \Gamma(1/4)$ was known by a result of G.V. Chudnovsky going back to 1976, but the algebraic independence of π, e^{π} was unknown before the work of Nesterenko.

We have already mentioned results in higher dimension. A related question is the study of other fuchsian triangle groups besides $\mathrm{PSL}(2, \mathbf{Z})$, in the spirit of work of the author and J. Wolfart [3]. A fuchsian triangle group Δ is a subgroup of $\mathrm{PSL}(2, \mathbf{R})$ determined up to conjugacy by the presentation, for integers $2 \leq p, q, r \leq \infty$ with $1/p + 1/q + 1/r < 1$,

$$\Delta = \Delta(p, q, r) = \langle M_1, M_2, M_3 \mid M_1^p = M_2^q = M_3^r = M_1 M_2 M_3 = 1 \rangle.$$

We say that Δ is of signature (p, q, r). The case $(p, q, r) = (2, 3, \infty)$ corresponds to $\mathrm{PSL}(2, \mathbf{Z})$. The action of $\Delta(p, q, r)$ on $\mathcal{H}$ is properly discontinuous and so we have a corresponding fundamental domain which is two copies of a hyperbolic triangle with angles $\pi/p, \pi/q, \pi/r$. The fuchsian triangle groups fall into 2 classes, arithmetic and non-arithmetic. In [6], [7] Takeuchi found those signatures (p, q, r) with $\Delta(p, q, r)$ arithmetic. They are a finite list of 85 arithmetic signatures and of course the signature $(2, 3, \infty)$ is among them. The infinitely many other signatures give rise to non-arithmetic groups, for example the group with signature $(2, 5, \infty)$. The arithmetic groups are those for which the orbits of the action on $\mathcal{H}$ can be made to correspond to certain isomorphism classes of abelian varieties (higher dimensional analogues of elliptic curves) with some given structure. In the non-arithmetic case, one can also find some underlying abelian varieties, but the situation is more involved (see [3]). There is a notion of complex multiplication for abelian varieties and one can therefore define "special points" on $\mathcal{H}$ corresponding to these abelian varieties: they are not in general the same as the special points for $\mathrm{PSL}(2, \mathbf{Z})$. There is a generalisation of the j-function, which we denote by j_Δ, to a general fuchsian triangle group Δ and we have an analogue of Schneider's theorem for this situation as a consequence of the methods of [2] and [3]. A special case of a conjecture of André-Oort applied to this situation leads to the following question.

Open Problem: *The set*

$$\mathcal{S} = \{j_\Delta(\tau) \mid \tau, j_\Delta(\tau) \in \overline{\mathbf{Q}}\}$$

is finite if and only if Δ is non-arithmetic.

If Δ is arithmetic then it is clear that $\mathcal{S}$ is infinite once one understands the role of $(\mathcal{H}, \Delta)$ as defining a moduli space for abelian varieties and indeed this case is not difficult. The fact that the situation is more involved in the non-arithmetic case, as mentioned above, leads to the finiteness prediction in this case. This finiteness statement will be the hard part to prove.

References

1. H. Cohen, *Elliptic Curves*, in From Number Theory to Physics, M. Waldschmidt, P. Moussa, J.-M. Luck, C. Itzykson, eds., Springer-Verlag, New York Berlin Heidelberg Tokyo Hong Kong Barcelona Budapest (1989).
2. P. B. Cohen, Humbert surfaces and transcendence properties of automorphic functions, Rocky Mountain J. Math. **26**, No.3 (1996), pp. 987-1001.
3. P. Cohen, J. Wolfart, Modular embeddings for some non-arithmetic Fuchsian groups, Acta Arith. **LVI**, (1990), pp. 93-110.
4. N. Koblitz, Introduction to Elliptic Curves and Modular Forms, Springer-Verlag, New York Berlin Heidelberg Tokyo (1984).
5. Th. Schneider, Arithmetische Untersuchungen elliptischer Integrale, Math. Ann **113**, (1937), pp. 1-13.
6. K. Takeuchi, A characterisation of arithmetic Fuchsian groups, J. Math.Soc. Japan **27**, No.4 (1975), pp. 600-612.
7. K. Takeuchi, Arithmetic triangle groups, J. Math. Soc. Japan **29**, (1977), pp. 91-106.

On Generalized Markoff Equations and Their Interpretation

Serge Perrine ⋆

5, Rue du Bon Pasteur, 57070 Metz (France)

Abstract. This article gives a generalization of the Markoff theory for diophantine equations:

$$m^2 + \varepsilon_2 m_1^2 + \varepsilon_1 m_2^2 = (a+1)mm_1m_2 + \varepsilon_2 \partial K m_1 m_2 - um$$

It is shown how these equations are linked to finite sequences of integers. The arborescent structure of their solutions is given. The link is given with the analysis of the Markoff spectrum, and with the representation of the free group F_2 in $M(2, Z)$, the algebra of 2×2 matrices with integers coefficients.

1 Introduction

The classical theory of Andrei A. Markoff [19] [3] [10] deals with the diophantine equation, with solutions $(m, m_1, m_2) \in (N^*)^3$:

$$M^{++}(2,0,0) : x^2 + y^2 + z^2 = 3xyz$$

- In two former articles [24] [26], it was built an analog theory with similar equations, as for example:

$$M^{++}(2,0,-2) : x^2 + y^2 + z^2 = 3xyz + 2x$$

For such equations, we present here an interpretation with matrices, generalizing the work made by Harvey Cohn [5] for the classical Markoff theory. The formalism that we build can be extended to more general equations $(M^{s_1 s_2}(a, \partial K, u))$ written, with s_1 the sign of $\varepsilon_1 \in \{\pm 1\}$, s_2 the sign of $\varepsilon_2 \in \{\pm 1\}$, $a \in N$, $\partial K \in Z$ and $u \in Z$:

$$m^2 + \varepsilon_2 m_1^2 + \varepsilon_1 m_2^2 = (a+1)mm_1m_2 + \varepsilon_2 \partial K m_1 m_2 - um \qquad (M)$$

We show how to build, for such equations, trees of solutions. We establish a link with the study of diophantine approximation for real algebraic numbers of degree 2, and the representation of free groups with two generators.

In the present article, we use sequences of positive integers:

$$W = (\alpha_i, \alpha_{i-1}, ..., \alpha_0)$$

⋆ Serge.Perrine@wanadoo.fr

M. Planat (Ed.): LNP 550, pp. 398–417, 2000.
© Springer-Verlag Berlin Heidelberg 2000

$W^* = (\alpha_0, \alpha_1, ..., \alpha_i)$ the mirror sequence of W

We associate two extended sequences, on left and right:

$$\triangleleft W = \left\{ \begin{array}{ll} (1, \alpha_i - 1, \alpha_{i-1}, ..., \alpha_0) & \text{if } \alpha_i \neq 1 \\ (\alpha_{i-1} + 1, ..., \alpha_0) & \text{if } \alpha_i = 1 \end{array} \right\}$$

$$W \triangleright = (\triangleleft W^*)^*$$

We define the matrix associated to W and its determinant:

$$M_{[W]} = \begin{bmatrix} \alpha_i & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{i-1} & 1 \\ 1 & 0 \end{bmatrix} ... \begin{bmatrix} \alpha_0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\varepsilon_{[W]} = \det(M_{[W]}) = (-1)^{i+1}$$

The tranposition of matrices is associated to the mirror operation for sequences. We use the same notation for it. Moreover, we make the following convention:

$$M_{[\triangleleft]} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} = M^*_{[\triangleright]}$$

The matrices that we consider are situated in the set $M(2, Z)$ of 2×2 matrices with integer coefficients. They operate on $R \cup \{\infty\}$ thanks to the relation:

$$\begin{bmatrix} a & b \\ d & c \end{bmatrix} (x) = \frac{ax + b}{cx + d}$$

So we have relations between matrices and continued fractions:

$$M_{[W]}(\infty) = [W] = [\alpha_i, \alpha_{i-1}, ..., \alpha_0]$$

$$M_{[W]}(0) = [\alpha_i, \alpha_{i-1}, ..., \alpha_1]$$

where we use the notation:

$$[W] = [\alpha_i, \alpha_{i-1}, ..., \alpha_0] = \alpha_i + \cfrac{1}{\alpha_{i-1} + \cfrac{1}{... + \cfrac{1}{\alpha_0}}}$$

Hence, we have:

$$M_{[W_1]}([W_2]) = [W_1, W_2]$$

$$M_{[\emptyset]} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1$$

$$[\emptyset] = \infty$$

It is necessary to avoid confusion with the commutator of matrices:

$$[A, B] = A^{-1} B^{-1} A B$$

2 The Case of the Classical Markoff Theory

Here we are concerned with the triples of solutions $(m, m_1, m_2) \in (N^*)^3$ of the equation:

$$M^{++}(2,0,0) : x^2 + y^2 + z^2 = 3xyz$$

For a permutation, we only consider Cohn's triples of solutions, defined by the condition:

$$m \geq m_1 \geq m_2 \geq 1$$

H. Cohn has shown [5] that any of them is associated to a triple of matrices (M, M_1, M_2) of $SL(2, Z)$, such that:

$$\mathrm{tr}(M) = 3m$$
$$\mathrm{tr}(M_1) = 3m_1$$
$$\mathrm{tr}(M_2) = 3m_2$$

The equation $M^{++}(2,0,0)$ is then the translation of the relation (FR_1) of Robert Fricke, demonstrated for to matrices A and B contained in $SL(2, Z)$:

$$\mathrm{tr}(AB)^2 + \mathrm{tr}(B)^2 + \mathrm{tr}(A)^2$$
$$= \mathrm{tr}(AB)\mathrm{tr}(B)\mathrm{tr}(A) + \mathrm{tr}(ABA^{-1}B^{-1}) + 2$$

Here, we set $AB = M, B = M_1$, and $A = M_2$. Simplifying, the relation $M^{++}(2,0,0)$ is equivalent to the equality:

$$tr(ABA^{-1}B^{-1}) = -2$$

The construction of M, M_1 et M_2 can be done as now described. Every Cohn triple (m, m_1, m_2) gives birth to an algebraic number of degree 2, whose development as a continued fraction is:

$$\theta_2(S) = [0, \underline{S^*, 2}]$$

The sequence S^* can be decomposed as:

$$S^* = (a_n, a_{n-1}, ..., a_0) = (X_1, 2, X_2)$$

So, we can consider the following matrices:

$$M_{[S^*]} = \begin{bmatrix} m & m - K_2 \\ K_1 & K_1 - l \end{bmatrix} \text{ (here } K_1 = K_2 \text{)}$$
$$M_{[X_1]} = \begin{bmatrix} m_1 & m_1 - k_{12} \\ k_1 & k_1 - l_1 \end{bmatrix}$$
$$M_{[X_2]} = \begin{bmatrix} m_2 & m_2 - k_2 \\ k_{21} & k_{21} - l_2 \end{bmatrix}$$
$$M = AB = M_{[S^*,2]} = M_{[\triangleleft S \triangleright, 2]}$$
$$M_1 = B = M_{[X_1^* \triangleright, 2]}$$
$$M_2 = A = M_{[\triangleleft X_2^*, 2]}$$

With the notation:

$$L = M_{[2]}M^{-1}_{[\triangleright,2,\triangleleft]} = \begin{bmatrix} -1 & -6 \\ 0 & -1 \end{bmatrix} = M_{[\triangleright]}L^{-1}M_{[\triangleright]}$$

we can write:

$$AB = M_{[\triangleleft X_2,2,T]}LM_{[\triangleright,2,\triangleleft,X_2,2]}$$
$$BA = M_{[\triangleleft X_2,2,T]}M_{[\triangleright,2,\triangleleft,X_2,2]}$$
$$\mathrm{tr}(ABA^{-1}B^{-1}) = \mathrm{tr}(M_{[\triangleleft X_2,2,T]}LM^{-1}_{[\triangleleft X_2,2,T]}) = \mathrm{tr}(L) = -2$$

The link with the formalism of H. Cohn [5] is now completely clear.

The former development gives some hyperbolic quadrilateral in the Poincare's half-plane H_2, as $p\,\beta\,s\,\alpha$, whose vertices are rational numbers situated on the edge of H_2:

$$\begin{array}{ccc} \alpha = [T] & \overset{A}{\longrightarrow} & s = [X_1^{*}\triangleright] = [X_1^{*}] \\ B\uparrow & & \uparrow B \\ p = [-1,1,1,X_2^{*}] & \overset{A}{\longrightarrow} & \beta = 0 \end{array}$$

It is easy to deduce from the above diagram a conformal torus perforated at a point, associated to the triple (m, m_1, m_2). In order to do it, we only identify, in the half-plane H_2, two geodesics $p\alpha$ and βs thanks to A, and the two other geodesics $p\beta$ and αs thanks to B.

So, we get the geometrical interpretation given by H. Cohn for the classical Markoff theory, classifying the hyperbolic quadrilaterals which are fundamental domains of H_2, thanks to conformal transformations of H_2.

The things happen as if, by an hyperbolical Tangram game, we were building quadrilaterals giving the same conformal perforated torus.For this, we use the inner automorphisms of $GL(2,Z)$:

$$\mathrm{Ad}(P) : x \longrightarrow Ad(P)(x) = PxP^{-1}$$

The operator Ad is the adjunct representation of the group $GL(2,Z)$ inside the algebra $M(2,Z)$ of 2×2 matrices with integer coefficients. And the set of these transformations $\mathrm{Ad}(P)$ such that $P \in GL(2,Z)$ is a group that we call here $\mathrm{Ad}(GL(2,Z))$.

It can be shown that any couple (A,B) built as before is a couple of generators of $[SL(2,Z), SL(2,Z)]$, the normal sugroup of commutators of $SL(2,Z)$.

The couples (A,B) and (A',B') of $[SL(2,Z), SL(2,Z)]$ can be linked by a matrix P of $GL(2,Z)$:

$$(A',B') = (Ad(P)(A), Ad(P)(B))$$

We deduce [22] a conformal equivalence (if P is in $SL(2,Z)$) or an anti-conformal one in H_2 defined by P. This allows to identify the quadrilaterals built with any of the two couples.

When it is not the case, we can cut the associated quadrilaterals associated to any of the couples (A,B) and (A',B') in a finite number of domains, respectively conformly or anti-conformly equivalent. It is the phenomenon of hyperbolic Tangram quoted before.

We also find this way a direct verification that the group $G = gp(A,B)$ generated by A and B in $SL(2,Z)$ is a Fricke group [7] [15], that is to say a group G, which is fuchsian and free with two generators A and B, such that the quotient H_2/G is a perforated torus. Hence we get the other notation ([17] p.7):

$$G = \langle A, B \rangle$$

In our present case, the group G is nothing but $[SL(2,Z), SL(2,Z)]$, a free group with two generators. It is a normal subgroup of $SL(2,Z)$, even normal in $GL(2,Z)$, and such that:

$$SL(2,Z)/[SL(2,Z), SL(2,Z)] \text{ cyclic group with 12 elements}$$

The classical Markoff theory gives then couples of generators for the group G that we consider. For example, with the triple $(5,2,1)$ corresponding to the sequences $X_2 = \emptyset$ and $X_1 = (1,1)$, we get:

$$G = [SL(2,Z), SL(2,Z)] = \langle \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \rangle$$

But any other Cohn's triple for the equation $M^{++}(2,0,0)$ gives another system of generators for this group, hence isomorphic to the free abstract group $F_2 = \langle x, y \rangle$ with two generators x and y.

The complete tree of solutions of the equation $M^{++}(2,0,0)$ can be identified (see [22]) with the quotient group:

$$\mathrm{Aut}(F_2)/Ad(GL(2,Z)) = Z/2Z * Z/2Z * Z/2Z$$

So we classify the couples of generators (A,B) of F_2 by inner automorphisms of $GL(2,Z)$. In fact, the group $\mathrm{Ad}(GL(2,Z))$ is normal in $\mathrm{Aut}(F_2)$, and choosing an automorphism of F_2 is the same thing as choosing a couple of generators of this last group.

The group $\mathrm{Ad}(GL(2,Z))$ has to be distinguished from the group $Int(F_2)$, defined thanks to the inner automorphisms of the group F_2. This last group gives the possibility to define [13]:

$$\mathrm{Out}(F_2) = \mathrm{Aut}(F_2)/\mathrm{Int}(F_2) = SL(2,Z)$$

We easily see that any inner automorphism of F_2 can be extended as an inner automorphism of $GL(2,Z)$ (see [12] [14]).

Moreover, it is known [6] [31] [29] that the simple closed geodesics of the conformal perforated torus associated to F_2 give the Cohn's triples of solutions for the equation $M^{++}(2,0,0)$. So we get the possibility to understand the apparition in Physics of the classical Markoff theory, with periodical trajectories of a dynamical system which can be built on a torus perforated by the extraction of a point [1] [27].

3 More General Diophantine Equations

Our aim is to get more general dynamical systems than before. So we consider here a sequence with positive integers:

$$S^* = (a_n, a_{n-1}, ..., a_0) = (X_1, a, X_2)$$

We suppose in what follows that the sequence X_1 is defined with a new sequence T:

$$X_1 = (\triangleleft X_2^*, a, T)$$

The sequence S^* allows to consider the three matrices $M_{[S^*]}$, $M_{[X_1]}$, $M_{[X_2]}$. We suppose that they have the same expressions as in the previous section. We also introduce:

$$M_{[T]} = \begin{bmatrix} \mu & \mu - \kappa_2 \\ \kappa_1 & \kappa_1 - \lambda \end{bmatrix}$$

It is easy to deduce relations between the parameters of the matrices that we consider. For example:

$$K_1 - K_2 = \varepsilon_{[X_2]}(\kappa_1 + \kappa_2 - \mu) = \varepsilon_{[X_2]}\partial K$$

where ∂K is equal to zero if and only if, like in the Markoff theory, we have:

$$T = T^*$$

We write now:

$$A = M_{[\triangleleft X_2^*, a]}$$
$$B = M_{[X_1^* \triangleright, a]}$$
$$AB = M_{[\triangleleft X_2^*, a]} M_{[X_1^* \triangleright, a]} = M_{[\triangleleft S \triangleright, a]}$$

These relations introduce new parameters:

$$\begin{aligned}
&\varepsilon_A = \det(A) = \varepsilon_{[X_2]} \text{ (also denoted } \varepsilon_2)\\
&\varepsilon_B = \det(B) = \varepsilon_{[X_1]} \text{ (also denoted } \varepsilon_1)\\
&\mathrm{tr}(AB) = (a+1)m + (K_2 - K_1) \in Z\\
&\mathrm{tr}(A) = (a+1)m_2 + T_2 \in Z\\
&\mathrm{tr}(B) = (a+1)m_1 - T_1 \in Z\\
&T_2 = k_2 + k_{21} - m_2\\
&T_1 = k_1 + k_{12} - m_1\\
&u = m_2T_1 - m_1T_2
\end{aligned}$$

The Fricke's formula, generalizing (FR_1) to $GL(2, Z)$ is now:

$$\begin{aligned}
\mathrm{tr}(ABA^{-1}B^{-1}) = &-2 + \varepsilon_A\mathrm{tr}(A)^2 + \varepsilon_B\mathrm{tr}(B)^2 + \varepsilon_A\varepsilon_B\mathrm{tr}(AB)^2\\
&-\varepsilon_A\varepsilon_B\mathrm{tr}(A)\mathrm{tr}(B)\mathrm{tr}(AB)
\end{aligned}$$

It gives a quite complicated formula:

$$\begin{aligned}
\mathrm{tr}(ABA^{-1}B^{-1}) + 2 = &\varepsilon_2[(a+1)^2m_2^2 + 2(a+1)m_2T_2 + T_2^2]\\
&+\varepsilon_1[(a+1)^2m_1^2 - 2(a+1)m_1T_1 + T_1^2]\\
&+\varepsilon_1\varepsilon_2[(a+1)^2m^2]\\
&-\varepsilon_1\varepsilon_2(a+1)^3mm_1m_2\\
&-\varepsilon_1\varepsilon_2(a+1)^2m[m_1T_2 - m_2T_1]\\
&+\varepsilon_1\varepsilon_2(a+1)mT_1T_2 + \Delta(ABA^{-1}B^{-1})
\end{aligned}$$

where:

$$\begin{aligned}
\Delta(ABA^{-1}B^{-1}) = &\varepsilon_1\varepsilon_2[2(a+1)(K_2 - K_1) + (K_2 - K_1)^2]\\
&-\varepsilon_1\varepsilon_2(a+1)^2m_1m_2(K_2 - K_1)\\
&-\varepsilon_1\varepsilon_2(a+1)[m_1T_2 - m_2T_1](K_2 - K_1)\\
&+\varepsilon_1\varepsilon_2T_1T_2(K_2 - K_1)
\end{aligned}$$

We now introduce $M_{[U]}$ and $M_{[V]}$ such that:

$$\begin{aligned}
&M_{[X_2]} = M_{[V]}M_{[\triangleleft X_2^*]}\\
&M_{[X_1]} = M_{[\triangleleft X_2^*, a, T]} = M_{[X_1^*\triangleright]}M_{[U]} = M_{[T^*, a, X_2]}M_{[U]}\\
&M_{[V]} = M_{[X_2]}M^{-1}_{[\triangleleft X_2^*]} = M_{[\triangleleft]} + \varepsilon_2T_2M_{[X_2]}M_{[-1]}\\
&M_{[U]} = M^{-1}_{[X_1^*\triangleright]}M_{[X_1]} = M_{[\triangleright]} - \varepsilon_1T_1M_{[-1]}M_{[X_1]}
\end{aligned}$$

Hence we have:

$$\begin{aligned}
&BA = M_{[X_1^*\triangleright]}M_{[a]}M_{[\triangleleft X_2^*, a]}\\
&AB = M_{[X_1^*\triangleright]}L^{-1}M_{[a]}M_{[\triangleleft X_2^*, a]}
\end{aligned}$$

where we can introduce the matrix depending on T:

$$L^{-1} = M_{[U]} M_{[T]}^{-1} M_{[T^*]} M_{[a]} M_{[V]} M_{[a]}^{-1}$$

The former expressions give:

$$[A^{-1}, B^{-1}] = ABA^{-1}B^{-1} = M_{[X_1^{*\triangleright}]} L^{-1} M_{[X_1^{*\triangleright}]}^{-1}$$

$$\mathrm{tr}(ABA^{-1}B^{-1}) = \mathrm{tr}(L^{-1})$$

In order to compute $\mathrm{tr}(L^{-1})$, we introduce:

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

$$N = M_{[T]}^{-1} M_{[T^*]} = 1 + \varepsilon_{[T]} \varepsilon_2 (K_1 - K_2) J M_{[T^*]}$$

The product giving L^{-1} can be expressed:

$$L^{-1} = L_0^{-1} + \varepsilon_1 T_1 L_1^{-1} + \varepsilon_2 T_2 L_2^{-1} + \varepsilon_1 \varepsilon_2 T_1 T_2 L_{12}^{-1}$$

with:

$$\begin{aligned}
&L_0^{-1} = M_{[\triangleright]} N M_{[a\triangleleft]} M_{[a]}^{-1} \\
&\mathrm{tr}(L_0^{-1}) = -2 + 2\varepsilon_1 (K_1 - K_2)(a+1)\mu + \varepsilon_1 \varepsilon_2 (K_1 - K_2)^2 \\
&L_1^{-1} = -M_{[-1]} M_{[X_1]} N M_{[a]} M_{[\triangleleft]} M_{[a]}^{-1} \\
&\mathrm{tr}(L_1^{-1}) = (T_1 - 2(a+1)m_1) + \varepsilon_2 (K_1 - K_2)((a+1)m_2 - T_2) \\
&L_2^{-1} = M_{[\triangleright]} N M_{[a]} M_{[X_2]} M_{[-1]} M_{[a]}^{-1} \\
&\mathrm{tr}(L_2^{-1}) = (T_2 + 2(a+1)m_2) - \varepsilon_1 (K_1 - K_2)((a+1)m_1 + T_1) \\
&L_{12}^{-1} = -M_{[-1]} M_{[X_1]} N M_{[a]} M_{[X_2]} M_{[-1]} M_{[a]}^{-1} \\
&\mathrm{tr}(L_{12}^{-1}) = (a+1)m + (K_1 - K_2)
\end{aligned}$$

The former relations give now, using the linearity of the trace:

$$\begin{aligned}
\mathrm{tr}(L^{-1}) + 2 = {} & \varepsilon_1 T_1^2 - 2\varepsilon_1 (a+1) m_1 T_1 + \varepsilon_2 T_2^2 \\
& + 2\varepsilon_2 (a+1) m_2 T_2 + \varepsilon_1 \varepsilon_2 (a+1) m T_1 T_2 + \Delta(L^{-1})
\end{aligned}$$

where:

$$\Delta(L^{-1}) = \varepsilon_1 \varepsilon_2 (K_1 - K_2)((a+1)(u + 2\varepsilon_2 \mu) + K_1 - K_2 - T_1 T_2)$$

Subtracting from the relation given before for $\mathrm{tr}(ABA^{-1}B^{-1}) + 2$, we obtain:

$$\begin{aligned}
&\Delta(L^{-1}) - \Delta(ABA^{-1}B^{-1}) \\
&= (a+1)^2 [\varepsilon_1 \varepsilon_2 m^2 + \varepsilon_1 m_1^2 + \varepsilon_2 m_2^2 - \varepsilon_1 \varepsilon_2 (a+1) m m_1 m_2 + \varepsilon_1 \varepsilon_2 u m]
\end{aligned}$$

But with the relations already given, we easily evaluate:

$$\Delta(ABA^{-1}B^{-1}) - \Delta(L^{-1}) = \varepsilon_1\varepsilon_2(K_1 - K_2)(a+1)^2 m_1 m_2$$

Simplifying with $(a+1)^2$, the equation $(M^{s_1 s_2}(a, \partial K, u))$ appears:

$$m^2 + \varepsilon_2 m_1^2 + \varepsilon_1 m_2^2 = (a+1)mm_1m_2 + \varepsilon_2 \partial K m_1 m_2 - um$$

For $\partial K = 0$, $\varepsilon_1 = \varepsilon_2 = 1$, $a = 2$ and $u = 0$, we get the equation $M^{++}(2,0,0)$. If $\partial K \neq 0$ and $\varepsilon_1 = \varepsilon_2 = 1$, we obtain the equation $M^{++}(a, \partial K, u)$ mentionned in [24]. If $\partial K = 0$, $\varepsilon_1 = -\varepsilon_2 = 1$ and $u = 0$, we find the equation $M^{+-}(a,0,0)$ completely solved in [20]:

$$m^2 - m_1^2 + m_2^2 = (a+1)mm_1m_2$$

It is very easy to see that the equation $(M^{s_1 s_2}(a, \partial K, u))$ is equivalent to the union of the two following equations:

$$m_1^2 - \partial K m_1 m_2 + \varepsilon_1\varepsilon_2 m_2^2 = \mu m$$
$$\varepsilon_2 \mu = (a+1)m_1 m_2 - m - u$$

In order to obtain it again, we only have to eliminate the terms μ. Conversely, the former calculus give the system of the two equations thanks to $(M^{s_1 s_2}(a, \partial K, u))$.

We find also the equation $(M^{-s_1 - s_2}(a, -\partial K, u + 2\varepsilon_2\mu))$ with the same solutions:

$$m^2 - \varepsilon_2 m_1^2 - \varepsilon_1 m_2^2 = (a+1)mm_1m_2 - \varepsilon_2 \partial K m_1 m_2 - (u + 2\varepsilon_2\mu)m$$

We remark that the former conjunction of two equations simply appears in the situation where we have:

$$m \mid m_1^2 - \partial K m_1 m_2 + \varepsilon_1\varepsilon_2 m_2^2$$

We need only to give ε_2 and to define:

$$\mu = ((m_1^2 - \partial K m_1 m_2 + \varepsilon_1\varepsilon_2 m_2^2)/m)$$

$$u = (a+1)m_1 m_2 - m - \varepsilon_2\mu$$

So, we can build a generalized Markoff equation, and the associated equation with the corresponding two matrices A and B. More generally, with the condition:

$$m \mid \alpha m_1^2 + \beta m_1 m_2 + \gamma m_2^2$$

We find analog equations, studied in [24] or [26]:

$$m^2 + \alpha m_1^2 + \gamma m_2^2 = (a+1)mm_1m_2 - \beta m_1 m_2 - um$$

Such equations contain all the information leading to the theory of quadratic forms [11] [32].

We saw that the number ∂K does not depend on T. For u, the dépendance on T is complicated with a dependance on a and X_2:

$$u + \varepsilon_2\mu = (a\kappa_2 + \lambda)m_2^2 - (a\mu + \kappa_1 - \kappa_2)m_2k_{21} - \mu k_{21}^2 = \Psi_{(T,a)}(m_2, k_{21})$$

The possible values u come from values represented by the form $\Psi_{(T,a)}$. But there exists [8] an algorithmic process determining whether, for any value u we can find m_2 and k_{21}, and also an effective method to compute these values. The number ε_2 being in $\{\pm 1\}$, we deduce a possible sequence X_2.

This gives the complete answer to the question of knowing the values u, hence diophantine equations $(M^{s_1s_2}(a, \partial K, u))$ having solutions. We have also the answer to the question of knowing the possible sequences X_2, when (T, a) is given (see [21]).

All the classical Markoff theory is, with $(T, a) = (\emptyset, 2)$, given by the following form, linked with the golden ratio:

$$\Psi_{(\emptyset,2)}(x, y) = x^2 - xy - y^2$$

4 Solving the Generalized Equations

It is possible to get arborescent structures for the solutions $(m, m_1, m_2) \in (N^*)^3$ of the equation $(M^{s_1s_2}(a, \partial K, u))$. In order to do it, we use the following three transformations:

$$\begin{aligned}
&X : (A, B) \longrightarrow (A^X, B^X)\\
&A^X = M_{[\lhd]}A^{-1}M_{[\lhd]}\\
&B^X = M_{[a]}^{-1}B^*M_{[a]}\\
&Y : (A, B) \longrightarrow (A^Y, B^Y)\\
&A^Y = M_{[\lhd]}M_{[a]}^{-1}M_{[\rhd]}(A^*)^{-1}M_{[a]}\\
&B^Y = M_{[a]}^{-1}A^*M_{[\rhd]}M_{[a]}M_{[\lhd]}M_{[a]}^{-1}B^*A^*M_{[a]}\\
&Z : (A, B) \longrightarrow (A^Z, B^Z)\\
&A^Z = M_{[a]}^{-1}B^*A^*M_{[\rhd]}M_{[a]}M_{[\lhd]}M_{[a]}^{-1}B^*M_{[a]}\\
&B^Z = M_{[a]}^{-1}(B^*)^{-1}M_{[a]}M_{[\lhd]}M_{[a]}^{-1}M_{[\rhd]}M_{[a]}
\end{aligned}$$

These transformations are idempotent, as is easy to verify. We can compose them in order to get transformations that can be applied to Cohn's triples, according to the methods of [25]. But these transformations change from one level to another level the equation (M), as we are going to see it.

• The lefthand construction is defined with the sequences X_2 and T as follows:

$$X_2^G = (\triangleleft T^*, a, X_2)$$
$$T^G = T$$

It gives:

$$A^G = M_{[\triangleleft X_2^{G*},a]} = M_{[\triangleleft]} M_{[a]}^{-1} B^* M_{[\triangleright]} M_{[a]}$$
$$B^G = M_{[X_1^{G*}\triangleright,a]} = B M_{[a]}^{-1} M_{[\triangleright]} (A^*)^{-1} M_{[a]} M_{[\triangleleft]} B$$

We deduce:

$$\varepsilon_2^G = \det(A^G) = \det(B^*) = \varepsilon_1$$
$$\varepsilon_1^G = \det(B^G) = \det((A^*)^{-1}) = \varepsilon_2$$
$$u^G = \varepsilon_1 \varepsilon_2 u$$
$$(\partial K)^G = \partial K$$

The resulting lefthand equation (M^G) can be computed from (M). It can be different from (M), but is to it equal when $\varepsilon_1 = \varepsilon_2 = 1$.

• Twice on the right, the construction is given by:

$$X_2^{DD} = X_2^*$$
$$T^{DD} = (\triangleleft X_2^*, a, T^*, a, X_2 \triangleright)$$

We find:

$$A^{DD} = M_{[\triangleleft X_2^{DD*},a]} = M_{[\triangleleft]} M_{[a]}^{-1} A^* M_{[\triangleright]} M_{[a]}$$
$$B^{DD} = M_{[X_1^{DD*}\triangleright,a]} = AB M_{[a]}^{-1} M_{[\triangleright]} A^* M_{[\triangleright]} M_{[a]}$$

With these expressions:

$$\varepsilon_2^{DD} = \det(A^{DD}) = \det(A^*) = \varepsilon_2$$
$$\varepsilon_1^{DD} = \det(B^{DD}) = \det(B) = \varepsilon_1$$
$$u^{DD} = u$$
$$(\partial K)^{DD} = \varepsilon_2 \partial K$$

The resulting equation (M^{DD}) coming from (M) can be computed. It can be different from (M), but is equal when we have $\varepsilon_1 = \varepsilon_2 = 1$ or $\partial K = 0$.

• The construction on the right after left is defined by:

$$X_2^{GD} = (\triangleleft X_2^*, a, T)$$
$$T^{GD} = (X_2^*, a, T^*, a, X_2)$$

It gives:

$$A^{GD} = M_{[\triangleleft X_2^{GD*},a]} = M_{[\triangleleft]} B M_{[a]}^{-1} M_{[\triangleright]} M_{[a]}$$
$$B^{GD} = M_{[X_1^{GD*}\triangleright,a]} = M_{[a]}^{-1} B^* A^* M_{[\triangleright]} M_{[a]} M_{[\triangleleft]} M_{[a]}^{-1} B^* M_{[a]}$$

With the expressions:

$$\varepsilon_2^{GD} = \det(A^{GD}) = \det(B) = \varepsilon_1$$
$$\varepsilon_1^{GD} = \det(B^{GD}) = \det(A^*) = \varepsilon_2$$
$$u^{GD} = \varepsilon_2 u \text{ ou } u^{GD} = \varepsilon_1\varepsilon_2 u$$
$$(\partial K)^{GD} = \varepsilon_2 \partial K$$

The equation (M^{GD}) coming from (M) can be deduced. It can also be different from (M), but is equal when $\varepsilon_1 = \varepsilon_2 = 1$.

The repeated application of the transformations that we have exhibited shows easily that any equation (M) having a Cohn's triple of solutions have an infinity of other solutions. We deduce from them longer and longer sequences X_2 corresponding to such solutions. The same transformations organize the set of triples of solutions as arborescent structures.

Also, considering for (M) the triples of solutions $(m, m_1, m_2) \in Z^3$, the transformations X, Y and Z constitute sets of triples as arborescent structures. It is easy to show that are a finite number of them (see [23]).

5 Application to the Analysis of the Markoff Spectrum

For the equation (M) we can realize the same calculus as the one made by Cassels for the classical Markoff equation [3]. We introduce for that the quadratic forms:

$$\phi(z, y) = z^2 + ((a+1)m + K_1 - K_2)zy + \varepsilon_1\varepsilon_2 y^2 = m^2 F(x, y)$$
$$F(x, y) = x^2 + \left[\frac{(a+1)m - K_1 - K_2}{m}\right] xy + \left[\frac{l - (a+1)K_1}{m}\right] y^2$$

where:

$$z = mx - K_1 y$$

We find properties similar to those already met in [21]. Moreover, thanks to (M), we have:

$$\begin{aligned}\phi(-\varepsilon_1 m_2, m_1) &= \varepsilon_2\varepsilon_1 m_1^2 - \varepsilon_1(a+1)mm_1m_2 - \varepsilon_2\varepsilon_1\partial K m_1 m_2 + m_2^2 \\ &= -\varepsilon_1 m^2((m+u)/m) \\ &= m^2 F(k_1, m_1)\end{aligned}$$

The discriminant of F can easily be computed:

$$\Delta(F) = \left[\frac{((a+1)m + K_1 - K_2)^2 - 4\varepsilon_1\varepsilon_2}{m^2}\right] = \frac{\Delta_a(S)}{m^2}$$

We have only to compute the arithmetical minimum of F:

$$m(F) = \inf\{|\ F(x, y)\ |: (x, y) \in Z^2 - \{(0, 0)\}\}$$

In order to deduce the Markoff constant of F :

$$C(F) = (m(F)/\sqrt{\Delta_a(S)})$$

This number measures the best way to approximate by rational numbers (p/q) the following irrational number associated to F:

$$\theta_a(S) = \frac{K_1 + K_2 - (a+1)m + \sqrt{\Delta_a(S)}}{2m}$$

Without difficulty, we find:

$$C(F) = \underline{\lim}_{q \in Z - \{0\}} \ (q^2 \left| \theta_a(S) - \frac{p}{q} \right|)$$

The Markoff spectrum is the sets of these values $C(F)$, also noted $C(\theta_a(S))$. In many cases, we have:

$$C(F) = \frac{\mid m + u \mid}{\sqrt{((a+1)m + K_1 - K_2)^2 - 4\varepsilon_1\varepsilon_2}} \leq \frac{m}{\sqrt{\Delta_a(S)}}$$

This readily yields:

$$F(k_1, m_1) = -\varepsilon_1((m+u)/m)$$
$$F(K_1, m) = F(K_2 - (a+1)m, m) = \varepsilon_1\varepsilon_2$$

The only cases to consider, in order to compute $C(F)$ are the cases where we have:

$$-2m < u \leq 0$$

We find a lot of possibilities as soon as we have an arborescent structure of solutions, because the values m increase then to infinity with the length of the sequences S^* that appear. But the number u is given by the form $\Psi_{(T,a)}$ depending only on T and a, with:

$$u + \varepsilon_2\mu = \Psi_{(T,a)}(m_2, k_{21})$$

Hence, the same sequence (T, a) gives the possibility to analyze a part of the Markoff spectrum constituted with constants having the form given before.

If we have an infinity of possibilities for m such that:

$$m(F) = F(k_1, m_1) = -\varepsilon_1((m+u)/m)$$

we find at the limit a unique accumulation point for the Markoff constants that we consider. It has the form:

$$\frac{1}{(a+1)}$$

The real difficulty for progressing in the analysis is to ensure that we get:

$$m(F) = F(k_1, m_1)$$

We have in fact encountered cases where this is not true [26].

6 A Link with the Representation of Free Groups

Let us consider two matrices A and B built as before. They generate a group $g = gp(A, B)$ in $GL(2, Z)$, which is an image of the abstract free group F_2 represented by:

$$\pi_{(A,B)} : F_2 = \langle x, y \rangle \longrightarrow GL(2, Z)$$

With every word $W = W(x, y)$ of F_2 written with x and y, we put:

$$\pi_{(A,B)}(W(x, y)) = W(A, B)$$

The group $\ker(\pi_{(A,B)})$ is a subgroup of the free group F_2. By a classical result of Nielsen and Schreier ([17] p.95), it is also a free group. The group $\ker(\pi_{(A,B)})$ is not necessarily of finite rank, though the rank is denumerable. If $W_1, W_2, ...$ are terms of F_2 generating $\ker(\pi_{(A,B)})$, we write:

$$\begin{aligned} G = gp(A, B) &= \Im(\pi_{(A,B)}) \\ &= \langle A, B; W_1(A, B) = W_2(A, B) = ... = 1 \rangle \\ &\simeq \langle x, y; W_1 = W_2 = ... = 1 \rangle \\ &= F_2 / \ker(\pi_{(A,B)}) \end{aligned}$$

and when $\ker(\pi_{(A,B)}) = \{1\}$, we have only:

$$G = \langle A, B \rangle \simeq \langle x, y \rangle = F_2$$

In order to use on G transformations X, Y, Z we introduce two matrices $M_{[\lhd]}$ and $M_{[a]}$. But this is not adapted here, because we want to have the guarantee that the computations are made inside the group G. Then we use a deformation building transformations again idempotent:

$$\begin{aligned} &\overline{X} : (A, B) \longrightarrow (A^{\overline{X}}, B^{\overline{X}}) \\ &A^{\overline{X}} = A^{-1} \\ &B^{\overline{X}} = B \\ &\overline{Y} : (A, B) \longrightarrow (A^{\overline{Y}}, B^{\overline{Y}}) \\ &A^{\overline{Y}} = A^{-1} \\ &B^{\overline{Y}} = ABA \\ &\overline{Z} : (A, B) \longrightarrow (A^{\overline{Z}}, B^{\overline{Z}}) \\ &A^{\overline{Z}} = BAB \\ &B^{\overline{Z}} = B^{-1} \end{aligned}$$

These notations are not precise enough. We must write them, for any word $W = W(x, y)$ of F_2 acting on G, in the following manner:

$$\overline{X}(W)(A, B) = W^{\overline{X}}(A, B) = W(A^{-1}, B)$$
$$\overline{Y}(W)(A, B) = W^{\overline{Y}}(A, B) = W(A^{-1}, ABA)$$
$$\overline{Z}(W)(A, B) = W^{\overline{Z}}(A, B) = W(BAB, B^{-1})$$

Thus, $A^{\overline{X}}$ designates $x^{\overline{X}}(A, B) = A^{-1}$, and $B^{\overline{X}}$ is the term $y^{\overline{X}}(A, B) = B$, etc.

The transformations $\overline{X}$, $\overline{Y}$, $\overline{Z}$, then become automorphisms of F_2. We call them elementary automorphisms:

$$W^{\overline{X}} = W^{\overline{X}}(x, y) = \overline{X}(W)(x, y) = \overline{X}(W) = \overline{X}(W(x, y)) = W(x^{-1}, y)$$
$$W^{\overline{Y}} = W^{\overline{Y}}(x, y) = \overline{Y}(W)(x, y) = \overline{Y}(W) = \overline{Y}(W(x, y)) = W(x^{-1}, xyx)$$
$$W^{\overline{Z}} = W^{\overline{Z}}(x, y) = \overline{Z}(W)(x, y) = \overline{Z}(W) = \overline{Z}(W(x, y)) = W(yxy, y^{-1})$$

For example, the transformation $\overline{Z}$ is operating on the word $W(x, y)$ sustituing in this word yxy to x, and y^{-1} to y.

Thanks to composition, we build the essential automorphisms of F_2:

$$\overline{N} = \overline{N}(\overline{X}, \overline{Y}, \overline{Z})$$

The group of essential automorphisms of F_2 constitutes a proper subgroup of the group $\mathrm{Aut}(F_2)$ of the automorphisms of F_2. It is an image of the triangle group $Z/2Z * Z/2Z * Z/2Z$, and a free product of three cyclic groups with two elements.

So, we find the possibility to define other couples of generators of G:

$$G = gp(\overline{N}(x)(A, B), \overline{N}(y)(A, B))$$

An interesting phenomenon is that the traces of the expressions that we consider are similar. They are equal for the classical Markoff theory. So, we have:

$$\mathrm{tr}(A^X) = \mathrm{tr}(A^{\overline{X}}) = \varepsilon_2((a+1)m_2 + T_2)$$
$$\mathrm{tr}(B^X) = \mathrm{tr}(B^{\overline{X}}) = ((a+1)m_1 - T_1)$$
$$\mathrm{tr}(A^Y) = \varepsilon_2((a+1)m_2 - T_2)$$
$$\mathrm{tr}(A^{\overline{Y}}) = \varepsilon_2((a+1)m_2 + T_2)$$
$$\mathrm{tr}(B^Y) = ((a+1)m)((a+1)m_2 + T_2) - \varepsilon_2((a+1)m_1 + T_1)$$
$$\mathrm{tr}(B^{\overline{Y}}) = ((a+1)m + K_2 - K_1)((a+1)m_2 + T_2) - ((a+1)m_1 - T_1)$$
$$\mathrm{tr}(A^Z) = ((a+1)m)((a+1)m_1 - T_1) - \varepsilon_1((a+1)m_2 - T_2)$$
$$\mathrm{tr}(A^{\overline{Z}}) = ((a+1)m + K_2 - K_1)((a+1)m_1 - T_1) - ((a+1)m_2 + T_2)$$
$$\mathrm{tr}(B^Z) = \varepsilon_1((a+1)m_1 + T_1)$$
$$\mathrm{tr}(B^{\overline{Z}}) = \varepsilon_1((a+1)m_1 - T_1)$$

For any automorphism $\tau \in \mathrm{Aut}(F_2)$, we write also τ the application:

$$(\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A)) \longrightarrow (\mathrm{tr}(\tau(xy)(A,B)), \mathrm{tr}(\tau(y)(A,B)), \mathrm{tr}(\tau(x)(A,B)))$$

We get with what we saw:

$$\overline{X} : (\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A)) \longrightarrow ((\mathrm{tr}(A)\mathrm{tr}(B) - \mathrm{tr}(AB)), \mathrm{tr}(B), \varepsilon_A \mathrm{tr}(A))$$
$$\overline{Y} : (\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A)) \longrightarrow (\mathrm{tr}(AB), \mathrm{tr}(AB)\mathrm{tr}(A) - \mathrm{tr}(B), \varepsilon_A \mathrm{tr}(A))$$
$$\overline{Z} : (\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A)) \longrightarrow (\mathrm{tr}(AB), \varepsilon_B \mathrm{tr}(B), \mathrm{tr}(AB)\mathrm{tr}(B) - \mathrm{tr}(A))$$

• For any field k, we write usually $R_k(G,2)$ the set of homomorphisms ρ of the group G in $SL(2,k)$. Clearly, $R_k(G,2)$ is a subset of the set of representations of degree 2 from G to k:

$$\rho : G \longrightarrow GL(2,k)$$

The former developments give, when the characteristic of k is equal to zero, informations about the characters of these representations:

$$\chi_\rho : C \in G \longrightarrow \chi_\rho(C) = tr(\rho(C)) \in k$$

In order to define a representation $\rho \in R_k(G,2)$, we have only to dispose of two matrices $\rho(A)$ and $\rho(B)$ contained in $SL(2,k)$.

Using $\pi_{(A,B)}$, and composing with ρ, we get an application:

$$\widetilde{\pi_{(A,B)}} : \rho \in R_k(G,2) \longrightarrow \rho' = \rho \circ \pi_{(A,B)} \in R_k(F_2,2)$$

This verifies the condition:

$$\ker(\pi_{(A,B)}) \subset \ker(\rho')$$

If k is a field with charateristic 0, the ring Z immerges in a natural way in k, and $SL(2,Z)$ appears in $SL(2,k)$.

If $\det(A) = \det(B) = 1$, we find a natural representation of G in $SL(2,k)$ with $G \subset SL(2,Z)$.

In the more general cases, it is easy to see [2] that $R_k(G,2)$ can be considered as an affine k-algebraic set.

When $k = C$, we also find the affine C-algebraic set [2]:

$$R_C(F_2,2) = SL(2,C) \times SL(2,C)$$

It has dimension 6 and can be projected on C^3 thanks to:

$$\varphi' : \rho' \in R_C(F_2,2) \longrightarrow \varphi'(\rho') = (tr\rho'(xy), tr\rho'(y), tr\rho'(x)) \in C^3$$

So we have an interest to consider the former triples.

• Another way to deal with the situation is to consider the $Z[GL(2,Z)]$-modulus Z^2. The isomorphism of two such structures associated to the two representations ρ_1 and ρ_2, is equivalent to the existence of $P \in GL(2,Z)$ a matrix such that:

$$\rho_1(M) = P\rho_2(M)P^{-1} \quad \text{for any } M \in GL(2,Z)$$

Hence, the classical Markoff theory also gives information about the structure of $Z[GL(2,Z)]$-modulus on Z^2.

• For any field k with characteristic zero, we have a natural extension with the representation Ad of $SL(2,k)$ in the k-space k^2. This can be considered as a $k[SL(2,k)]$-modulus. The extension of the field k gives ramification phenomenons and resolution of singularities. A representation $\rho \in R_C(G,2)$ gives also on k^2 a structure of $k[G]$-modulus. The isomorphism of the two structures of $k[G]$-moduli on k^2, associated to ρ_1 and ρ_2, is given thanks to a matrix $P \in SL(2,k)$ verifying:

$$\rho_1(M) = P\rho_2(M)P^{-1} \qquad \text{for any } M \in G$$

With [2] p.101, we get a parametrization of all the classes of semi-simple representations for this equivalence thanks to points of an affine algebraic set. In the case when $k = C$ or $G = F_2$, we find all the corresponding characters. With the notations of [2] and what we have seen before:

$$X(F_2,2) = \varphi'(R_C(F_2,2)) = C^3$$

We refer the reader to [16] for more details on the characters.

We designate by Φ_2 the group $\mathrm{Aut}(F_2)$ of automorphisms of the group F_2. It acts on the couples of generators of F_2.

Thanks to a result of B.H. Neumann, we have the following presentation [17] p.169, where the product that we use is the composition of automorphisms:

$$\Phi_2 = \langle \sigma, P, U; \sigma^2 = P^2 = (UP\sigma P)^2 = (\sigma PU)^3 = (P\sigma)^4 = [U, \sigma U\sigma] = 1 \rangle$$

Its uses the three Nielsen transformations, defined by:

$$\sigma : W(x,y) \longrightarrow W(x^{-1},y)$$
$$P : W(x,y) \longrightarrow W(y,x)$$
$$U : W(x,y) \longrightarrow W(xy,y)$$

Changing U, we prefer to consider here V such that $V^3 = 1$:

$$V = (U\sigma P) : W(x,y) \longrightarrow W(y, y^{-1}x^{-1})$$

These transformations act on the traces $(\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A))$ with:

$$\sigma : (\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A)) \longrightarrow (\mathrm{tr}(A)\mathrm{tr}(B) - \mathrm{tr}(AB), \mathrm{tr}(B), \varepsilon_A \mathrm{tr}(A))$$
$$P : (\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A)) \longrightarrow (\mathrm{tr}(AB), \mathrm{tr}(A), \mathrm{tr}(B))$$
$$V : (\mathrm{tr}(AB), \mathrm{tr}(B), \mathrm{tr}(A)) \longrightarrow (\varepsilon_A \mathrm{tr}(A), \varepsilon_{AB} \mathrm{tr}(AB), \mathrm{tr}(B))$$

Composing, we find all the possible permutations of the three traces. This result gives the possibility to write expressions similar to the action of $\overline{X}$, $\overline{Y}$, $\overline{Z}$, on the triples of traces. We deduce:

$$\sigma = \overline{X} : W(x, y) \longrightarrow W(x^{-1}, y)$$
$$PVVP\sigma VPVVP = Ad(yxy)\overline{Z} : W(x, y) \longrightarrow (yxy)W(yxy, y^{-1})(yxy)^{-1}$$
$$P\overline{Z}P = \overline{Y} : W(x, y) \longrightarrow W(x^{-1}, xyx)$$

Composing, we get now expressions for all the essential transformations $\overline{N}(\overline{X}, \overline{Y}, \overline{Z})$ with the three Nielsen transformations σ, P, V, and inner automorphisms of F_2.

The interest of the former development is that we can change the couple of generators in the group G, keeping under control, step by step, all the successive changings that we do. For this, we define a notion of height:

$$h(A, B) = (\mid \mathrm{tr}(AB) \mid, \mid \mathrm{tr}(B) \mid, \mid \mathrm{tr}(A) \mid) \in N^3$$

The elementary transformations $\overline{X}$, $\overline{Y}$, $\overline{Z}$, are used in order to minimize this height according to the methods of [9] [23]. We deduce an algorithm for reducing couples of generators (A, B) in G. This algorithm identifies for any couple an essential automorphism reducing it, and the corresponding couple with a less height. With this algorithm, we get:

Theorem 1. *We consider two matrices built as before:*

$$A = M_{[\triangleleft X_2^*, a]} \in SL(2, Z)$$
$$B = M_{[X_1^* \triangleright, a]} \in SL(2, Z)$$

The group G generated by A and B in $SL(2, Z)$ is free and of rank 2. This construction gives the faithful representations of the free group with two generators F_2 in $SL(2, Z)$. They are all given by this way, at a $SL(2, Z)$-equivalence. Moreover, when two of them are taken different, they are not equivalent.

This result gives, at an equivalence, the faithful representations of F_2 in the modular group $PSL(2, Z)$.

Lifting the restriction to cases where $\det(\varepsilon_A) = \det(\varepsilon_B) = 1$, we then generalize to faithful representations of F_2 in the group $GL(2, Z)$.

As a fuchsian group, the image of F_2 acts in the half plane of Poincaré H_2. The Riemann surface which is deduced is a torus perforated by the extraction of a closed disk. So we have the garantee that the image group of F_2 in $PSL(2, Z)$ is a Fricke group [7]. Hence, we determine all the Fricke

groups given by subgroups of $GL(2, Z)$. Conversely, for any torus perforated by the extraction of a closed disk, the fundamental group is isomorphic to F_2. The mapping class group is isomorphic to the quotient of $\mathrm{Aut}(F_2)$ by the subgroup of inner automorphisms [18].

Our construction generalizes completely to conformal toruses, and to equations $(M^{s_1 s_2}(a, \partial K, u))$ linked to them, the approach of Harvey COHN [5]. It gives some complementary aspects of other articles quoted in our bibliography.

References

[1] V. I. ARNOLD, A. AVEZ, Problèmes ergodiques de la mécanique classique, Gauthier-Villars (Paris), 1967
[2] G. BAUMSLAG, Topics in combinatorial group theory, Lectures in Mathematics, ETH Zürich, Birkhauser, 1993
[3] J. W. S. CASSELS, An introduction to diophantine approximation, Cambridge Tracts in Mathematics and Mathematical Physics (47), 1957
[4] M. COHEN, W. METZLER, A ZIMMERMANN, What does a basis of F(a,b) look like, Mathematische Annalen (257), 1981, pp. 534-445
[5] H. COHN, Approach to Markoff's minimal forms through modular functions, Annals of Mathematics, vol 61 (1), 1955, pp. 1-12
[6] H. COHN, Representation of Markoff's binary quadratic forms by geodesics on a perforated torus, Acta Arithmetica (XVIII), 1971, pp. 123-136
[7] H. COHN, Remarks on the cyclotomic Fricke groups, Kleinian groups and related topics, Proceedings Oaxtepec 1981, Lecture Notes in Mathematics (971), Springer Verlag, 1982
[8] J. H. CONWAY, The sensual (quadratic) form, The Carus Monograph (26), M.A.A., 1997
[9] T. W. CUSICK, On Perrine's generalized Markoff equation, Aequationes Mathematicae 46 (3), 1993, pp. 203-211
[10] M. E. FLAHIVE, T. W. CUSICK, The Markoff and Lagrange spectra, Mathematical Surveys and Monographs 30, A.M.S., 1989
[11] C. F. GAUSS, Recherches arithmétiques, traduction Poullet-Delisle, Paris, 1807, réédition A. Blanchard, 1953
[12] L. K. HUA, I. REINER, Automorphisms of the unimodular group, Trans. Amer. Math. Soc. 72, 1952, pp. 467-473
[13] N. V. IVANOV, Automorphisms of Teichmuller modular groups, topology and geometry (Rohlin seminar), Lecture Notes in Mathematics (1346), Springer Verlag, 1989
[14] G. A. JONES, Congruence and non-congruence subgroups of the modular group: a survey, A survey on groups with a single defining relation, Proceedings of the International Conference of St Andrews, London Mathematical Society Technical Seminars (121), 1986, pp. 38-58
[15] L. KEEN, On Fricke moduli, Advances in the theory of Riemann surfaces (L.V. Ahlfors), Annals of Mathematical Studies 66, Princeton University Press, 1971, pp.205-224
[16] L. LE BRUYN, Trace rings of generic 2 by 2 matrices, Memoirs of the A.M.S. (363), 1987

[17] W. MAGNUS, A. KARRASS, D. SOLITAR, Combinatorial group theory, Dover, 1976, (2ème édition)
[18] J. NIELSEN, Die Isomorphismen der allgemeinen, unendlichen Gruppe mit zwei Erzeugenden, Math. Ann. (78), 1917, pp. 385-397
[19] A. A. MARKOFF, Sur les formes quadratiques binaires indéfinies, Mathematische Annalen, Band VI, 1879, pp. 381-406, Band XVII, 1880, pp. 379-399
[20] S. PERRINE, Approximation diophantienne (Théorie de Markoff), Thèse présentée à l'Université de Metz, décembre 1988
[21] S. PERRINE, Sur une généralisation de la théorie de Markoff, Journal of Number Theory, vol 37 (2), 1991, pp. 211-230
[22] S. PERRINE, La méthode de Poincaré appliquée à l'arithmétique, Contribution au Congrès International Henri Poincaré, Nancy, 14-18 mai 1994
[23] S. PERRINE, Sur des équations de Markoff généralisant celle de Markoff, Annales de la Faculté des Sciences de Toulouse, vol VI (1), 1997, pp. 124-141
[24] S. PERRINE, Trees of approximation constants, Contributed talk at the conference on continued fractions: From analytic number theory to constructive approximation, University of Missouri, Columbia, May 20-23, 1998 (to be published in Contemporary Mathematics)
[25] S. PERRINE, Un arbre de constantes d'approximation analogue à celui de l'équation diophantienne de Markoff, Journal de Théorie des Nombres de Bordeaux 10, 1998, pp. 321-353
[26] S. PERRINE, About the solutions of the diophantine equation: $x^2+y^2+z^2 = 3xyz+2x$, submitted
[27] M. PLANAT, Synchronization, topology and oscillators, Ann. Telecommun., 51(7-8), Juillet Aout 1996, pp. 391-406
[28] M. PLANAT, S. DOS SANTOS, J. CRESSON, S. PERRINE, 1/f frequency noise in a communication receiver and the Riemann hypothesis, Proc. 15th International Conference on Noise in Physical Systems and $1/f$ Fluctuations, C. Surya (Eds), Bentham Press, 1999, pp 409-412
[29] P. SCHMUTZ SCHALLER, Geometry of Riemann surfaces based on closed geodesics, Bull. Amer. Math. Soc, vol 35 (3), july 1998, pp. 193-214
[30] M. R. SCHROEDER, Number theory in Science and Communication, Springer Verlag, 1984
[31] M. SHEINGORN, Characterization of simple closed geodesics on Fricke surfaces, Duke Math. J. 52, 1985, pp. 535-545
[32] M. SHEINGORN, Rational solutions of $\sum_{i=1}^{3} a_i x_i^2 = dx_1x_2x_3$ and simple closed geodesics on Fricke surfaces, Holomorphic functions and moduli (1), Mathematical Sciences Research Institute Publications 10, Springer Verlag, 1988, pp. 229-236
[33] J. P. SERRE, Arbres, amalgames, SL_2, Astérisque 46, S.M.F., 1983