ISTITUTO LOMBARDO
ACCADEMIA di SCIENZE e LETTERE

Rendiconti

CLASSE DI SCIENZE MATEMATICHE E NATURALI

Vol. 146
(2012)

Istituto Lombardo di Scienze e Lettere

MILANO
ISSN 1974-6989

ISTITUTO LOMBARDO - ACCADEMIA di SCIENZE e LETTERE

RENDICONTI

CLASSE DI SCIENZE MATEMATICHE E NATURALI

Vol. 146
(2012)

Istituto Lombardo di Scienze e Lettere

MILANO

ISTITUTO LOMBARDO - ACCADEMIA DI SCIENZE E LETTERE

RENDICONTI

CLASSE DI SCIENZE MATEMATICHE E NATURALI

Responsabile scientifico
Leone Corradi Dell'Acqua

Comitato di redazione
Gianpiero Sironi
Angelo Stella
Leone Corradi Dell'Acqua
Giuseppe Frasso

Pubblicato con il contributo di
Fondazione A. Baselli
Politecnico di Milano
Università degli Studi dell'Insubria, Varese
Università degli Studi di Milano
Università degli Studi Milano-Bicocca
Università degli Studi di Pavia

ISSN 1974-6989
Copyright 2015

PAGEPress Publications
via G. Belli 7
27100 Pavia, Italy
T. +39.0382.1751762
F: +39.0382.1750481

NUOVI APPROCCI TEORICI E SPERIMENTALI ALLA SCOPERTA DI FARMACI

Nota del s.c. MARTINO BOLOGNESI (*)

(Adunanza del 19 gennaio 2012)

SUNTO. – Diversi motivi ambientali, demografici ed ecologici suggeriscono che virus noti o sconosciuti continueranno ad emergere negli anni a venire in tutto il mondo, presentando una continua minaccia per la popolazione umana. Gli interventi terapeutici oggi disponibili presentano diverse limitazioni; le campagne di vaccinazione nel terzo mondo spesso incontrano problemi di distribuzione, e raggiungono una frazione insufficiente della popolazione. Per molti virus un vaccino non è disponibile, una carenza che si traduce in elevata mortalità nei paesi tropicali e tra i neonati. Su un diverso fronte, la resistenza dei batteri agli antibiotici è in preoccupante crescita, ed e' spesso associata a infezioni nosocomiali. Di conseguenza, nuove idee e approcci alla scoperta o progettazione di farmaci efficaci rappresentano una priorità della ricerca, per offrire strumenti terapeutici e di prevenzione contro potenziali pandemie. In questo contesto, il nostro gruppo presso l'Università di Milano, in collaborazione con diversi laboratori europei ed extraeuropei, studia le proprietà strutturali e funzionali di enzimi coinvolti nella replicazione di virus a RNA, quali bersagli per la progettazione di farmaci antivirali. La logica alla base della scelta di uno specifico enzima virale parte dall'idea che questi enzimi siano sufficientemente diversi dalle loro controparti umane; in tal modo, bloccando l'enzima del virus con un nuovo farmaco, non si dovrebbero registrare reazioni avverse nell'ospite umano. L'approccio alla scoperta di nuovi farmaci, applicato nel nostro laboratorio, è basato su una precisa serie di fasi sperimentali: i) l'analisi della struttura tridimensionale degli enzimi liberi, mediante cristallografia a raggi X; ii) screening preliminare *in silico* (computazionale) delle regioni degli enzimi selezionati verso cui la ricerca del farmaco è mirata (ad esempio, i siti attivi dell'enzima); iii) test biochimici e biofisici di inibizione dell'enzima; iv)

(*) Dipartimento di Bioscienze, Università di Milano, Italy.
E-mail: martino.bolognesi@unimi.it ; http://users.unimi.it/biolstru/Home.html

struttura cristallina e analisi di complessi enzima/inibitori; v) test di attività in cellula/*in vivo*; vi) estrazione di informazioni per l'ottimizzazione della molecola-farmaco. Questo metodo di ricerca si è dimostrato efficace nella scoperta di inibitori a basso peso molecolare di due enzimi chiave del virus della Febbre Gialla (e in parte per il virus Dengue), e di Norovirus. In particolare, per Norovirus abbiamo focalizzato la nostra attenzione sul farmaco Suramina, una molecola attiva e nota da tempo (come farmaco contro la "malattia del sonno"), che era stato selezionato attraverso la procedura biofisica e simulativa sopra descritta. I nostri test cristallografici e di inibizione hanno permesso di evidenziare le modalità di legame dell'inibitore all'enzima, e parametri inibitori funzionali adeguati allo sviluppo di un farmaco. Successivamente, nel contesto di una collaborazione internazionale, abbiamo sperimentato una serie di frammenti molecolari della Suramina, alla ricerca di nuovi composti attivi dotati degli opportuni parametri farmacologici. L'attività di ricerca, che, come descritto, si basa su nuovi approcci concettuali e multidisciplinari, ha portato alla produzione di diversi composti a basso peso molecolare che saranno ulteriormente sviluppati per la produzione di composti antivirali.

***

ABSTRACT. – *New theoretic and experimental approaches to drug discovery.*
Environmental, demographic and ecological reasons suggest that either novel or known viruses will continue to emerge worldwide, posing new threats to the human population. Additionally, therapeutic interventions present different outcomes; for example, vaccination campaigns in the Third World often encounter local distribution problems and reach an insufficient fraction of the population. For many viruses no vaccine is available, such that the toll death is high, particularly in tropical countries and among infants. On the other hand, resistance of bacteria to known antibiotics is increasingly a serious threat, often associated to nosocomial infections. As a result, new ideas and approaches to the discovery or design of new effective drugs are a high priority in all civilized countries, to prevent and be ready to face potential pandemics. In this context, our group at the University of Milano, in collaboration with several European and extra-European labs, has been exploring the structural and functional properties of enzymes involved in the replication of (+)stranded RNA viruses as targets for the design of antiviral drugs. The rationale behind the choice of specific target viral is the idea that these are structurally and functionally sufficiently different from their human counterparts; thus, blocking the virus enzyme with a new drug should not be reflected by adverse reactions in the human host. The discovery approach applied in our laboratory has been based on a series of specific experimental steps: i) the analysis of crystal structure of the free enzymes, through X-ray crystallography; ii) *in silico* (computational) preliminary screening of selected enzyme regions towards which the drug search is targeted (*e.g.* mostly the enzyme active sites); iii) biochemical and biophysical tests of enzyme inhibition; iv) crystal structure analyses of enzyme/inhibitor complexes; v) in cell/*in vivo* testing; vi) inference for drug-lead optimization. This research method proved effective in discovering low molecular weight inhibitors of two key enzymes from Yellow fever virus (and partly for Dengue virus), and for Norovirus. Specifically, we targeted Norovirus studying the long known drug Suramin (used in the therapy of 'sleeping sickness'), which was selected through the procedure described above through our *in silico* docking screenings. Our crystallographic and inhibition assays allowed to highlight the inhibitor binding mode and satisfactory functional inhibitory pa-

rameters. Subsequently, in the context of an international collaboration, we could test a series of Suramin molecular fragments, in search of new active compounds endowed with suitable pharmacological parameters. The described research activity, which is based on new conceptual and multidisciplinary approaches to drug discovery, has led to the production of several small molecules that will be further developed into antiviral compounds.

BIOMOLECOLE DA BASIDIOMICETI COME SORGENTE DI POTENZIALI NUOVI FARMACI

Nota del s.c. GIANLUCA NASINI (*)

(Adunanza del 19 gennaio 2012)

SUNTO. – Sono riportati l'isolamento e la struttura di alcuni composti, biogeneticamente correlati, prodotti in coltura da vari ceppi di Basidiomiceti appartenenti ai generi *Armillaria*, *Laurilia*, *Clitocybe* e *Clavicorona*. I metaboliti isolati sono sesquiterpenoidi contenenti l'anello protoilludanico e sono stati anche correlati con l'attività biologica e schemi biosintetici. Abbiamo analizzato la chimica dei melleolidi, sulcatine, tsugicoline e clavilattoni. In particolare, la spirolaxina, isolata da colture di *Sporotrichum laxum*, è attiva sul batterio Helicobacter pylori ed è un potente inibitore dell'angiogenesi. La sua struttura è stata determinata con evidenze chimiche e spettroscopiche e confermata mediante analisi con i raggi X.

***

ABSTRACT. – We reported the isolation and structural characterization of a new biogenetical related compounds, produced in cultures by various strains of Basidiomycetes of the genera *Armillaria*, *Laurilia*, *Clitocybe* and *Clavicorona* . The isolated metabolites are sesquiterpenoids containing the protoilludane ring system and are even correlated with some biological activity and biogenetic steps. We have analyzed the chemistry of the melleolides, sulcatines, tsugicolines and clavilactones. In particular, the spirolaxine, isolated from cultures of *Sporotrichum laxum*, is active on the bacterium *Helicobacter pylori* and is a potent inhibitor of angiogenesis. Its structure was determined on the basis of chemical and spectroscopic evidence and confirmed by X-ray analysis.

(*) CNR-Istituto di Chimica del Riconoscimento Molecolare. Dipartimento di Chimica, Materiali ed Ingegneria Chimica del Politecnico, Via Mancinelli 7, Milan, Italy. E-mail: gianluca.nasini@polimi.it

1. Introduzione

Scopi di questa ricerca sono stati la coltivazione del micelio, l'isolamento, la determinazione della struttura, la stereochimica, le correlazioni chimiche di nuove sostanze e lo studio dell'attività biologica di nuovi metaboliti prodotti da Basidiomiceti; inoltre la possibilità di coltivare il fungo su diversi substrati aumenta la possibilità di isolare nuovi metaboliti, sfruttando il ricco patrimonio enzimatico del fungo stesso.

Una serie di nuovi sesquiterpeni di origine protoilludanica sono stati isolati da funghi dei generi, *Armillaria, Laurilia, Clitocybe* e *Clavicorona*.

Nel nostro progetto sono stati studiati inizialmente i metaboliti prodotti dal fungo *Armillaria mellea*, un basidiomicete che è noto produrre sostanze che possiedono attività antibatterica; abbiamo isolato su terreno GYA (glucosio-lievito-agar) i melleolidi B-H mentre su terreno contenente MPG (malto-peptone-glucosio), si sono ottenuti i melledonali B-D; questi composti, che hanno una insolita idrossilazione al C-10 e C-13 sono nuovi membri della classe dei sesquiterpeni protoilludanici O-metilorsellinati (*Fig. 1*). Le loro strutture sono state determinate sulla base di correlazioni chimiche e dati spettroscopici. Le configurazioni assolute sono state ottenute mediante cristallografia a raggi X e procedure NMR.[1]

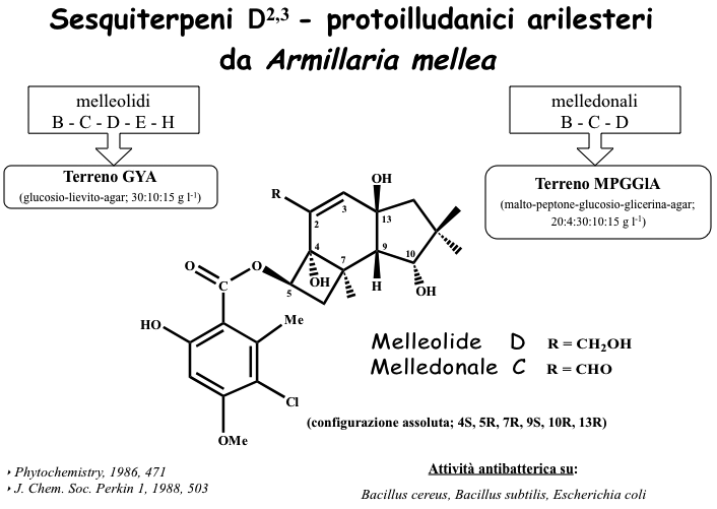


Fig. 1.

2. CENERE: *LAURILIA ; CLAVICORONA*

Cinque nuovi sesquiterpeni, le sulcatine A-E sono state isolate da colture liquido-stazionario (MPGB) di *Laurilia sulcata (Stereum sulcatum)* (*Fig. 2*); il ceppo proviene dalla micoteca di Baarn (CBS-Olanda) e le strutture sono state risolte da misure spettroscopiche e correlazioni HETCOR e COLOC. Le strutture delle sulcatine C-E sono insolite nei sesquiterpeni di origine protoilludanica; l'apertura dell'anello a quattro termini nel co-metabolita sulcatina A può portare ad una reazione retroaldolica con l'apertura dell'anello ciclobutanico dando luogo ad una nuova serie di sesquiterpeni che chiameremo nor-isoilludalani (vedi schema biosintetico) (*Fig. 3*).[2]

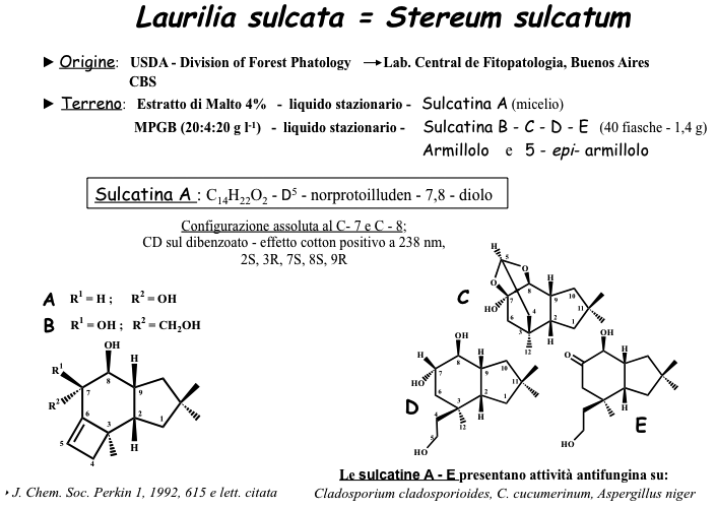


Fig. 2.

Le tsugicoline A-D (1a, 2, 3a, 4) provengono da fermentazione di *Laurilia tsugicola* su terreno MPGB (*Fig. 4*); in particolare 1a è un precursore biogenetico per ottenere altri sesquiterpeni con diversi scheletri; la presenza anomala di una funzione carbonilica nell'anello ciclobutanico, è la chiave della reattività di questo metabolita che

porta ad una facile apertura del legame 6-7 con la formazione di una larga serie di nuovi metaboliti.[3]

La tsugicolina A riarrangia a pH 13 nel composto 8a con una struttura simile ai sesquiterpeni isolattarani e nel composto 9a come la sostanza naturale tsugicolina H 9c; in condizioni basiche differenti, si formano i composti 4a, 5, 6a e 7. L'isolamento della sostanza 2 è il primo esempio di una conversione chimica di un protoilludano in un sesquiterpene sterpuranico (*Fig. 5*).[4]

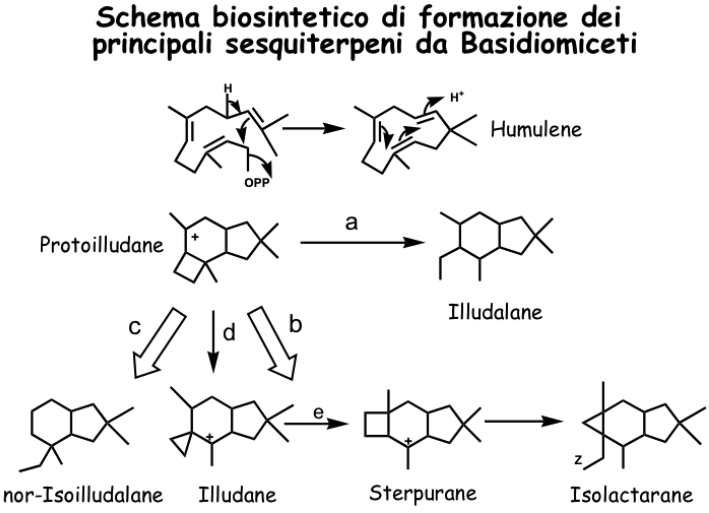


Fig. 3.

Dal fungo *Clavicorona divaricata* (*Fig. 6*) in coltura agarizzata a base di malto-peptone-glucosio (MPG) sono stati isolate principalmente le divaricatine A 3a e B 3b insieme alle tsugicoline L 2b e M 5b a struttura furo-sesquiterpenoidica; le loro strutture sono state determinate con l'aiuto di studi NMR. Tutti i metaboliti sono attivi contro alcuni batteri e inibiscono la germinazione di *Lepidum sativum*.[5]

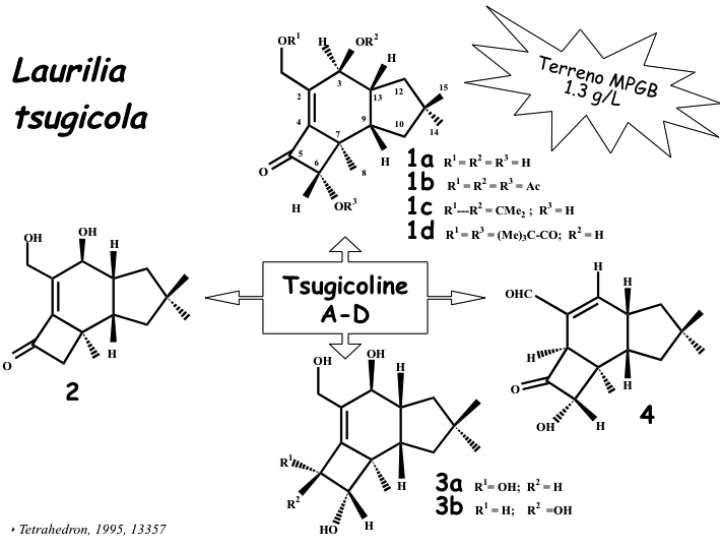

Laurilia tsugicola
Terreno MPGB 1.3 g/L
1a R1 = R2 = R3 = H
1b R1 = R2 = R3 = Ac
1c R1---R2 = CMe2 ; R3 = H
1d R1 = R3 = (Me)3C-CO; R2 = H
Tsugicoline A-D
2
4
3a R1= OH; R2 = H
3b R1 = H; R2 =OH
Tetrahedron, 1995, 13357

Fig. 4.

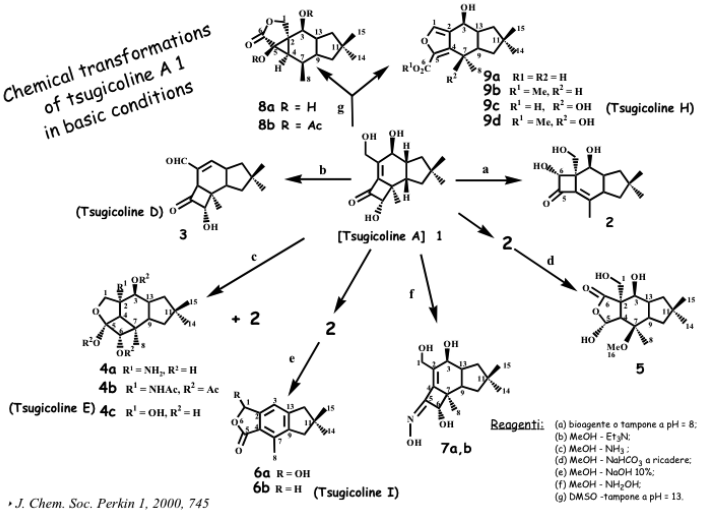

Chemical transformations of tsugicoline A 1 in basic conditions
8a R = H
8b R = Ac
9a R1 = R2 = H
9b R1 = Me, R2 = H
9c R1 = H, R2 = OH (Tsugicoline H)
9d R1 = Me, R2 = OH
(Tsugicoline D)
3
[Tsugicoline A] 1
2
4a R1 = NH2, R2 = H
4b R1 = NHAc, R2 = Ac
(Tsugicoline E) 4c R1 = OH, R2 = H
6a R = OH
6b R = H (Tsugicoline I)
7a,b
5
Reagenti:
(a) bioagente o tampone a pH = 8;
(b) MeOH - Et3N;
(c) MeOH - NH3 ;
(d) MeOH - NaHCO3 a ricadere;
(e) MeOH - NaOH 10%;
(f) MeOH - NH2OH;
(g) DMSO -tampone a pH = 13.
J. Chem. Soc. Perkin 1, 2000, 745

Fig. 5.

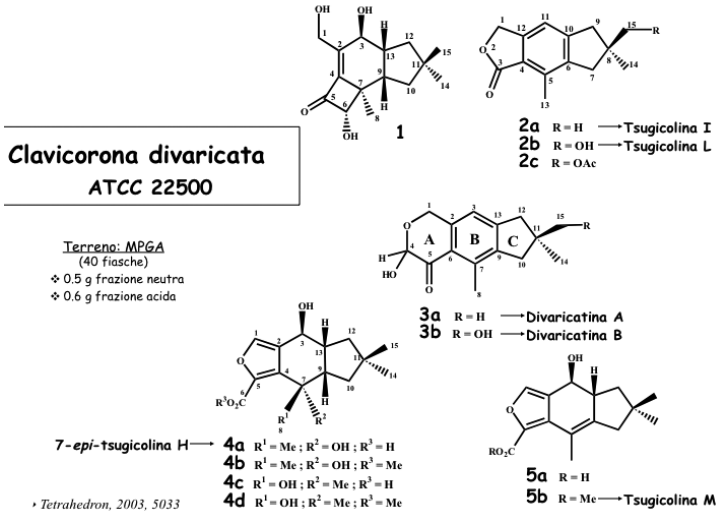


Fig. 6.

3. GENERE: *CLITOCYBE ; SPOROTRICHUM*

La forte tossicità della illudina M 1 naturale (*Clitocybe illudens*) porta ad escludere il suo uso come agente antitumorale; allo scopo di ridurre la sua tossicità che si ritiene sia correlata alla capacità del doppio legame coniugato di dare reazioni di tipo Michael, il metabolita è stato fatto reagire con diazometano ottenendo i prodotti 2-4 mentre per ossidazione con m-CPBA è stato ottenuto il composto 5a, questi sono risultati meno attivi del loro precursore (*Fig. 7*). Nella tabella sono mostrate le tossicità di questi derivati se testati sulla linea cellulare A 2780.[6]

Altri interessanti metaboliti, i clavilattoni A-E sono stati isolati da colture di *Clitocybe clavipes*; il clavilattone D ha mostrato inibire la proteina tyrosine kinasi.[7]

La spirolaxina 1, isolata dal fungo *Sporotrichum laxum* (*Figg. 8,9*), è un metabolita contenente un nucleo della 3-metossi-5-idrossi-ftalide legato con una catena metilenica ad un gruppo spiroacetalico; la sua stereochimica è stata determinata mediante cristallografia a raggi X e da misure di CD. La spirolaxina mostra attività sul batterio *Helicobacter*

pylori, responsabile di ulcere gastriche; insieme al suo co-metabolita sporotricale 2 risulta anche essere un potente inibitore dell'angiogenesi. L'attività della spirolaxina viene rilevata *in vitro* su cellule endoteliali e su di una varieta di linee cellulari tumorali. Si è dimostrato che ad una dose non tossica (1-0.1 μM), inibisce la crescita delle cellule endoteliali da microcircolo di surrene bovino (BMEC) e delle cellule endoteliali umane (UVEC); dopo 24-48 h con dosi da 1 a 40 μM la molecola determina inibizione della proliferazione sulle linee tumorali MESSA (sarcoma uterino), LoVo e HL60. La sintesi della (+)-spirolaxina metiletere chiude la ricerca su questo interessante metabolita.[8,9]

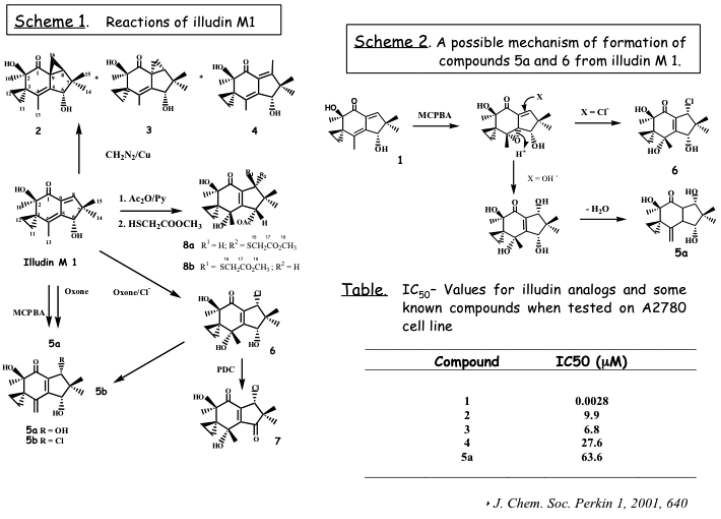


Fig. 7.

4. I Clavilattoni

I clavilattoni A-E sono metaboliti prodotti dal Basidiomicete *Clitocybe clavipes* su colture agarizzate (MPGA); questi metaboliti hanno attività antibatterica ed antifungina ed inoltre inibiscono la crescita di *Lepidium sativum*.

Il fungo su di un diverso terreno (GMYA) produce il clavilattone D un metabolita rosso la cui struttura è stata determinata mediante studi ^{1}H e ^{13}C NMR per identificare soprattutto la posizione del gruppo ossidrilico aromatico. Il CD ha mostrato attività inibitrice 5 volte superiore ai CA e CB dell'enzima Tyrosin-Kinase che partecipa alla trasmissione dei segnali di controllo della crescita; poiché diverse Tyr-Kinases sono il prodotto di protoncogeni, inibitori di queste molecole possono rilevarsi degli importanti agenti anti-proliferativi.

La biogenesi dei clavilattoni può essere razionalizzata partendo dal geranil-idrochinone attraverso numerosi passaggi ossidativi (*Fig. 10*).

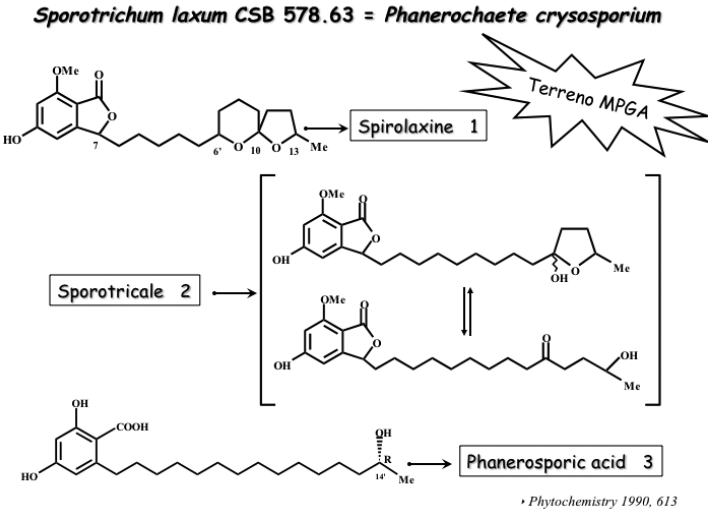


Fig. 8.

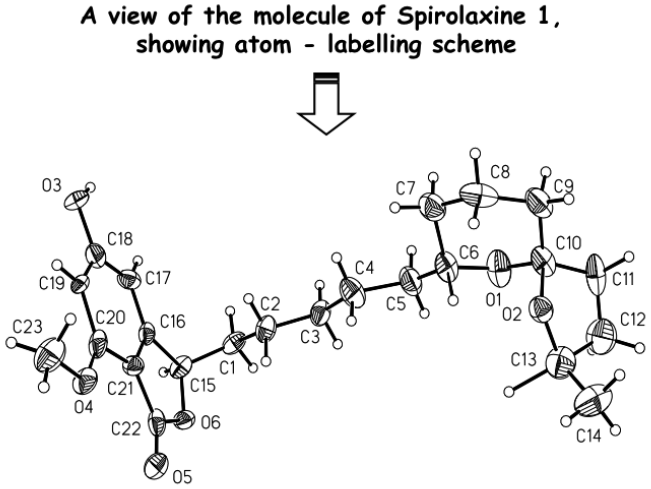
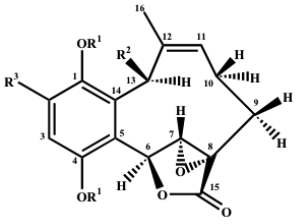


Fig. 9.



CA : $R^1 = R^2 = R^3 = H$
CB : 1,4-dioxo, $R^2 = R^3 = H$
CC : $R^1 = R^3 = H$, $R^2 = OH$
CD : 1,4-dioxo, $R^2 = H$, $R^3 = OH$
CE : $R^1 = R^3 = H$, $R^2 = OMe$

- *J. Chem. Soc. Perkin 1, 1994, 2165*
- *Phytochemistry, 2000, 1039*
- *Biochemical Pharmacology, 2000, 1539*

Fig. 10.

BIBLIOGRAFIA

[1] A. Arnone, R. Cardillo, G. Nasini, J.Chem.Soc. Perkin Trans.1, 1988, 503.
[2] A. Arnone, G. Nasini, O. Vajna de Pava, G. Assante, J.Chem.Soc. Perkin Trans.1, 1992, 615.
[3] A. Arnone, U. Brambilla, G. Nasini, O. Vajna de Pava, Tetrahedron, 1995, 48, 13357.
[4] A. Arnone, C. De Gregorio, A. Mele, G. Nasini, O. Vajna de Pava, J.Chem.Soc. Perkin Trans,1, 2000, 745.
[5] A. Arnone, G. Candiani, G. Nasini, R. Sinisi, Tetrahedron, 2003, 59, 5033.
[6] A. Arnone, L. Merlini, G. Nasini, O. Vajna de Pava, F. Zunino, J.Chem.Soc. Perkin Trans.1, 2001, 610.
[7] L. Merlini, G. Nasini, L. Scaglioni, G. Cassinelli, C. Lanzi, Phytochemistry, 2000, 53, 1039.
[8] A. Bava, M. Clericuzio, G. Giannini, L. Malpezzi, S.V. Meille, G. Nasini, Eur.J. Org.Chem. 2005, 2292.
[9] R. Nannei, S. Dallavalle, L. Merlini, A. Bava, G. Nasini, J.Org.Chem.2006, 71, 6277, e lett. citata.

CARBON ATOMIC WIRES: FROM STARS TO NANOTECHNOLOGY

CARLO S. CASARI (*)

Nota presentata dal s.c. Carlo Bottani
(Adunanza del 2 febbraio 2012)

SUNTO. – Il carbonio può presentarsi in una grande varietà di sistemi e nanostrutture di diversa dimensionalità come nel caso di fullereni, nano tubi e grafene, tutti con proprietà specifiche di grande interesse per la scienza di base e per quella applicata. In questo contesto atomi di carbonio con ibridazione di tipo *sp* possono aggregarsi per formare fili atomici le cui proprietà risultano fortemente dipendenti dalle dimensioni. La ricerca del carbonio *sp* ha una lunga storia che ha toccato anche l'astrofisica e i minerali trovati in crateri da impatto meteoritico e oggigiorno, anche se strutture a base di carbonio sp sono tuttora generalmente considerati sfuggenti e instabili, diversi sistemi nano-strutturati interamente costitutiti da (o contenenti) fili atomici di carbonio vengono prodotti e studiati. Così come il grafene viene considerato il materiale più sottile mai ottenuto (spessore di un atomo) i fili atomici di carbonio rappresentano veri sistemi mono-dimensionali (diametro di un atomo) con grandi potenzialità per le nanoscienze e le nanotecnologie. In questa Nota si discute lo stato attuale della ricerca sui fili atomici di carbonio, partendo dalla descrizione dei sistemi ideali come termine di paragone per strutture reali, allo scopo di mostrare come le proprietà possano essere modulate controllando la lunghezza (cioè il numero di atomi di carbonio) e il tipo di terminazione (atomi o gruppi molecolari o nanostrutture). Vengono presentate le principali tecniche di sintesi e le strategie per ottenere strutture stabili. Particolare enfasi viene messa nell'uso delle spettroscopie vibrazionali per ottenere informazioni sulle proprietà strutturali ed elettroniche, infine, vengono discusse le prospettive aperte per lo sviluppo di dispositivi innovativi basati sull'utilizzo di questi sistemi.

***

(*) Dipartimento di Energia, Politecnico di Milano, Italy.
E-mail: carlo.casari@polimi.it

ABSTRACT. – Carbon can produce a wide variety of systems and nanostructures with different dimensionality as exemplified by fullerenes, nanotubes and graphene, all with peculiar properties of great interest for fundamental and applied research. In this framework *sp*-hybridized carbon atoms can arrange to form atomic wires with size dependent functional properties. The *sp* form of carbon has a long story passing through astrophysics and minerals found in meteor craters and, although still considered elusive and unstable, nowadays a number of nano-systems constituted by or containing sp-carbon wires have been produced and characterized. As graphene is considered the thinnest material (one-atom thick), carbon atomic wires represent true 1-dimensional systems (one- atom diameter) with a great potential for nanoscience and nanotechnology. The status of research on carbon atomic wires is here discussed, starting from the description of ideal systems to real structures, showing that the properties can be tuned by controlling the wire length (*i.e.* number of carbon atoms) and termination (atom, molecular group or nanostructure). Synthesis techniques are presented as well as strategies to have stable wires. A particular attention will be given to the use of vibrational spectroscopy to provide insight on the structure and electronic properties of these systems. Perspectives for novel devices based on the exploitation of the electronic properties of these systems are also discussed.

1. INTRODUCTION

Carbon is at the origin of life, it is a versatile element since it can arrange its valence electrons in different ways to produce a wide variety of molecules and structures. This is clear in organic chemistry where the three possible sp^3, sp^2, and sp^1 hybridizations lead to alkanes (*e.g.* methane), alkenes (*e.g.* ethylene) and alkynes (*e.g.* acetylene) molecules, respectively. The different geometry of the hybridized orbitals results in linear (1-dimensional), planar (2-dimensional) or tetrahedral (3-dimensional) spatial organization. In astrophysics the presence of short linear carbon molecules has been detected in the interstellar medium and experimental studies to understand their formation drove the discovery of fullerenes. Moving to solid state systems, carbon shows only two allotropes: graphite and diamond corresponding to sp^2 and sp^3 hybridization. The lack of a third crystalline form based on sp^1 hybridization (linear carbon or carbyne) has attracted the attention of scientist for the last two centuries. The discovery of this new carbon was claimed in mineral samples found in a meteor crater, nevertheless the existence of such a new pure carbon solid has still today to be confirmed and the efforts in this direction have been progressively reduced. In this context the advent of nanotechnology and the great impact of the discovery of carbon nanostructures such as fullerenes, nanotubes and graphene have

raised a renewed interest in nanoscale carbon systems with linear structures.[1] Carbon nanostructures can be classified according to their hybridization, as shown in *Fig. 1*. Graphene, one isolated honeycomb layer of graphite, is a 2-dimensional structure that can be placed at the sp^2 vertex. Nanotubes and fullerenes are curved structures that can be put in the pathway from sp^2 to sp^3 where the sp^3 character increases for smaller structures (the fullerene C_{20} has no hexagons in its structure). Finally ending at the sp^3 vertex with nanodiamond. Moving from sp^2 to sp^1 we can find a mixed sp^2-sp^1 planar structure (graphyne), ring structures and linear 1-dimensional atomic chains. Approaching sp^1 from sp^3 we find hypothetical possible mixed systems sometimes called superdiamond or sp-sp^3-yne-diamond.

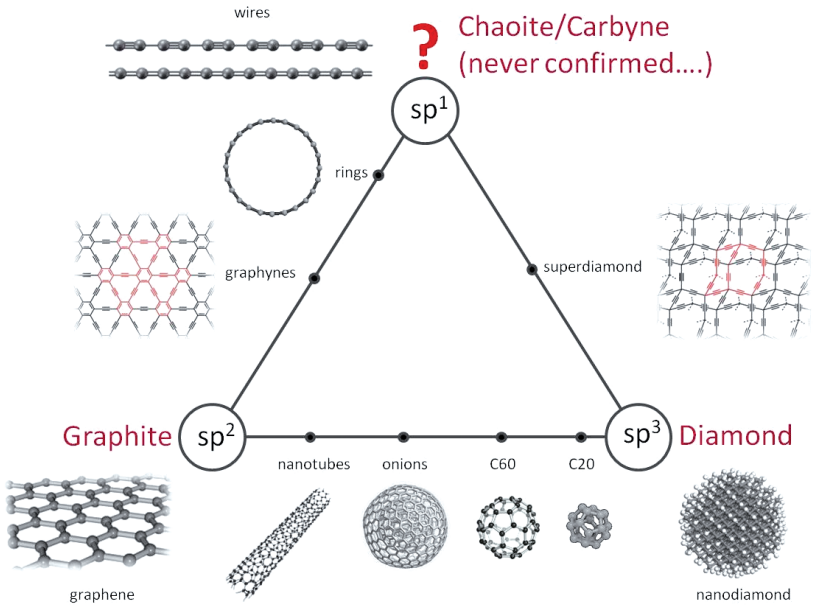


Fig. 1 – Three phase diagram showing possible classification of carbon nanostructures according to the hybridization state of carbon atoms.

Carbon atomic wires possess a large conjugation because of the two p orbitals provided by each atom. This gives rise to a strong structure-property relationship that makes these systems very appealing for

fundamental and applied research. Ideal wires are infinite long systems and present only two possible geometric configurations: an alternated structure with single and triple bonds, also called polyyne and an equalized structure with a sequence of double bonds, also named cumulene (see *Fig. 2*). As discussed in section 3, cumulene is a 1-dimensional metal while polyyne is a semiconductor. In real systems the wire has a finite length and a defined termination. As a consequence the electronic (metal or semiconductor) and optical properties (optical absorption) are determined by the size and the type of termination. The control of wire structure can open the way to novel functional systems to be used as building blocks for advanced materials with tunable functionality. Of course, one needs to know how to fabricate wires, how to characterize their structure and functional properties before being able to make devices and there are still some important challenges to this goal. Here the present status of research about carbon atomic wires is presented to discuss the potentialities of these systems as advanced materials for future technological applications.

2. A BRIEF HISTORY OF SP-HYBRIDIZED CARBON

Graphite and diamond are known since the ancient time, but it was in 1796 that S. Tennant experimentally demonstrated that diamond and graphite were made of the same substance.[2] At the end of nineteenth's century some works were made trying to synthesize linear or polymeric forms of carbon by Glaser in 1870 and Baeyer in 1885. As a consequence of Baeyer's theories stating the impossibility to produce linear carbon the research on sp-carbon suffered a strong damping.

A new impulse was given in the sixties by claims of discovery of a natural form (a new allotrope) of carbon with linear structure. The first paper claiming this discovery (named carbyne, while the mineral form was called chaoite) dates back to the sixties when El Goresy and Donnay while analyzing by x-ray diffraction graphite-rich material coming from a meteor impact crater (the Noerdlinger Ries crater in Germany) found a primitive cell size different from graphite and other carbon-based materials.[3] Other authors started to investigate the conditions for the formation of this material and the related position in the phase diagram of carbon.[4,5] In the same period Russian scientists reported the synthesis of carbyne . Kudryavtsev and coworkers registered the discovery

of carbyne (in russian, see *e.g.* [6]), a sort of patent issued by the USSR Committee of Inventions and Discoveries. In the eighties strong criticisms were raised for example by Smith and Buseck about the interpretation of the diffraction results and the existence of carbyne was object of intense debates, as witnessed by the papers appeared on Science with titles like 'Carbyne: do they really exist?'.[7-9] In the meantime the search for sp carbon linear clusters to understand some spectral features observed in the interstellar medium was a hot topic in astrophysics. Researcher were trying to reproduce in laboratory the conditions for carbon cluster formation in the stars, including Sir H. Kroto at the University of Sussex in UK. He was trying to synthesize HC_5N and longer polyyne-like systems by laser vaporization source when in 1985, together with Richard Smalley and Robert Curl, observed the presence of large clusters made by 60 carbon atoms (20 hexagons and 12 pentagons forming a soccer ball). In 1996 Kroto, Smalley and Curl were awarded with the Nobel Prize in Chemistry 'for their discovery of fullerenes'; the story of the discovery is nicely reported in the Nobel Lecture given by Kroto.[10]

In 1995 R. Lagow and coworkers reported the synthesis of 'liner acetylenic carbon' up to 28 carbon atoms.[11] They optimistically stated that systems of more than 300 atoms were potentially possible to be synthesized. At present, the occurrence of sp-hybridized carbon has been observed in many carbon-based materials and structures, embedded in cold gas matrices,[12] in free carbon clusters in the gas phase,[13,14] in pure sp-sp^2 systems,[15] in liquids,[16] inside carbon nanotubes (CNTs) [17] and connecting graphene sheets.[18] Nowadays, polyynes in liquids (even number of carbon atoms up to 20) are synthesized even as size-selected samples [19-21] and with well defined end groups.[22] Solid state samples have been also produced [23] and recently R.R. Tykwinsky and co-workers reported the synthesis of chains up to 44 carbon atoms terminated by bulky groups.[24] The latter system is stable in air and at room temperature in the form of a solid sample. Tykwinski reported also the synthesis of long cumulenes [25] even though they seem more difficult to produce and a few works are present reporting their observation. By playing with termination-induced electronic arrangement, short cumulenic structures can be produced, as reported by Cataldo [26] and extensively discussed in the review by Cadierno *et al.*.[27] Both cumulenes and polyynes were detected in a pure sp-sp^2 cluster-assembled amorphous carbon sys-

tem.[15,28-30] More recently reversible transformation of polyyne into cumulenes under UV irradiation has been also reported.[31]

3. Structure and properties of carbon wires

The ideal model of carbon atomic wire is an infinite chain of carbon atoms. In this case there are only two ways for carbon atoms to arrange in the wire: a sequence of double bonds (cumulene) and single and triple alternated bonds (polyyne), see *Fig. 2.* Cumulenes have one atom per unit cell and each contributes with two electrons, one for each π-orbital. This results in a half-filled conduction band and the system is a 1-dimensional metal. On the contrary, polyynes have two atoms per unit cell thus filling completely the valence band and leaving an empty conduction band. As a result polyyne is a 1-dimensional semiconductor. The same approach applies also for the vibrational properties (*i.e.* phonons). Cumulene is an example of a 1-dimensional homo-atomic chain usually used to describe the vibrational properties of more complex systems while polyyne is a homo-atomic chain with different bond lengths (and strength) that is the analogue of a hetero-atomic chain. Simple textbook calculations show that cumulene has only acoustic phonons while polyynes have acoustic and optic phonon branches. More detailed calculations lead to corrections in the optical phonon frequencies at the center of the Brillouin zone when conjugation effects are considered.

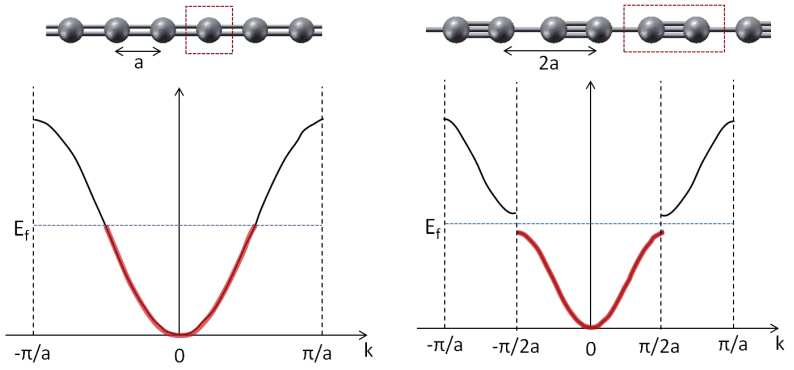


Fig. 2 – Schematic electronic band structure for infinite cumulene and polyyne showing the metal and semiconductive behavior.

The link between the two possible wire structures is the onset of Peierl's distorsion, a theoretical prediction made by Peierl for 1-D metals.[32] According to Peierl a 1-D metal can undergo a metal-to insulator transition since a dimerized chain structure is energetically more favorable than an equalized one. The alternated bond structure (dimerized) is the result of a balance between the elastic strain introduced to distort the chain and the gain obtained by lowering the energy of electrons. In the case of carbon wires such instability would mean that cumulenes are unstable and tend to rearrange in polyyne form. This actually occurs only for the infinite wire while for real systems the role of Peierls distorsion is less important as discussed in the following.

The distinction of wires in the two discussed configurations, alternated and equalized chains, strictly holds only for ideal infinite systems, while for real finite wires the distinction between cumulenic and polyynic configuration is somehow relaxed since in both cases approaching the ends the bond lengths can change due to a different termination. The termination has usually a different bond length and this affects the entire wire structure so that the shorter the wire the larger is the terminating effect. A distribution of bond lengths along the wire is typically found when simulating the structure of finite and terminated wires. In this case the bond length alternation (BLA, the length difference between two adjacent bonds) is a more correct parameter to describe the wire structure. The ideal cumulene has BLA=0 while simulations of the structure of real and finite cumulene-like systems show that the BLA is small but not zero. *Fig. 3* shows a comparison of the BLA evaluated from first principles calculations for polyynes and cumulenes as a function of the wire length (*i.e.* number of carbon atoms).

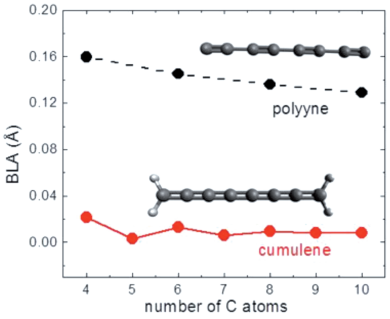


Fig. 3 – Dependence of bond length alternation (BLA) values for finite-length polyynes and cumulenes as a function of the number of carbon atoms. The values are obtained from first principles calculations.[47]

In the same way the energy gap and the optical and transport properties depend not only on the type of structure but also on chain length (*i.e.* the number of carbon atoms) and type of end groups.[33-35] The electronic bands show a gap, which depends on the BLA value.[33] Although the electronic properties have been theoretically addressed by many authors, the experimental evidences are scarce. Only a few works reported the study of electronic core levels and the valence band properties by electron spectroscopy techniques.[13,36,37]

Vibrational properties show a similar dependence in the phonon dispersion relation. In particular the frequency of the longitudinal optical phonons at the center of the Brillouin zone scales from 1800 to more than 2200 cm^{-1}, depending on the wire structure, thus suggesting why vibrational spectroscopy techniques are extremely useful for the characterization of sp-carbon wires, as discussed in section 5.

4. Synhthesis techniques

Today, many techniques, both physical and chemical, have been demonstrated to produce sp carbon wires in different forms, mainly in a bottom-up approach.[19] Physical techniques are mainly based on the rapid quenching of a carbon vapour in different environments (see *Fig. 4*). Supersonic carbon cluster sources based on the arc discharge between graphite electrodes (*i.e.* pulsed microplasma cluster source – PMCS developed by P. Milani) demonstrated the capability to produce sp-rich amorphous carbon films (estimated to be up to 40%).[28] A similar approach was exploited using thermal or laser vaporization cluster sources.[12,38] Sp carbon has been also produced by ion irradiation of amorphous carbon [39] and by femtosecond (fs) laser irradiaton of a graphite target.[40] fs laser pulses have been used to produce amorphous carbon films containing sp, sp^2 and sp^3 phase.[41] For all the above-mentioned techniques sp carbon is present in amorphous (mainly sp^2 carbon moieties) and a fine control of the chain structure and length is still to be achieved. Isolated wires can be produced by laser ablation (with both fs and ns pulses) of carbon solid targets (graphite) or suspensions (fullerenes, diamond particles) in liquids and particularly in the polyyne-like form with an even number of carbon atoms.[16, 40] With reference to polyynes in solution an easy-to-use and cost-effective technique is the submerged arc discharge in liquids developed by F. Cataldo.[42] This technique allows also controlling, to some

extent, the chain termination by selecting suitable solvents.[43] As an example short cumulenic structures terminated by sulfur atoms have been produce by arcing in carbon sulfide.[26]

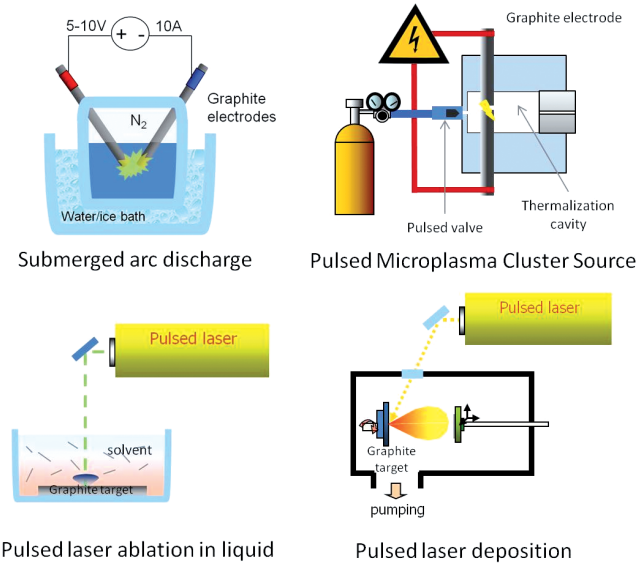


Fig. 4 – Examples of physical techniques for the synthesis of carbon atomic wires

Different chemical techniques have been used to synthesize a large number of sp carbon chains terminated with different molecular groups as reported in the review by Gladysz *et al.*.[22] Among the most employed chemical routes we cite the dehydropolycondensation of acetylene such as the Glaser reaction based on the oxidative coupling reaction of ethynyl groups by copper salts; polycondensation reactions of halides; dehydrohalogenation of polymers such as the chemical carbonization of poly(vinylidene halides) (PVDH). Tykwinski reported the synthesis of very long and stable polyynes and cumulenes.[25]

Top-down methods start from a different carbon structure that can be turned into a wire by means of mechanical stress, electric field or removal of selected atoms. One feasible possibility is represented by electron bombardment of a single graphene flake in a transmission electron microscope (TEM) thus allowing to selectively remove carbon atoms until a single atomic chain is connecting two separate graphene edges.[18] However the method is extremely time-consuming and requires the elec-

tron beam and the image capabilities of a TEM. Such approach allows to obtain very nice TEM images of wires suspended between graphene edges even though the unambiguous identification of the chain structure was not possible. Some of the systems that can be produced by these techniques are reported in *Fig. 5*.

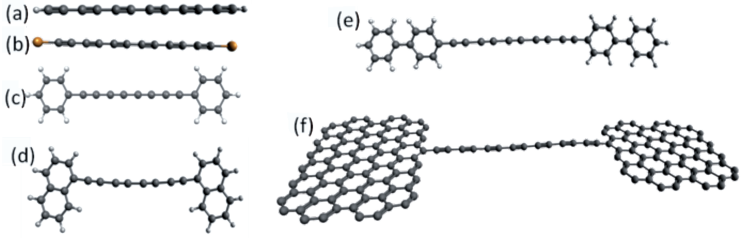


Fig. 5 – Structure of real wires with different terminations, from top to bottom (a to f) polyyne terminated by hydrogen, copper, phenyl, bis-phenyl, naphtyl and graphene.

5. Characterization techniques

Vibrations in molecules and solids are strongly dependent on the atomic arrangement and chemical bond. In carbon nanostructures the local atomic organization can be very different due to the hybridization state and local order, hence a vibrational technique sensitive to the local vibrations can provide a wealth of structural and chemical information. Raman spectroscopy is a technique based on inelastic scattering of light by the thermally excited vibrations (*i.e.* phonons in quantum solid state approach) of the material. The technique is based on the analysis of the spectrum of an incident monochromatic light (laser light with a given wavelength) after the interaction with a sample. The outcoming light can have wavelengths higher and lower than the incident one. In a quantum particle description incident photons are inelastically scattered by phonons and come out from the material with a different energy.[44] The difference equals the energy exchanged in the process in creating or annihilating a phonon. In this way the spectrum contains information on the phonon energy that can be a fingerprint of some characteristics of the investigated material. Raman turns out to be extremely useful to investigate carbon-based materials and nanostructures as exemplified in *Fig. 6*. Different types of carbon nanostructures are characterized by Raman signatures that can be used to investigate the structural properties, a chief example being

graphene for which the number of layers can be unambiguously detected by inspecting the Raman spectrum.[45,46] Sp carbon shows Raman peaks in a specific spectral region (1800-2200 cm^{-1}) where no other peaks from sp^2 carbon (1300-1600 cm^{-1} and 2700-3200 cm^{-1}) or from diamond (1332 cm^{-1}) are present, thus permitting to identify its presence, even when other carbon phases coexist, as in the case of sp-sp^2 carbon moieties. The Raman peak is associated to the carbon-carbon stretching mode and polyynes are expected to have higher Raman frequencies than cumulenes, permitting to be sensitive to the structural arrangement.[47] In addition the peak frequency decreases with increasing the chain length due to modulation of the BLA with the wire length and approaches the value of the ideal structure. In addition effects of chain conformations such as chain bending, torsions and different arrangement of the terminating groups and to what extent they affect the Raman spectrum have been studied both experimentally and theoretically.[21,48,49] This allows to get a lot of information by a detailed inspection of the Raman spectrum also with the support of theoretical simulations. As shown in *Fig. 6* different types of polyynes display specific Raman spectra depending on wire length, termination and local order. For instance, from the analysis of the spectrum of phenyl-terminated polyynes it is possible to evidence the vibration of the phenyl ring (at about 1600 cm^{-1}) and to associate the different peaks in the 2000-2200 cm^{-1} region to the contribution of wires of different length from 4 to 12 carbon atoms.[48]

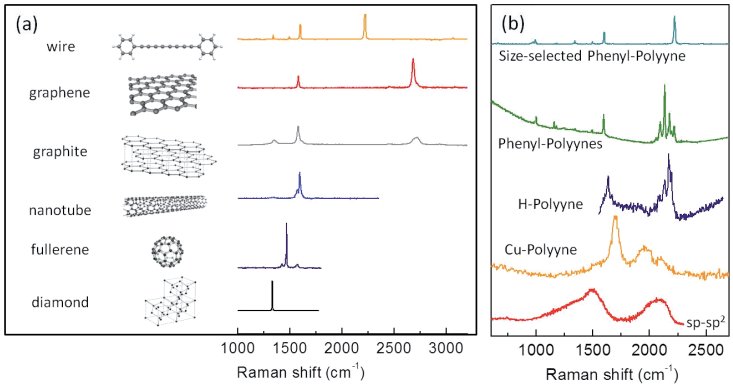


Fig. 6 – (a) Raman spectra of graphite, diamond and carbon nanostructures (fullerene C60, single wall carbon nanotubes, graphene and phenyl-polyynes). (b) Raman spectra of different type of wires.

Stability can be easily studied by Raman spectroscopy as one of the major problems arising when dealing with sp carbon wires. A part from systems with suitable terminations, the majority of sp carbon suffers from a poor stability. What typically happens is a sp-to-sp^2 transition due to cross-linking reactions. To this problem, a suitable strategy is the use of bulky terminating groups able to prevent cross-linking due to steric hindrance. An alternative route to the direct synthesis of stable structures is the stabilization of already formed wires. For instance H-terminated polyynes can be embedded in a solid assembly of Ag nanoparticles, resulting in a sample stable for several weeks at room temperature and in air conditions. The stabilization steps can be followed by Raman spectroscopy, as shown in *Fig. 7*.[50] More recently Hayashi and co-workers produced a polymeric composite (*i.e.* poly(vinyl alcohol)) containing polyynes stabilized by Ag nanoparticles.[51]

The stability has been also studied, in particular the cumulene/polyyne stability, the sp-to-sp^2 transition in different conditions and the variation of length distribution when interacting with metal nanoparticles.[28,29,52,53]

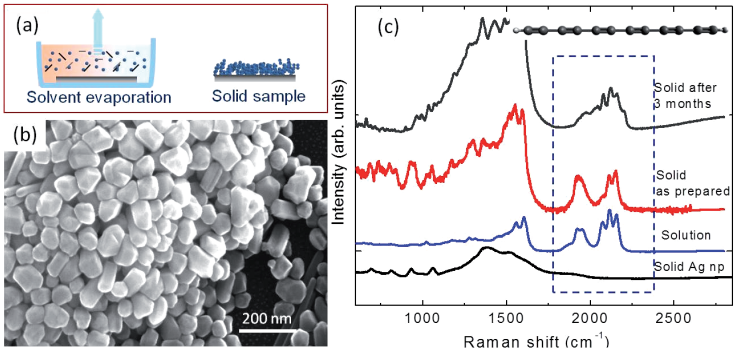


Fig. 7 – (a) Schematic process of stabilization of H-polyynes. (b) Scanning electron microscope image of the silver nanoparticle assembly and (c) Raman spectra of the solid silver assembly alone, the solution of mixed polyynes and nanoaprticles, the solid assembly with polyynes as prepared and the same sample after 3 months showing that the signal from sp carbon is still present.

Actually, the interaction with metal nanoparticles can be exploited to increase the Raman signal by means of surface enhanced Raman scattering (SERS). Metal nanoparticles have collective excitations of the surface conduction electrons (*i.e.* surface plasmons) that depend on the par-

ticle size. If the surface plasmon absorption matches the incident laser wavelength used for Raman, the high electromagnetic field close to the particle can enhance the Raman signal for order of magnitudes. SERS is usually employed to detect very small amount of sample, in the best conditions single molecule spectroscopy has been reported.[54] Two are the mechanisms involved in the SERS process: electromagnetic and chemical effect. The former is the due to the local enhancement of the electric field while the latter is active when some chemical interaction of the sample with the SERS active medium (*i.e.* nanoparticles) takes place. Carbon atomic wires in solution can be studied by SERS, the procedure is based on mixing the sample solution (H-terminated polyynes) with noble metal nanoparticles in solution (*e.g.* silver or gold colloids) as shown in *Fig. 8*. Silver colloids with diameters in the 20-40 nm range show an absorption peak (surface Plasmon) at about 420 nm while after mixing the Plasmon absorption is extended down to the infrared region (see *Fig. 8b,e*). This is an indication of a chemical effect in which polyynes aggregate when interacting with metal nanoparticles and allows to perform SERS with different excitation wavelengths.[52] The SERS effect is shown in *Fig. 8c*, where the comparison between Raman and SERS indicates an enhancement factor of about one million times.

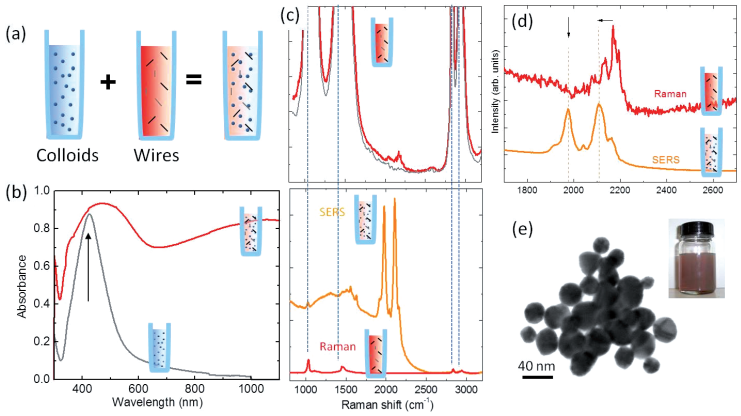


Fig. 8 – Raman and SERS of H-terminated polyynes. (a) Scheme of SERS experiment; (b) UV-vis absorption spectra of silver nanoparticles; (c) Raman and SERS of polyynes: Raman spectrum compared with the spectrum of the solvent alone (top) and SERS compared with Raman where the peaks from the solvent are used for normalization (bottom); (d) particular of the SERS and Raman spectra in the sp-carbon spectral region and (e) scanning electron microscopy image of Au nanoparticles as a SERS active medium.

A closer inspection to the SERS and Raman spectra reveals that the two spectra are not the same. In addition to the strong enhancement it is possible to observe a redshift of the main Raman peak together with the appearance of a new feature at lower wavenumbers (below 2000 cm^{-1}). With the support of numerical calculations it is possible to explain such changes by a charge transfer from the nanoparticle to the wire. One electron moves to the wire and this induces a rearrangement of the structure with a strong reduction of the BLA.[48] The change in the BLA is responsible for the modification of the SERS spectrum with respect to the Raman one. Wire modification due to charge transfer is an interesting effect that can be used to tune the properties of the wires. In fact reduction of the BLA means a tendency towards a cumulene-like structure driven by interaction with noble metal particles. Such mechanism was observed not only in H-terminated polyynes but also in other systems such as phenyl-polyynes thus opening a way for the external tuning of wire structure, without the need to change the termination to induce cumulene-like character.

6. Towards applications

Carbon atomic wires have a great potential interest for technological applications. The possibility to tune the properties opens the way to novel functional systems to be used as building blocks for advanced materials and devices. Wires may be employed in a solution-processable approach with low cost and environmental friendly techniques. In this context the availability of metallic, semiconducting and insulating properties in the same system and the possibility to tune the energy and optical gap can overcome some problems in molecular and flexible electronics.

Appealing electron transport properties are expected for these wires as outlined by a number of theoretical investigations.[55,56] A pioneer work has been done by Lang and Avouris who computed the conductance of a wire connected to metal electrodes and outlined an oscillatory behavior depending on the even or odd number of carbon atoms.[57,58] In particular, cumulenes as 1-D metals are expected to display ballistic conductance, a process where electrons move without scattering, that results in a constant conductance, not depending on

the wire length, at the value of $2G_0$ where $G_0 = 2e^2/h = 77\ \mu S$ (equivalent to a resistance of about *12 kΩ)* is the quantum of conductance.[59] A very few papers recently reported the measurement of the conductivity of single wires. O. Cretu *et al.* measured the conductivity of a wire connecting two graphene flakes in the transmission electron microscope used for the top-down fabrication of the wire, as shown in *Fig. 9a-c*. Other groups used polyynes suspended between a metal surface and the metal tip of a scanning tunneling microscope as the contact electrodes. It was observed a larger resistance than the expected value, this is possibly due to the contact resistance between the wire and the electrodes.[60,61] Other considered systems include wires connecting graphene sheets and carbon nanotubes. In the latter case Yuzvinsky *et al.* reported the observation of a non monotonic current-voltage (I-V) curve with an apparent negative differential resistance (*i.e.* the current decreases when increasing the applied voltage).[62] Such behavior can be explained by invoking the possibility of a resonant tunneling behavior in which electrons go through the contacts by means of a quantum tunneling process. A memory device has been proposed based on the formation and disruption of a wire across fractured graphene which results is a change in the conductivity, as shown in *Fig. 9d-f*.[63] The wire formation and disruption process is repeatable and is controlled through the voltage applied on the two sides of the device. In such a device the reported conductance seems to match very well with a ballistic transport regime showing a characteristic staircase behavior in the I-V curve (see *Fig. 9f*). A similar device was proposed by Li *et al.* even though a nanoelectromechanical mechanism was considered responsible for the change in conductivity and only in a note the authors refer to the possibility of wire formation.[64]

One of the major problems arising in these proposed devices is that the wire structure (cumulene/polyyne) is poorly controlled or even not known. In this framework it has to be noticed that structural rearrangements of polyynes towards a more equalized structure (*i.e.* cumulene-like) occurring upon external perturbations such as UV light irradiation or charge transfers after interaction with metal nanoparticles have been recently observed by IR and surface enhanced Raman spectroscopy (SERS).[31,48] To this regard the manipulation of the structure of wires by external signals (optical and/or electrical) is a new and exciting challenge.

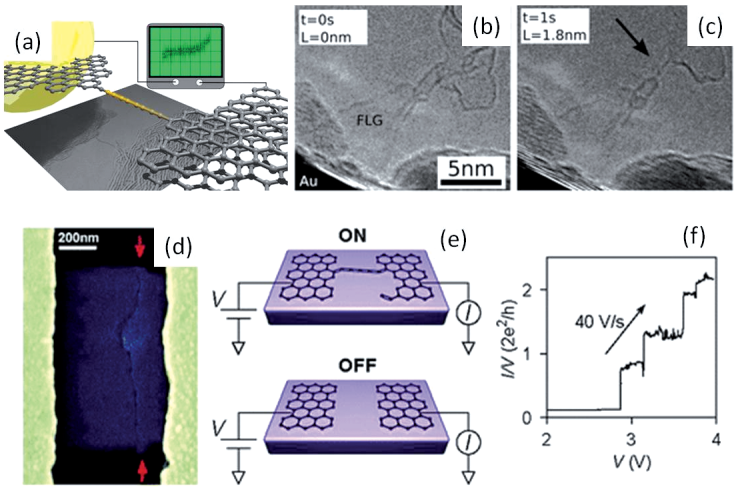


Fig. 9 – (a-c) Electrical measurements of a carbon atomic wire connecting two graphene edges. Reprinted (adapted) with permission from (B. Standley, W. Bao, H. Zhang Nano Letters 8, 3345 (2008)). Copyright (2008) American Chemical Society. (d-f) Memory device based on the formation and disruption of wires across graphene. Reprinted (adapted) with permission from (O. Cretu, A.R. Botello-Mendez, I. Janowska et al. Nano Lett. 13, 3487-3493 (2013)). Copyright (2013) American Chemical Society.

7. Conclusions and perspectives

The science of carbon nanostructures is a wide field embracing physics, chemistry, materials science and nanotechnology. Carbon nanostructures are amongst the largest investigated systems in the last three decades, as witnessed by the literature on fullerenes, nanotubes and graphene. In this framework there is still space for novel systems and allotropes, one possibility is represented by carbon atomic wires, linear 1-dimensional systems made by sp-hybridized carbon atoms, as discussed in this paper. Such systems are appealing for fundamental studies related with the formation of carbon clusters (the so-called fullerene road), with quantum effects in 1-dimensional systems and for applications due to the tunability of their functional properties. Carbon atomic wires, though with some difficulties, can be nowadays synthesized with different techniques in different forms (length and termination), characterization techniques such as Raman spectroscopy allow to

detect the structure of the wires and recently the feasibility of measuring the electrical properties down to the single wire has been demonstrated, together with the realization of nano-scale devices. In this context carbon atomic wires have the potential to play a role as novel carbon nanostructures where a lot of work is still to be done to understand the properties and to synthesize systems with tailored properties. Very recently Yakobson and co-workers have underlined the mechanical properties that are expected, from theoretical predictions, to outperform graphene and nanotubes.[65] Moreover they show that cumulene structures are stabilized by vibrations and mechanical strain plays a key role in inducing a metal-to-insulator transition thus opening new opportunities for electromechanical switching and conductive polymer applications.[66] Theoreticians have been also attracted by hybrid sp-sp^2 systems in the form of 2-dimensional layers, the chief example being graphyne. Simulations have shown that such systems can have very interesting electronic properties such as an electron mobility even higher than graphene as reported by Malko *et al.*.[67] The synthesis of such systems in extended form is still to be developed while the chemical synthesis as large molecules have been reported by Diederich and coworkers.[68] The quest for new carbon forms is still open, researchers have been looking for carbon atomic wires sp-hybridized carbon systems between the stars then to land in the nano-world where atomic wires and novel hybrid allotropes as 2-dimensional sp-sp^2 systems still wait to be studied, developed in detail and maybe applied in advanced technologies.

ACKNOWLEDGEMENTS

C.S. Casari would like to acknowledge C.E. Bottani, G. Zerbi, P. Milani and C. Castiglioni for useful discussion and M. Tommasini, A. Milani, A. Lucotti, V. Russo, A. Li Bassi, L. Ravagnan, C. Lenardi, P. Piseri, F. Siviero, for the work done in collaboration from 2000 to present.

REFERENCES

[1] A. Hirsch. Nature Mat. 9, 868 (2010).
[2] S. Tennant. Phil. Trans. of the Roy. Soc. of London, 87, 123-127(1797).
[3] A. El Goresy and G. Donnay. Science 161, 363 (1968).
[4] A.G. Whittaker, P.L. Kintner. Science 165, 589 (1969).
[5] A.G. Whitteker. Science 200, 763 (1978).
[6] Carbyne and Carbynoid structures, R.B. Heimann, S.E. Evsyukov, L. Kavan Eds, Kluwer 1999.
[7] P.P.K. Smith, P.R. Buseck. Science 216, 984 (1982).
[8] A.G. Whittaker. Science 229, 485 (1985).
[9] P.P.K. Smith, P.R. Buseck. Science 229, 486 (1985).
[10] H. Kroto. Nobel Lecture December 7 1996 (http://www.nobelprize.org)
[11] R.J. Lagow, J.J. Kampa *et al.* Science 267, 362 (1995).
[12] T. Wakabayashi, A.-L. Ong *et al.* J. Phys. Chem. B 108, 3686-3690 (2004).
[13] L. Ravagnan, T. Mazza *et al.* Chemical Communications 47, 2952-4 (2011).
[14] A.A. Zaidi, A. Hu, M.J. Wesolowski *et al.*, Carbon 48, 2517-2520 (2010).
[15] L. Ravagnan, F. Siviero *et al.* Physical Review Letters, 89, 285506 (2002).
[16] H. Tabata, M. Fujii and S. Hayashi. Carbon 44, 522-529 (2006).
[17] X. Zhao, Y. Ando, Y. Liu *et al.* Physical Review Letters 90, 187401 (2003).
[18] C. Jin, H. Lan, L. Peng *et al.* Physical Review Letters 102, 205501 (2009).
[19] Polyynes: Synthesis Properties and Applications, F. Cataldo Ed., Taylor & Francis 2006.
[20] H. Tabata, M. Fujii, S. Hayashi *et al.* Carbon 44, 3168-3176 (2006).
[21] A. Lucotti, M. Tommasini, D. Fazzi *et al.* J. Am. Chem. Soc. 131, 4239-44 (2009).
[22] S. Szafert and J.A. Gladysz. Chemical Reviews 103, 4175-205 (2003).
[23] F. Cataldo, C.S. Casari. J. Inorg. Organ. Pol. and Mater. 17, 641-651 (2007).
[24] W.A. Chalifoux and R.R. Tykwinski. Nature Chemistry 2, 967-71 (2010).
[25] J.A. Januszewski, R.R. Tykwinski. Chem. Soc. Rev. 43, 3184-3203 (2014).
[26] F. Cataldo. J. of Inorg. and Organomet. Polym. and Mater. 16, 15 (2006).
[27] V. Cadierno and J. Gimeno. Chemical Reviews 109, 3512-60 (2009).
[28] L. Ravagnan, P. Piseri, M. Bruzzi *et al.* Physical Review Letters 98, 216103 (2007).
[29] C.S. Casari, A. Li Bassi, L. Ravagnan *et al.* Physical Review B, 69, 075422 (2004).
[30] C.S. Casari, A. Li Bassi, A. Baserga *et al.* Physical Review B 77, 195444 (2008).
[31] M.M. Yildizhan, D. Fazzi *et al.* Journal of Chemical Physics 134, 124512 (2011).
[32] R.E. Peierls, Quantum Theory of Solids (Clarendon, Oxford, 1955), p. 108.
[33] A. Milani, M. Tommasini *et al.* Physical Review B 74, 153418 (2006).
[34] A. Milani, M. Tommasini, G. Zerbi. J. of Chem. Phys. 128, 064501 (2008).
[35] A. Milani, M. Tommasini, G. Zerbi. J. of Raman Spectroscopy 40, 1931-1934 (2009).
[36] L. Zhang, H. Ma, N. Yao *et al.* J. Vac. Sci. Tech. 25, 545 (2007).

[37] A. Hu, S. Griesing, M. Rybachuk *et al.* J. of Applied Physics 102, 074311 (2007).
[38] L. D'Urso, G. Compagnini, O. Puglisi *et al.* J. of Phys. Chem. C 111, 17437 (2007).
[39] G. Compagnini, S. Battiato, O. Puglisi *et al.* Carbon 43, 3025-3028 (2005).
[40] A. Hu, M. Rybachuk, Q.-B. Lu *et al.* Appl. Phys. Lett. 91, 131906 (2007).
[41] A. Hu, Q.-B. Lu, W.W. Duley, M. Rybachuk. J. Chem. Phys. 126, 154705 (2007).
[42] F. Cataldo. Carbon 41, 2671 (2003).
[43] F. Cataldo. Tetrahedron 60, 4265-4274 (2004).
[44] Phonons, Theory and Experiments II, P. Brüesch Springer 1982.
[45] A.C. Ferrari, D.M. Basko. Nature Nanotechnology 8, 235 (2013).
[46] A.C. Ferrari *et al.* Phys. Rev. Lett. 97, 187401 (2006).
[47] F. Innocenti, A. Milani, C. Castiglioni. J. Raman Spectrosc. 41, 226-236 (2010).
[48] A. Milani, A. Lucotti, V. Russo *et al.* J. of Phys. Chem. C, 115(26), 12836 (2011).
[49] L. Ravagnan, N. Manini, E. Cinquanta *et al.* Phys. Rev. Lett. 102, 245502 (2009).
[50] C.S. Casari, V. Russo, A. Li Bassi *et al.* Applied Physics Letters 90, 13111 (2007).
[51] S. Okada, M. Fujii, S. Hayashi. Carbon 49, 4704-4709 (2011).
[52] A. Lucotti, M. Tommasini, M. Del Zoppo *et al.* Chem. Phys. Lett. 417, 78-82 (2006).
[53] A. Lucotti, C.S. Casari, M. Tommasini *et al.* Chem. Phys. Lett. 478, 45-50 (2009).
[54] K. Kneipp, Y. Wang, H. Kneipp *et al.* Phys. Rev. Lett. 78, 1667 (1997).
[55] S. Tongay, R. Senger, S. Dag, and S. Ciraci. Phys. Rev. Lett. 93, 136404 (2004).
[56] Z. Zanolli, G. Onida, J.-C. Charlier. ACS Nano 4, 5174-80 (2010).
[57] N.D. Lang and Ph. Avouris. Phys. Rev. Lett. 81, 3515-3518 (1998).
[58] N.D. Lang and Ph. Avouris. Phys. Rev. Lett. 84, 358 (2000).
[59] O. Cretu, A.R. Botello-Mendez, I. Janowska *et al.* Nano Lett. 13, 3487-3493 (2013).
[60] P. Moreno-García, M. Gulcur *et al.* J. Am. Chem. Soc. 135, 12228-12240 (2013).
[61] F. Chen, X. Li, J. Hihath *et al.* J. Am. Chem. Soc. 128, 15874-15881 (2006).
[62] T.D. Yuzvinsky, W. Mickelson, S. Aloni. Nano Letters 6, 2718-22 (2006).
[63] B. Standley, W. Bao, H. Zhang. Nano Letters 8, 3345 (2008).
[64] Y. Li, A. Ainitskii, J.M. Tour. Nature Materials 7, 966 (2008).
[65] M. Liu *et al.* ACS Nano 7, 10075-10082 (2013).
[66] V.I. Artyukhov *et al.* Nano Lett. 14, 4224-4229 (2014).
[67] D. Malko, C. Neiss. Phys. Rev. Lett. 108, 086804 (2012).
[68] F. Diederich, M. Kivala. Adv. Mater. 22, 803-812 (2010).

Istituto Lombardo (Rend. Scienze) 146, 37-64 (2012)

LE MURA SPAGNOLE DI MILANO: GLI INTERVENTI CONSERVATIVI (2006-2009) E LA LORO NEGAZIONE

REBECCA FANT (*)

Nota presentata dal m.e. Amedeo Bellini
(Adunanza del 1 Marzo 2012)

SUNTO. – In questo lavoro si presenta l'approccio metodologico conservativo applicato all'intervento di restauro, che nasce da un percorso di conoscenza e valorizzazione dei resti del sistema difensivo delle Mura Spagnole. Tale approccio deriva da un riconoscimento del valore testimoniale, nei secoli scorsi negato, e dall'applicazione di indagini e attenzione al dato materiale che hanno indirizzato il progetto esecutivo verso metodologie più corrette. Il vero problema della permanenza materiale del monumento però si pone nell'ottica della "conservazione preventiva e programmata", ovvero della cura costante, del passaggio da una fase di "correzione", prevista con l'intervento di restauro, alla fase di "mantenimento". Questa fase ancora oggi risulta negata.

***

ABSTRACT. – In this paper is presented the methodological approach to the conservation works of the Spanish defensive walls. The approach arise from a path of knowledge and valorisation, deriving its meaning from the witnessing value; this latter, being denied during the past centuries, applies to surveys and pays attention to materials data, which addressed the conservation project towards more correct methodologies. The problem of the monument material persistence applies in the direction of "planned conservation", that is a constant care, passing form a "correction" phase, foreseen with the restoration works, to a "maintenance" phase. Such a phase is currently still denied.

(*) Architetto specialista in Restauro dei Monumenti. Docente a contratto del Politecnico di Milano, Italy.
E-mail: fant@rebeccafantarchitetto.it

1. Introduzione

Perché sono state restaurate le Mura spagnole?

In realtà l'incarico del rilievo in previsione di un intervento di restauro (2001) e la progettazione definitiva per i lavori "di restauro, consolidamento e recupero all'uso delle Mura Spagnole nei tratti di viale Filippetti, Beatrice d'Este, piazza Medaglie D'oro e di viale Caldara" da parte dell'arch. Pietro Ripa (2003) sono stati determinati da "cedimenti della cortina muraria" in via Filippetti che hanno costretto l'Amministrazione Comunale a porre una "cesata protettiva della salute pubblica".[1] Questo progetto definitivo prevedeva un'ipotesi complessiva ripartita in due lotti: il primo lotto riguardante la struttura muraria e il secondo lotto la sistemazione delle aree a verde circostanti. Di questo progetto solamente il primo lotto sarà oggetto del bando di gara per "l'*Affidamento mediante procedura di sponsorizzazione per la progettazione esecutiva e l'esecuzione dei lavori di restauro, consolidamento e recupero all'uso delle Mura Spagnole nei tratti di viale Filippetti, Beatrice d'Este, Piazza Medaglie D'oro, viale Caldara*". Nel marzo del 2006 la società T.M.C. Pubblicità S.r.l., a seguito dell'aggiudicazione della gara, stipulava una Convenzione con la Direzione Centrale Tecnica - Settore Edilizia Patrimoniale e Demaniale del Comune di Milano e fin da subito sono state eseguite le indagini diagnostiche, gli studi e gli approfondimenti per la conoscenza dello stato di conservazione del manufatto, e la campionatura degli interventi, al fine di redigere il progetto esecutivo[2] (*Tab. 1*).

Inoltre, a seguito del montaggio del ponteggio sull'Arco di Porta Romana, era stato indicato nell'autorizzazione del settembre 2006 che

---

1 P. RIPA (a cura di), Fascicolo n. 1, *Relazione storica e descrittiva progettuale dell'intervento. Progetto definitivo. Il restauro, consolidamento e recupero all'uso delle mura spagnole "per un parco della ritrovata cinta difensiva delle mura spagnole"*, marzo 2003, p. 3.

2 Per l'intervento di restauro, argomento del presente contributo, si rimanda alla documentazione di progetto depositata presso gli Enti di controllo: Comune di Milano e Soprintendenza ai Beni Architettonici e del Paesaggio di Milano; unico lavoro approfondito sull'argomento risulta la tesi di Specializzazione di A. MASTRONARDI, *Il progetto di conservazione delle Mura spagnole e di Porta Romana a Milano: aspetti teorici ed operativi,* Scuola di specializzazione in Beni Architettonici e del Paesaggio del Politecnico di Milano, Relatore Prof. A. BELLINI e Correlatore R. FANT, A.A. 2007-2008. Negli anni 2007-2009 sono stati fatti convegni e visite guidate, di cui resta traccia anche in www.milanoneicantieridellarte.it.

"*…l'intervento richiesto sull'Arco da parte della Soprintendenza ai Beni Architettonici e del Paesaggio di Milano va necessariamente attuato*"; quindi è stato richiesto e predisposto anche il progetto definitivo-esecutivo per il *Restauro dell'Arco di Porta Romana,* sito in prossimità del Bastione di piazza Medaglie d'Oro, i cui lavori sono stati realizzati tra il 2008 e il 2009.

INTERVENTO PER LA REALIZZAZIONE, TRAMITE SPONSORIZZAZIONE, DELLA PROGETTAZIONE ESECUTIVA E DELL'ESECUZIONE DEI LAVORI DI RESTAURO, CONSOLIDAMENTO, RECUPERO E VALORIZZAZIONE DELLE MURA SPAGNOLE NEI TRATTI COMPRESI TRA VIALE FILIPPETTI, VIALE BEATRICE D'ESTE, PIAZZA MEDAGLIE D'ORO E VIA CALDARA
Società sponsorizzatrice: TMC Pubblicità S.r.l. – Corso Indipendenza 18 Milano **Comune di Milano,** Direzione Centrale Tecnica, Settore Edilizia Patrimoniale e Demaniale Responsabile del Procedimento: Ing. Armando Lotumolo, Collaboratore: Arch. Corrado De Luca; dal 2006 Arch. Pasquale Mariani Orlandi - Area Tecnica Settore Tecnico Cultura e Beni Comunali Diversi
PROGETTO DEFINITIVO: Arch. Pietro Ripa Progettista interventi di eradicazione, infestanti arboree ed arbustive: Dott. G. Claudio Corazzin Progettista interventi strutturali: Dott. Ing. Enrico Ferrari Ricerca storica: Arch. Francesco Repishti
PROGETTO ESECUTIVO E DIREZIONE LAVORI : ***Intervento di restauro conservativo e consolidamento delle Mura Spagnole nei tratti compresi tra Viale Filippetti, Viale Beatrice D'Este, Piazza Medaglie D'oro, Via Caldara*** Prof. Arch. Amedeo Bellini - Prof. Ing. Arch. Stefano Della Torre, Prof. Arch. Valerio Di Battista - Arch. Paolo Favole - Arch. Lorena Bauce - D.S. Arch. Rebecca Fant Consulenti per gli aspetti biologici: Dott. Alberto Villa – Dott. Marco Realini Indagini diagnostica: C.S.G. Palladio S.r.l. di Vicenza per le analisi microstratigrafiche e chimico fisiche Laboratorio Sperimentale B.E.S.T. – Politecnico di Milano per le indagini termoigrometriche
PROGETTO DEFINITIVO-ESECUTIVO E DIREZIONE LAVORI : ***Intervento di restauro conservativo Arco di Porta Romana e Opere Aggiuntive Mura Spagnole*** Prof. Arch. Amedeo Bellini - Prof. Ing. Arch. Stefano Della Torre, Prof. Arch. Valerio Di Battista - Arch. Paolo Favole - Arch. Lorena Bauce - D.S. Arch. Rebecca Fant Rilievo dell'Arco di Porta Romana: Sat Survey S.r.l. di Mestre (Ve) Indagini diagnostica Arco di Porta Romana: Istituto per la Conservazione e la Valorizzazione dei Beni Culturali – CNR, Responsabile dell'Unità di Milano Dott. Marco Realini
IMPRESA ESECUTRICE: Gasparoli S.r.l di Gallarate (Va)

Tab. 1 – Filiera di cantiere.

Intervenire sulle Mura Spagnole ha implicato il cercare di comprendere e riconoscere le peculiarità di questo ambiente, dei resti di una architettura fortificata e delle sue vicende e trasformazioni, ma soprattutto il confrontarsi con la sua connotazione nel contesto urbano contemporaneo. Parlare oggi delle Mura Spagnole significa parlare di pochi tratti rispetto alla loro lunghezza originaria che copriva un perimetro di circa 11 chilometri. Costruite in una decina di anni a partire dal 1549, sotto la direzione dell'ingegnere Gian Maria Olgiati e per iniziativa del governatore Ferrante Gonzaga, la nuova cinta bastionata raccordava al Castello Sforzesco includendo i borghi esterni. Tali mura

"alla moderna" costituivano un'opera militare di grandezza solenne ed erano all'epoca della loro costruzione le più imponenti d'Europa, insieme a quelle di Anversa, anche se prive di decorazioni monumentali in corrispondenza degli ingressi, almeno fino al 1598.[3] La concezione simbolica risulta ben rappresentata a partire dalla stampa cartografica prodotta da Nunzio Galizia (*Fig. 1*) e a seguire in tutte quelle realizzate tra il Cinquecento e il Settecento.[4]

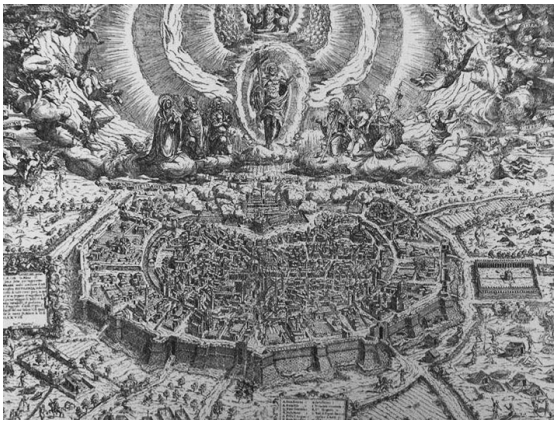


Fig. 1 – Nunzio Galizia, Milano liberata dalla peste*, 1578.*

[3] Per le vicende storiche si vedano principalmente: A. SCOTTI, *Per un profilo dell'architettura milanese (1535-1565)*, in *Omaggio a Tiziano. La cultura artistica milanese nell'età di Carlo V*, catalogo della mostra, Milano 1977, pp. 100-101; S. LEYDI, *Le cavalcate dell'ingegnero. L'opera di Gianmaria Olgiati, ingegnere militare di Carlo V*, Modena-Ferrara, 1989; A. FILASETA, *I Bastioni Spagnoli: organizzazione di cantiere e maestranze dal 1548 al 1553*, in *Milano città fortificata, vent'anni dopo*, Atti del convegno (Milano, 1 ottobre 2003), Quaderni del Castello Sforzesco, 5, Milano 2005, pp. 48-63.

[4] All'interno dell'Associazione Temporanea tra Professionisti del gruppo di progettazione, le vicende storiche sono state curate soprattutto per la parte cinquecentesca dal Prof. Stefano Della Torre e per gli interventi successivi dalla sottoscritta e dall'Arch. Lorena Bauce, insieme ad Antonella Mastronardi. Per le dettagliate vicende storiche di Porta Romana si rimanda a quanto prodotto nella relazione storica in parte ripresa in questa sede, alla bibliografia particolareggiata e alle seguenti fonti documentarie: Archivio Storico del Comune di Milano (ASCM), Località milanesi, 281; Archivio di Stato di Milano (ASMi), Fondi camerali p.a., 256; Archivio della Soprintendenza dei Beni Architettonici e del Paesaggio di Milano (AS BB AA PP Mi), cartella E/5/160; Archivio Borromeo Isola Bella, *Chiese nei territori dalla A alla B*, *Arona;* Biblioteca Ambrosiana di Milano (BAM), S 131 Sup., LVI-LVII e S 132 Sup., CXXIX.

Le porte apparivano essere "piutosto fori, & buche nelle mura, che porte convenienti a così gran Metropoli".[5] Il problema di una sistemazione del principale ingresso in città era già stato posto negli anni precedenti, ma fu la programmata venuta nel 1597 di Gregoria d'Austria, figlia dell'arciduca Carlo di Stiria e promessa sposa del principe Filippo, a dare lo spunto per festeggiamenti anche con l'erezione di archi trionfali (*Fig. 2*). Ma la sua morte prematura bloccò il progetto fino all'anno successivo, quando il principe si fidanzò con la sorella minore di Gregoria, Margherita, e i preparativi ripresero: l'ingresso trionfale di Margherita a Milano sarebbe avvenuto durante il suo viaggio verso la Spagna per conoscere lo sposo. L'evento è celebrato anche nella moneta-medaglia custodita nel medagliere del Castello Sforzesco (*Fig. 3*).

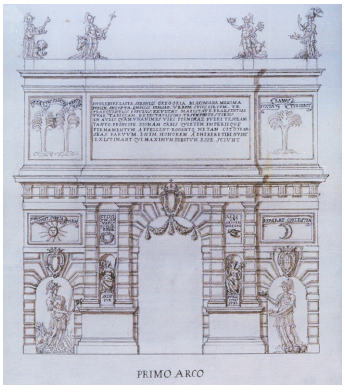


Fig. 2 – Porta Romana nel disegno inserito nel manoscritto di Guido Mazenta (?), Idea per l'arco da erigersi a Porta Romana in onore della principessa Gregoria d'Austria, 1597 (Madrid, Biblioteca Nacional, MS 2908).

Per decidere le iniziative della Città in onore dell'illustre ospite fu eletta una commissione scelta tra i membri del Consiglio Generale e i Decurioni, i quali il 28 luglio 1598 riferivano di aver stabilito che "la porta della città per la quale doverà entrare la detta Signora, che sarà

5 G. MAZENTA, *Apparato fatto dalla città di Milano per ricevere la serenissima Regina D. Margarita d'Austria sposata al potentiss. Re di Spagna D. Filippo III nostro signore*, Milano, nella stamperia del q. Pacifico Pontio, 1598 (seconda edizione 1599).

Porta Romana, si faccia di vivo et materia stabile a perpetua memoria" ossia in pietra, riprendendo il disegno fatto l'anno precedente.[6]

Fig. 3 – Moneta commemorativa dell'ingresso in città di Margherita d'Austria Regina di Spagna con l'effige della regina e con l'arco di Porta Romana (Milano, Civiche Raccolte Numismatiche, Medagliere Castello Sforzesco). Si trova anche in L. Candrini, L'Arco di Porta Romana*, cit. p. 399.*

Dopo essere state "pubblici passeggi" tra il Sette e l'Ottocento, le Mura non sono state comprese dai principali personaggi milanesi né reinserite nel circuito vitale urbano, forse anche perché non hanno mai svolto una reale funzione difensiva, ma solo daziaria; attribuendo loro probabilmente un significato negativo anche come limite al progresso ottocentesco e testimonianza di un periodo di dominazione straniera, di fatto ne hanno decretato la fine.[7] Le demolizioni e la mancanza di cura che dal Piano Beruto (1884-88) si sono susseguiti in modo costante, per piccoli e grandi episodi, dimostrano, in generale, il mancato riconoscimento dei loro molteplici valori testimoniali di carattere storico-simbolico e architettonico-costruttivo.[8] Vale la pena ricordare che Porta Romana ha rappresentato da sempre l'ingresso principale alla città di Milano, e ricordiamo che già nella cinta muraria di età comunale era

6 Relazione e fascicoli allegati del Progetto definitivo, 2006: *Restauro conservativo dell'Arco di Porta Romana*, da p. 3 e *Documentazione iconografica.*

7 Per le vicende degli ultimi due secoli si è consultato, presso gli archivi della Soprintendenza per i Beni Architettonici e per il Paesaggio, l'unica cartella presente, segnatura C/1/594 Milano- Mura Spagnole. In realtà ci sono poche testimonianze di voci in difesa delle Mura come i Soprintendenti Modigliani e Crema che però poco o nulla hanno potuto ottenere nel secolo XIX.

8 Si rimanda a M.T. FIORIO, *"Opus turrium et portarum": le sculture di Porta Romana*, in *Milano e la Lombardia in età comunale. Secoli XI-XIII*, Catalogo della mostra, Milano 1993, pp. 189-192.

stata realizzata la struttura monumentale di più alto valore simbolico della città medievale, demolita anch'essa alla fine del Settecento. Anche l'Arco di Porta Romana risalente al 1598 ha subito nel 1907 le operazioni di isolamento dalle cortine murarie del sistema difensivo (*Fig. 4*) e la risistemazione complessiva dell'area nel 1955, sistemazione che ha conferito al monumento il ruolo attuale di spartitraffico (*Fig. 5*), causandone un incremento di degrado: il deposito superficiale, causato dal particellato atmosferico, si trasforma in deposito più compatto e spesso, fino a diventare crosta con la presenza di gesso di neoformazione e di sali a causa delle condizioni ambientali e dall'elevato livello di inquinamento, causato principalmente dal traffico automobilistico.

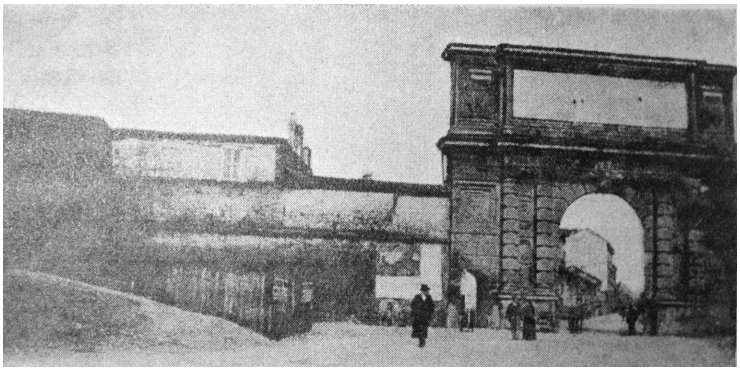


Fig. 4 – Fotografia storica di Porta Romana verso il 1900, il monumento è ancora fiancheggiato dalle Mura. Da L. Candrini, L'Arco di Porta Romana*, p. 400.*

Fig. 5 – Fotografia di Porta Romana del 2009, il monumento è isolato secondo la sistemazione attuale della piazza.

Ecco che l'intervento di conservazione delle Mura Spagnole di Milano, avvenuto tra il 2006 e il 2009, si è basato sul tentativo di cambiare tendenza: da una parte riconoscere i valori monumentali del complesso e portarli all'attenzione della comunità, dall'altro indirizzare il restauro verso le metodologie più corrette rispetto al progetto definitivo attraverso i criteri più aggiornati del processo conservativo di conoscenza, diagnostica e cura. Tale processo si è relazionato anche a un sentire tali luoghi "emozionanti" grazie al potere del frammento che evoca il tutto, del "margine" del tracciato urbano dalla cerchia delle mura ormai poco riconoscibile, della "soglia" come passaggio di un dentro e di un fuori del sistema difensivo decontestualizzato. Alla scala architettonica, come rudere, le Mura diventano documento di un saper fare, memoria di cui non si vorrebbe perdere oltre i molteplici valori testimoniali.

2. Stato di conservazione precedente all'intervento

2.1 Localizzazione e descrizione dei tratti di mura

Nella *Fig. 6* sono identificati in rosso i tratti delle Mura oggetto dell'intervento eseguito tra il 2006 e il 2009, compresi in una porzione posta circa a sud-est del sistema fortificato cinquecentesco (attuale circonvallazione esterna):

Tratto T1 cavaliere tra viale Caldara e viale Monte Nero;
Tratto T2 bastione di Porta Romana, circoscritto tra piazza Medaglie D'oro, viale Filippetti e viale Sabotino, e adiacente tratto di mura;
Tratto T3 mura bastionate con cavaliere compreso tra viale Filippetti e viale Sabotino;
Tratto T4 cavaliere tra viale Beatrice d'Este e via Patellani;
Tratto T5 punto di difesa avanzata corrispondente a viale Beatrice d'Este e largo Isabella d'Aragona;
Tratto T6 punto di difesa avanzata e tratto di mura comprese tra viale Beatrice d'Este fino a porta Ludovica.

Il primo tratto T1 corrisponde al cavaliere di viale Caldara-viale Monte Nero, presenta l'impianto di "giardino pubblico con memoria monumentale affiorante" con scale a rocaille e parterre a ghiaia, delimitato da un parapetto in ferro e tratti di Mura. Completano l'arredo urbano la creazione di aree di sosta e terrazzamenti.

Il secondo tratto T2, comprende il bastione di Porta Romana, unico bastione cittadino conservato. Si caratterizza per un rivestimento basamentale in ceppo d'Adda, un paramento esterno con tratto a scarpa, separato dal tratto superiore di muratura ad andamento verticale mediante un cordolo marcapiano a toro sempre in ceppo; all'epoca del restauro era concluso da cordolature cementizie e copertina sommitale risalente ad anni recenti. Verso l'interno del bastione è possibile trovare l'unico esempio di barbacani e dei relativi archi di scarico, originariamente protetti da terrapieni.

Fig. 6 – Schema dei lotti di intervento: tratti T1-T6.

Il terzo tratto di viale Filippetti T3 è connotato dall'intervento degli anni Sessanta, attuazione del progetto del 1953, che ha trasformato l'area corrispondente al cavaliere in giardino pubblico. Le parti interne della muratura risultano essere state interpretate come quinte scenografiche poste a sfondo della fontana.

Anche i resti delle mura presenti lungo viale Filippetti T4, largo Isabella d'Aragona T5, viale Beatrice D'Este fino a Porta Ludovica T6, sono il risultato delle trasformazioni del ventennio compreso tra il 1950 e il 1970; in particolare, le varianti ai Piani particolareggiati del 1953 e le molte autorizzazioni, anche da parte della Soprintendenza, a parziali demolizioni di tratti, persino nell'ordine degli 80 metri lineari. È evidente che le Mura erano viste come semplici cinte di delimitazione per le proprietà private in costruzione, non riconoscendone il loro valore monumentale.

2.2 Materiali e fenomeni di degrado

Con la consulenza del dott. Marco Realini e in particolare del

dott. Alberto Villa[9] si è preso in considerazione l'interrelazione tra il manufatto e le specie arboree.

I materiali che costituiscono la muratura, malte e mattoni, rappresentano infatti un substrato favorevole allo sviluppo della vegetazione per l'elevata porosità presente, che favorisce sia la ritenzione di umidità sia la penetrazione meccanica delle radici, anche per il fatto che i giunti di malta e i vari interstizi presenti nelle murature sono luoghi privilegiati per il deposito di polveri, materiali organici, ecc.. La compresenza di adeguata illuminazione che permette l'attività fotosintetica, di elevato tenore di umidità del substrato causato dalla pioggia, dall'umidità di risalita, da percolazione o condensa, e le condizioni ambientali favorevoli, hanno portato a una crescita biologica originata da comunità sostanzialmente erbacee, ma giunta fino a comunità arbustive e arboree più complesse e aggressive rispetto al substrato. Ricordiamo che, la crescita biologica è favorita da quei processi di degrado che portano all'aumento di porosità, al decoesionamento e alla fratturazione dei materiali esistenti, aumentando la "bioricettività"[10] del substrato; inoltre, le specie arboree risultano essere in grado di vivere in ambienti con un regime idrico discontinuo e poveri dal punto di vista nutritivo.

L'esecuzione delle operazioni di controllo degli agenti del degrado a prima vista potrebbero risultare relativamente semplici, invece sono alquanto complesse poiché esistono da un lato la resistenza allo strappo delle radici, che potrebbero ricacciare se dovessero restare vive in sito, e dall'altro l'impossibilità di raggiungere con mezzi meccanici le radici penetrate nelle strutture senza recare danni ai manufatti.

I tratti di mura presentavano emergenze decisamente differenziate in relazione alle vegetazione spontanea: a rischio di crollo, in parte già avvenuto (*Fig. 7*), fratturati con disgregazione delle malte e distacchi di intere parti del paramento murario in mattoni. Lo sviluppo vegetale

---

[9] Esperto di Beni Culturali e membro della Commissione Normal; ha collaborato alla realizzazione di molti progetti nazionali per la protezione di aree archeologiche dalla vegetazione infestante. Colgo l'occasione per ringraziarlo della disponibilità e ricchezza di dati e testimonianze personali che mi ha trasferito durante le nostre passeggiate (circa 2 km di tratti), nelle visite di sopralluogo al cantiere.

[10] CANEVA G., "La colonizzazione biologica della cinta muraria di Lucca: linee guida di un progetto di studio di biodeterioramento e di biomonitoraggio", in Atti del convegno *Le Mura di Lucca*, Lucca 17-19 maggio 2001, GIUSTI M.A. (a cura di), Alinea, Firenze, 2005, p. 194.

"non ostacolato" ha portato al collasso strutturale dei manufatti, determinando il loro "ruinare",[11] soprattutto nei tratti T2 e T3 (*Fig. 8*).

Fig. 7 – Fotografia del 2006, barbacane del tratto T2 di viale Filippetti; in particolare è evidente il danno prodotto da una grossa radice.

Fig. 8 – Fotografia del 2006, fianco del cavaliere del tratto T3 di viale Filippetti; è visibile la crescita incontrollata di arbusti di grandi dimensioni, alloggiati in particolar modo all'interno del fianco, che si comporta come una "vasca" di contenimento.

11 CANEVA G., op. cit., 2005, p. 195.

A partire dal progetto definitivo, si è proceduto a una serie di sopralluoghi con l'obiettivo di verificare le entità vegetali che crescono sui paramenti, maggiormente responsabili delle azioni di degrado, e nelle loro immediate vicinanze. A seconda del tratto analizzato, le specie individuate apparivano in molti casi uniformemente diffuse mentre altre erano più localizzate.

Tra le specie distribuite con una buona frequenza ovunque sulle mura c'erano: Ailanto, Bagolaro, Sedum, Vite americana, Edera e Parietaria. Queste ultime due si trovavano per parti di murature molto estese, lungo ogni tratto, particolarmente presenti sulle cortine orientate a nord e sud (prive di illuminazione e umide perché protette da vegetazione); nel tratto T6 Edera e Vite americana si presentavano folte e rigogliose su entrambi i lati delle mura, denotando una scelta estetica da parte dei privati che le avevano piantate. Particolarmente preoccupante per quanto riguarda la sua diffusione e collocazione era l'Ailanto: introdotto dalla Cina in Europa verso la fine del XVIII secolo, ha trovato in ambiente mediterraneo un perfetto acclimatamento, diventando una specie infestante di difficile controllo sia per la rapidità di sviluppo vegetativo che per i danni determinati dal suo insediamento sulle strutture murarie.

Si sottolinea che pur essendo il termine di *specie infestante* riferito a un concetto variabile e soggettivo, di norma lo si utilizza in quei casi in cui lo sviluppo di una specie interferisce con le attività umane o rappresenta un elemento di squilibrio nella naturale dinamica degli ecosistemi.[12]

Per quanto riguarda la vegetazione circostante le Mura, le piante e gli arbusti presenti costituivano una popolazione caotica e anomala per la compresenza di specie piantate con intento ornamentale, ad esempio gli Allori tra i barbacani dei tratti T2 e T3 o, arbitrariamente, da parte di cittadini (Aucuba, piccolo Abete, ecc..), e altre cresciute per semina spontanea, soprattutto Ailanti. L'assenza di uno studio agronomico delle specie idonee all'ambiente esistente ha creato casi di sofferenza e attacchi virali per le quali si è previsto l'abbattimento[13], di concerto con gli enti di tutela: Forestale e Soprintendenza.

---

[12] CANEVA G., "Il problema della crescita di Ailanthus altissima (Miller) Swingle nelle zone archeologiche e monumentali" in Atti del Convegno *Le Pietre nell'Architettura: Struttura e Superfici*, Bressanone, 25-28 Giugno, 1991, pagg. 225-233.

[13] La classificazione e ulteriori dettagli sulla vegetazione si può trovare nel fascicolo *Allegati alla Relazione* del Progetto esecutivo, 2006.

2.3 Stato di conservazione della muratura

Si è visto come il degrado causato dagli organismi biologici si articola attraverso i meccanismi fisici di decoesione e disgregazione del substrato effettuati dalla spinta e dalla penetrazione di strutture vitali nel materiale, che portano a danni di natura meccanica o da alterazione di tipo chimico.

La scelta di liberare dai terrapieni e lasciare a vista la maggior parte dei muri originariamente costruiti per essere ricoperti di terra, ha accelerato il processo di degrado perché parti del manufatto, non costruite per essere esposte agli agenti atmosferici, presentano una porosità maggiore rispetto a quella di un paramento esterno, realizzato con funzione di protezione del nucleo. Inoltre, non prevedere per tutto lo sviluppo una protezione superiore e l'assenza di una progettazione delle aree a verde e del controllo della crescita della vegetazione, ha determinato le situazioni di dissesti, crollo e perdita di consistenti porzioni di materia già denunciate.

Tra le principali cause di degrado delle murature troviamo l'acqua, sia allo stato liquido che a quello di vapore e, quindi, condensa. Esistono differenti vie di ingresso dell'acqua all'interno delle strutture murarie: essendo i mattoni e le malte materiali porosi sono permeabili all'acqua stessa. In molti casi, soprattutto sulla cortina a scarpa, i giunti di malta erano talmente degradati da risultare fortemente arretrati rispetto alla superficie, favorendo la penetrazione dell'acqua meteorica. Un'altra via privilegiata è la risalita dal terreno per capillarità oppure la discesa dall'alto per gravità, soprattutto in corrispondenza delle fessure e delle fratture createsi per la presenza della vegetazione e della mancata manutenzione.

Dalle analisi termografiche all'infrarosso e dai test gravimetrici, effettuati sia sul bastione che in un tratto di viale Filippetti, è risultato che le infiltrazioni erano presenti nella parte sommitale, tra il paramento e il riempimento della muratura e tra le mura e i contrafforti. La distribuzione dell'acqua all'interno delle strutture era presente con una maggiore concentrazione alla base delle murature, nei punti in cui si generavano dei piani interni di scorrimento che permettevano alle infiltrazioni di diffondersi rapidamente per gravità.

I processi degenerativi o degradativi di natura chimico-fisica sono accelerati dalle particolari condizioni climatiche di Milano, determinate in gran parte dalla conformazione orografica dell'area, essendo la pianura

padana circondata su tre lati da catene montuose che si estendono fino a quote elevate, determinando così peculiarità climatologiche sia dal punto di vista fisico che da quello dinamico, influenzando in modo determinante le capacità dispersive dell'atmosfera e quindi le condizioni di accumulo degli inquinanti, soprattutto nel periodo invernale.

Per la molteplicità delle situazioni riscontrate rispetto alle tessiture murarie, il degrado degli elementi, variabile per intensità da zona a zona, era presente sia nei giunti sia sugli strati più superficiali dei mattoni, particolarmente su quelli meno cotti o nei casi in cui erano presenti stilature cementizie. La muratura si presentava con caratteristiche differenti a seconda che fosse conservato il paramento a vista, a tessitura regolare, con corsi allineati e malta di allettamento di spessore uniforme, oppure che fossero esposte parti più interne. In questi casi già presente sopra i mattoni non erano più complanari, disposti senza corsi preferenziali e con la malta di allettamento stesa in modo disomogeneo. Lo stato delle mura era però anche il risultato di una lunga assenza di manutenzione che ha contribuito ad accelerare il degrado dell'apparato murario fino ai casi di dissesto e di perdita di ampie porzioni del costruito.

Proprio per la stretta relazione tra la patologia di degrado e il substrato interessato si è proceduto alla schedatura degli elementi e dei loro materiali, dei fenomeni di degrado che descrivono ed esemplificano quanto riscontrato rispetto al materiale specifico, al rilevamento del quadro fessurativo e del dissesto (analisi mancanti nel progetto definitivo).

Ricordiamo che se la superficie dell'opera è il luogo di tutte le trasformazioni fisiche e chimiche prodotte dall'interazione tra l'ambiente e il materiale, essa è anche il luogo della percezione estetica e la superficie da conservare in termini materici.

3. Intervento

3.1 Premessa metodologica

Il progetto definitivo è stato considerato alla luce dell'analisi dello stato di fatto del manufatto, definendo con molta attenzione eventuali rischi, controindicazioni o effetti collaterali dannosi. Nei casi di incertezza dei risultati alla prova del tempo, alla luce della complessità della reale condizione riscontrata, si è optato per una soluzione di "non intervento" in quanto è parsa la scelta più conservativa. Infatti, ogni inter-

vento di restauro deve essere valutato con la necessaria prudenza, evitando i rischi derivanti da una teorica fiducia nella tecnica, illudendosi che si possa sempre risolvere una problematica complessa con un singolo intervento "miracoloso". Ovviamente, il progetto esecutivo e l'intervento realizzato sono stati la conseguenza di un determinato approccio metodologico e la scelta di principi guida definiti: l'istanza dell'intervento è stata la conservazione dei materiali presenti ossia la salvaguardia della durata fisica dell'opera e, quindi, delle tracce dell'uomo e anche degli eventuali mutamenti causati dal trascorrere del tempo, per il periodo che sia il più lungo possibile.

3.2 Lavorazioni

L'intervento si è articolato in operazioni preliminari, sulla vegetazione, puliture, stuccature e sigillature, consolidamenti, opere di protezione, opere complementari. All'interno di queste categorie i singoli interventi sono stati relazionati ai differenti materiali presenti nel monumento, nonché alle relative patologie di degrado. Tutte le operazioni di restauro sono state eseguite da strutture operative con Sistema di Qualità certificato, dotato di SOA per la categoria di lavorazione OG2 e OS2.[14]

Di seguito riportiamo le lavorazioni eseguite.[15]

Le *operazioni preliminari* hanno compreso: l'eliminazione degli arbusti situati all'interno delle aree comprese tra i barbacani o adiacenti al manufatto, in modo da liberare le mura e le zone limitrofe, consentendo l'allestimento del cantiere e l'esecuzione delle operazioni di restauro e consolidamento del manufatto; il recupero dei mattoni, loro accatastamento su bancali nelle immediate vicinanze, pulitura e successivo loro reimpiego nelle lavorazioni di consolidamento e piccole integrazioni; la messa in sicurezza di porzioni di muratura pericolanti, nei casi di rischio di crollo o di forte distacco di parti di muratura in mattoni o di singoli elementi (mattoni o pietra), mediante opere provvisionali quali puntelli

---

[14] Si tratta delle categorie che si possono ottenere con l'attestazione SOA: OG2 Restauro e manutenzione dei beni immobili sottoposti a tutela e OS2 Superfici decorate di beni immobili del patrimonio culturale e beni culturali mobili di interesse storico, artistico, archeologico ed etnoantropologico

[15] Il presente paragrafo riprende ampiamente quanto già riportato nel Consuntivo scientifico, redatto nel 2009, al termine dei lavori, a cura della sottoscritta e dell'Arch. Lorena Bauce.

(*Fig. 9*), legature e stuccature provvisorie, a garanzia della stabilità per consentire le operazioni successive; il preconsolidamento delle superfici lapidee particolarmente degradate con applicazione di prodotto consolidante (esteri dell'acido silicico), applicato a pennello, al fine di consentire le successive operazioni di pulitura senza perdita di materiale.

Fig. 9 – Sequenza fotografica della messa in sicurezza dei barbacani a rischio di crollo, tratto T3 corrispondente al cavaliere di viale Filippetti.

Per quanto riguarda la *vegetazione e l'attacco biologico*, si ricorda che i principali biodeteriogeni sono parte di un vasto ambito che comprende batteri (autotrofi ed eterotrofi), attinomiceti, funghi, cianobatteri e alghe (cianoficee e cloroficee), licheni (crostosi, fogliosi e fruticosi), briofite (muschi ed epatiche), piante superiori (erbacee, arbustive e arboree) e animali (insetti, lumache e uccelli); per ogni categoria di biodeteriogeni esistono appropriati metodi di controllo, che possono essere di tipo meccanico, fisico e chimico selettivi.

Nel caso specifico si è proceduto richiamando trattamenti che vengono eseguiti in ambito archeologico e monumentale: metodi meccanici, attentamente integrati con un misurato uso di quelli chimici. In particolare si ricorda che esiste una categoria di specie, detta "ruderale" che alligna in preferenza proprio sul costruito. Tali specie ruderali si nutrono a spese dei materiali da costruzioni quali le malte, gli elementi lapidei artificiali e anche naturali. Da tutto ciò è emersa la necessità di definire i migliori metodi di asportazione per le specie da eliminare e quelle che potevano sopravvivere accanto alle mura, fino alla questione più grave che ha riguardato i casi degli Ailanti i cui tronchi si sono intersecati con l'apparecchio murario.

Le operazioni di restauro sulle strutture murarie, in laterizi a vista, hanno previsto trattamenti preliminari per l'eliminazione sia della vegetazione erbacea che di quella arbustiva e ad alto fusto.

Il primo intervento ha previsto il taglio dei rami e la rimozione delle masse non aderenti alla muratura; successivamente si è proceduto all'eliminazione della vegetazione ruderale e infestante mediante l'applicazione di prodotti chimici sistemici. Per quanto riguarda patina biologica, funghi, batteri, muffe, alghe e muschio si è proceduto su tutte le superfici interessate con un trattamento a spruzzo; un secondo trattamento è stato eseguito con un erbicida sistemico liquido solubile, a base di glifosate,[16] scelto per la bassa tossicità e assenza di attività residuale. Il prodotto agisce su sedum, parietaria, edera, vite americana, per assorbimento da parte delle foglie e viene traslocato per via sistemica sino alle radici e agli organi perennanti (*Fig. 10*).

Fig. 10 – Sequenza fotografica del trattamento del verde con glifosate (erbicida sistemico).

Dopo un tempo di attesa di circa 3 settimane si è passati alla rimozione della vegetazione umifera ed erbacea "morta" per l'azione del biocida.

Come già evidenziato nel testo, per motivi di maggiore efficacia e minor rischio di lesioni al substrato, si è preferito un trattamento chimico rispetto al solo intervento meccanico in quanto la semplice asportazione meccanica non risulta efficace, a meno di non poter rimuovere in maniera completa gli apparati radicali, e inoltre l'asportazione diretta di strutture vegetanti con radici significative e profonde può ledere il substrato a causa della resistenza delle radici stesse. Invece, dopo l'ap-

16 Per il Glyphosate è stato condotto uno studio in occasione del restauro delle Mura di Ferrara. Cfr. PASETTI A., MASSA V., COZZI E., "Studio dell'impatto del 'diserbo chimico' su paramenti in laterizio a vista. Il caso delle Mura di Ferrara" in Atti del Convegno *Le Superfici dell'Architettura: il cotto. Caratterizzazioni e trattamenti*, Bressanone, 30/6 – 3/7, 1992, pagg. 743-747.

plicazione di uno specifico biocida, radici e capillizio subiscono una notevole riduzione di volume dovuta alla totale perdita di acqua, consentendo di poter disalveolare, senza danni ai manufatti, il materiale vegetale che vi si era insinuato.

Per quanto riguarda le piante che avevano sconnesso le strutture murarie, compenetrandole profondamente, e che non consentivano la loro estirpazione senza provocare danni alla muratura stessa, si è proceduto con il taglio della pianta, lasciandone temporaneamente parte *in situ*, e applicando il successivo trattatamento di "devitalizzazione". Tale procedimento è stato eseguito mediante iniezioni di diserbante liquido (principio attivo Picloram), previa realizzazione di opportuni fori, eseguiti con trapano, al fine di raggiungere i vasi conduttori (trachee) delle infestanti arbustive e legnose (*Fig. 11*).

Fig. 11 – Sequenza fotografica del trattamento di "devitalizzazione" delle radici con le iniezioni di Picloram.

Dopo la totale "devitalizzazione" del legno (con un tempo variabile in funzione della stagione e della risposta dell'elemento vegetale), si è proceduto in due differenti modi: mediante rimozione completa del tronco, ove era possibile, attraverso lo smontaggio di piccole porzioni di muratura e tagli successivi; con la "mummificazione" della pianta ossia la sua conservazione in situ quando, invece, l'eliminazione della massa vegetale avrebbe comportato la perdita di elevate parti di muratura. Questa seconda tecnica prevede l'annegamento della massa vegetale all'interno della malta o materiali da costruzione, per privare di ossigeno e di luce l'apparato radicale "morto", al fine di evitare il degrado dello stesso (marcescenza). Il procedimento è visibile anche nell'ultima fotografia della sequenza illustrata nella *Fig. 12*.

A seguire sono state eseguite le *puliture* utilizzando i criteri di selettività, gradualità e controllabilità dei diversi metodi, e rispettando la superficie del materiale, senza produrre discontinuità superficiali o provocare la

formazione di sottoprodotti dannosi per la conservazione del materiale lapideo. L'intervento sulla muratura (malta, mattoni e pietra) ha previsto una pulitura ad acqua diffusa e generalizzata, mediante idropulitrice erogante acqua deionizzata, integrata con spazzolatura (spazzole morbide) e AB57 nelle zone con depositi più resistenti. L'impiego successivo e puntuale di microaeroabrasivatrice di precisione è stato utilizzato nei casi di incrostazioni e depositi tenaci, non rimossi con la pulitura ad acqua e AB57. L'inerte utilizzato è stato il carbonato di calcio, mentre la pressione di impatto è stata molto bassa e variabile a seconda delle zone. Si fa presente che per la forte adesione del deposito ai mattoni compatti, non è stato possibile ovunque rimuovere completamente la crosta nera, pena la perdita degli strati superficiali del mattone.

Fig. 12 – Sequenza fotografica della fase di consolidamento mediante barre in acciaio inox filettate. Si può notare nella terza foto la fase di "mummificazione" della radice.

Impacchi di polpa di cellulosa a fibra media di 200 micron e acqua distillata sono stati eseguiti per l'estrazione di sali, mentre per la rimozione delle scritte vandaliche si è proceduto con una prima applicazione di prodotto a base di cloruro di metilene, un lavaggio successivo e un ritocco con la microaeroabrasivatrice. Inoltre, per quanto riguarda l'abbassamento delle macchie di ossidi di ferro si sono effettuati impacchi a base di EDTA bisodico, usando sempre la polpa di carta come ispessente.

Nella scelta progettuale le *stuccature e le sigillature* hanno rivestito un ruolo fondamentale per la protezione delle mura, soprattutto in riferimento al fenomeno della semina ambientale ossia la capacità di diffondere da parte delle piante i propri semi anche lontano dal luogo di emissione e

la capacità di questi ultimi di attendere le condizioni ottimali per germinare. Essendo il muro interessato dalla presenza di umidità, prevalentemente di natura discendente, il tipo di malta del giunto ha assunto un'importanza notevole sia per quanto riguarda la distribuzione dell'acqua all'interno del muro sia per quanto riguarda l'evaporazione di quest'ultima e quindi del degrado. Il fenomeno infatti è strettamente connesso anche alla velocità del flusso evaporativo che in condizioni termoigrometriche esterne costanti dipende dal tipo di porosità e soprattutto dal coefficiente di assorbimento capillare. Il criterio utilizzato per la scelta della malta del restauro si è basato principalmente sulla compatibilità chimico-fisica e meccanica rispetto all'esistente, sia nel senso dell'eventuale interferenza tra il materiale aggiunto e quello esistente sia nel senso dell'affinità di comportamento al variare delle condizioni esterne (igroscopicità, modulo elastico, deformabilità termica). Soddisfatta la compatibilità, sono stati privilegiati i criteri di durabilità, manutenibilità, formale e cromatico, orientandosi verso una riconoscibilità delle nuove malte, ottenuta però evitando violente intromissioni nella percezione dell'immagine dell'opera nella sua integrità e in relazione all'ambiente circostante. Per questo motivo si è optato per una differenziazione delle nuove malte formulate con calce idraulica naturale, polvere di marmo nera, gialla e rossa, graniglia e/o sabbia vagliata, a seconda del materiale di supporto. Stuccature delle fessurazioni, risarciture e stilature delle malte di allettamento dei giunti sono state precedute dalla rimozione di quelle malte cementizie o degradate (*Fig. 13*), dannose alla muratura o non più idonee, probabilmente eseguite in occasione di precedenti interventi, effettuata con microscalpelli.

Fig. 13 – Fase di rimozione delle malte di allettamento e di stilatura non più idonee.

L'applicazione delle nuove malte, eseguita per strati successivi stesi a mano o con la fugatrice nei tratti regolari della cortina muraria, è stata eseguita con spatolini metallici per consentire una stilatura puntuale e molto accurata. Dopo un periodo di tempo che ha consentito un primo indurimento dello strato finale, si è proceduto con la lavatura e tamponatura delle superfici con spugne di mare e acqua deionizzata al fine di porre in risalto l'aggregato, la sua dimensione e la sua specifica cromia (*Fig. 14*). La lavorazione si è prefissa anche lo scopo di conservare e rendere leggibile i segni, per così dire "calligrafici", che si riscontrano osservando con attenzione le diverse stilature e rifiniture dei giunti che, allo stato attuale delle conoscenze, sono l'unica testimonianza materiale delle trasformazioni o manutenzioni più o meno recenti.

Fig. 14 – Fase di rifacimento delle malte di stilatura: malta applicata con fugatrice (a sinistra) e malta a lavorazione conclusa.

Per le stuccature e i riempimenti delle fessurazioni profonde si è proceduto mediante colatura o iniezione in profondità di un impasto a base di calce idraulica naturale e sabbie vagliate (*Figg. 15, 16*).

Anche sui fenomeni di alveolizzazione si è operato restando sottotono, dal punto di vista cromatico, con una formulazione a base di grassello di calce, cocciopesto, polvere di marmo nera e rossa, terre coloranti.

Il *consolidamento e la sigillatura delle fessurazioni profonde* e dei distacchi del rivestimento in ceppo sono stati eseguiti con un impasto a base di calce idraulica naturale, polvere di marmo e sabbia vagliata. Inoltre il consolidamento profondo della muratura si è articolato attraverso due differenti procedure: con riempimenti di malta e riadesioni e con la tecnica del cuci-scuci nei casi di paramento decoeso, ammalorato o friabile. Per quanto riguarda il cuci-scuci si è proceduto con la rimozione degli elementi esterni fortemente degradati o sconnessi, la rimozione dei

giunti di allettamento inconsistenti, la successiva pulitura delle superfici, la preparazione delle stesse, il riutilizzo o l'integrazione con mattoni di recupero. I nuovi elementi sono stati fatti fare a mano, realizzati secondo un campione prelevato dalle mura, riconoscibili per fattura, ma simili all'esistente per dimensione, colore e composizione.

Fig. 15 – Sequenza fotografica della fase di consolidamento mediante iniezioni della muratura.

Fig. 16 – Dettagli fotografici della fase di consolidamento della muratura.

A seguito della verifica puntuale dello stato di consistenza della muratura, messo in luce dopo l'intervento sulla vegetazione, si sono inserite barre in acciaio inox filettate (*Fig. 12*) nei casi in cui si sono riscontrati paramenti esterni dei barbacani, probabilmente costruiti in epoca successiva rispetto alla parte centrale. Tali rivestimenti non ammorsati, ma solo accostati, hanno favorito l'azione degli apparati radicali che risalendo hanno distaccato il paramento stesso.

L'inserimento di elementi metallici è stato eseguito anche su quattro

barbacani situati nell'angolo sinistro del cavaliere verso via Filippetti; questi barbacani si caratterizzavano per la presenza di radici e tronchi d'albero di dimensioni notevoli, che hanno distaccato la porzione mediante fessurazioni profonde, diventate le sedi dei tronchi stessi, con fenomeni di inclinazione e fuori piombo, e dissesti localizzati. Poiché si trattava di porzioni di muratura relativamente compatta e resistente, si è potuto procedere alla ricostituzione muraria del paramento mediante "cuci-scuci" e interventi di riempimento delle fessurazioni profonde e dei distacchi, aggiungendo tirantature metalliche laterali contro il ribaltamento delle porzioni ricucite (*Fig. 17*), anche nei casi di spinta accidentale (ad esempio per atto vandalico o urto di un mezzo di trasporto).

Fig. 17 – Sequenza fotografica del consolidamento dei barbacani con tiranti in acciaio inox.

In generale, l'intervento non ha previsto l'applicazione di un *protettivo* sull'intero paramento murario per facilitare l'evaporazione dell'acqua dalla muratura, ma il trattamento si è reso necessario puntualmente ove l'azione della pioggia di rimbalzo era più frequente.

Particolare attenzione si è posta nell'individuare differenti tipologie di copertine per la protezione del manufatto da infiltrazioni, ruscellamenti e depositi d'acqua, ad esempio la protezione della parte inferiore delle buche pontaie e dei mattoni in aggetto con realizzazione di copertina in malta, modellata per raccordare e agevolare l'allontanamento delle acque meteoriche, con caratteristiche simili a quella utilizzata nelle parti circostanti per le stilature e sigillature (calce idraulica naturale, sabbia vagliata e polvere di marmo); la copertina della parte sommitale realizzata in malta con andamento a dorso d'asino o in pendenza, stesa in due strati su supporto preventivamente pulito e preparato: lo strato di fondo è costituito da calce idraulica naturale e sabbia vagliata, mentre l'impasto della malta di finitura è stato realizzato con calce idraulica naturale, cocciopesto e polvere di marmo giallo oro e nera (*Fig. 18*). Le copertine hanno avuto anche lo scopo di colmare le discontinuità e le piccole mancanze presenti sulle superfici, in modo da ridurre la possibilità di infiltrazioni e ristagni.

Fig. 18 – Sequenza fotografica dei vari strati della copertina sommitale posta sulle creste delle mura.

A chiusura delle buche pontaie è stato inserito un mattone posto sotto livello (arretrato di circa una testa) che consente di leggere l'andamento percettivo delle forature nel paramento murario.

Pur non contemplando in questo primo lotto la sistemazione ambientale dell'area adiacente alle mura, rimandata a un secondo lotto di intervento (riguardante sistemazione delle aree a verde e drenaggio delle acque, pavimentazioni esterne, formazione di impianti di illuminazione, di irrigazione e di recinzione perimetrale), il progetto esecutivo ha previsto la creazione di una "zona di rispetto" da realizzare attraverso una fascia perimetrale in ghiaia per evitare il contatto diretto dell'attuale tappeto erboso con la muratura (*Figg. 19, 20*).

Fig. 19 – Fotografia del bastione al termine dei lavori (2009).

Fig. 20 – Fotografia del tratto T3 di viale Filippetti al termine dei lavori (2009).

4. Conclusioni

Le Mura Spagnole sono un complesso monumentale all'aperto e l'ambiente è costituito da uno spazio prevalentemente naturale. Di conseguenza la problematica si identifica con il controllo dello sviluppo dei biodeteriogeni presenti. Infatti sia gli spazi circostanti sia le strutture sono progressivamente interessate dal fenomeno dell'avanzare o del ritorno di tutte le forme di vita a suo tempo presenti; tale processo naturale, si rivela dannoso ai fini della conservazione del patrimonio storico-artistico.

Per la problematica della vegetazione sulle Mura, considerato il legame che si è venuto a creare tra opera di natura e opera dell'uomo, un corretto intervento sulla vegetazione dovrebbe prevedere delle fasi successive: "correzione, conversione e mantenimento".[17] La prima fase è finalizzata all'eliminazione della vegetazione pericolosa per il manufatto, attraverso l'uso di biocidi, la seconda si propone di agire tempestivamente durante l'eventuale ripresa vegetativa, attraverso ulteriori trattamenti chimici, e l'ultima fase riguarda un programma di manutenzione da condurre attraverso il monitoraggio costante. In tale ottica l'intervento e il

[17] BENENTE M., MATTONE M., "Il verde e il rudere: un legame indissolubile", in Atti del convegno *Le Mura di Lucca*, Lucca 17-19 maggio 2001, GIUSTI M.A. (a cura di), Alinea, Firenze, 2005, p. 210.

progetto di manutenzione sono da intendersi come un fatto unitario che si allontana dall'idea dell'intervento di restauro come evento eccezionale, volto a ricondurre l'opera a uno stato "definitivo", e afferma invece l'idea di un processo di cura costante che tende a garantire la più lunga permanenza, regolando le modificazioni che inevitabilmente sono il risultato del divenire. Da una fase di "correzione" prevista con l'intervento di restauro si sarebbe dovuto passare a una fase di "mantenimento". Ad esempio, alla fase di eliminazione della vegetazione considerata pericolosa per il manufatto (trattamento chimico e meccanico della vegetazione superiore), si sarebbe dovuto far seguire un monitoraggio delle specie e una verifica del quadro fessurativo, con eventuale successiva eliminazione degli infestanti nuovamente insediati.

Perché sono state restaurate le Mura spagnole se a otto anni dall'inizio del restauro del primo tratto delle mura e cinque anni dalla conclusione dei lavori ci si trova di fronte alla negazione di una loro cura? Il tempo che è testimone di memorie e monumenti, nel caso delle Mura ha documentato un fare e un abbandono. Lo stato di abbandono in cui versano i tratti di muratura e le frequentazioni di "giardini" e tratti pubblici confermano che non basta ri-pensare la manutenzione in un'ottica di visione processuale della conservazione, ma si deve applicare questa strategia condivisa di sviluppo della conservazione nei termini più ampi possibili, costituita dalle fasi di prevenzione, restauro, manutenzione e gestione del bene, ove l'attività di studio e valorizzazione sono il collante del processo.

Fig. 21 – Fotografia del tratto T3 di viale Filippetti nel 2012.

Negli anni del restauro si sono realizzate diverse iniziative culturali da parte dei professionisti coinvolti e della società di sponsorizzazione: convegni e visite guidate dedicate ai cittadini, agli studenti universitari e post laurea, a stranieri, ad associazioni italiane sui castelli, ma nel caso delle Mura non è esistita la volontà di gestione pubblica delle aree, peraltro sancito anche nell'art. 9 della Costituzione italiana (*Fig. 21*).

BIBLIOGRAFIA

Di seguito si riporta una breve bibliografia a cui si è fatto riferimento nella Relazione storica del progetto.

G. MAZENTA, *Apparato fatto dalla città di Milano per ricevere la serenissima Regina D. Margarita d'Austria sposata al potentiss. Re di Spagna D. Filippo III nostro signore*, Milano, nella stamperia del q. Pacifico Pontio, 1598 (seconda edizione 1599).

S. LATUADA, *Descrizione di Milano*, II, Milano 1738.

SCOTTI, *Per un profilo dell'architettura milanese (1535-1565), in Omaggio a Tiziano. La cultura artistica milanese nell'età di Carlo V*, Catalogo della mostra, Milano 1977.

B.G. VIGOTTI, *Impronte romane, medioevali e moderne sull'area dell'Istituto Gaetano Pini*, Milano 1981.

PELLEGRINO, *Porta Romana*, Libreria Milanese, Milano 1985.

S. LEYDI, *Le cavalcate dell'ingegnero. L'opera di Gianmaria Olgiati, ingegnere militare di Carlo V*, Modena-Ferrara 1989.

A. LANZA, M. SOMARE, *Porta Orientale, Romana e Ticinese*, Libreria Milanese, Milano 1992.

G. RICCI, *"Et parea che Milano diuenuto fosse di tutto il Mondo amplissimo Theatro… ". Festa e teatro a Milano negli ultimi anni del Cinquecento*, in *La fabbrica, la critica, la storia. Scritti in onore di Carlo Perogalli*, a cura di G. COLMUTO ZANELLA, F. CONTI, V. HYBSCH, Milano 1993.

M.T. FIORIO, *"Opus turrium et portarum": le sculture di Porta Romana*, in *Milano e la Lombardia in età comunale. Secoli XI-XIII*, Catalogo della mostra, Milano 1993;

S. FROMMEL, *Sebastiano Serlio architetto*, Milano 1998;

G. GENTILE, D. FEA BIANCOLINI, *Chiesa di Santa Maria di Loreto (Santa Marta)*, in *Arona sacra: l'epoca dei Borromeo*, catalogo della mostra di Arona, a cura di G. ROMANO, Torino 1977.

Guida d'Italia del Touring Club Italiano, *Milano*, Milano 1998.

A. SCOTTI, N. SOLDINI, *Borromini milanese,* in *Il giovane Borromini : dagli esordi a San Carlo alle quattro fontane,* catalogo della mostra di Lugano, a cura di M. KAHN-ROSSI e M. FRANCIOLLI, Milano 1999.

V. MILANO, *I fratelli Mazenta negli episcopati di Gaspare Visconti e Federico Borromeo*, in "Arte lombarda", 131, 2001. F. MINCHIOTTI, *Un precedente del pronao di Santa Maria Nuova di Abbiategrasso: Tolomeo Rinaldi e il progetto per l'oratorio di Sant'Agostino a Treviglio*, in "Arte lombarda", 131, 2001.
A. MORANDOTTI*, Milano profana nell'età dei Borromeo*, Milano 2005.
P. VENTRONE, F. MARCHESI, *Il passaggio di Margherita d'Austria a Milano tra lutto e festeggiamenti d'occasione*, in atti delle giornate di studio *Il viaggio attraverso l'Italia di Margherita d'Austria, regina di Spagna, 1598-1599*, S. Giuliano Terme (Pisa), 22-23 settembre 2006, (all'epoca in corso di stampa).

Fonti Documentarie

Archivio Storico del Comune di Milano (ASCM), Località milanesi, 281.
Archivio di Stato di Milano (ASMi), Fondi camerali p.a., 256.
Archivio della Soprintendenza dei Beni Architettonici e del Paesaggio di Milano (AS BB AA PP Mi), cartella E/5/160.
Archivio Borromeo Isola Bella, *Chiese nei territori dalla A alla B, Arona.*
Biblioteca Ambrosiana di Milano (BAM), S 131 Sup., LVI-LVII e S 132 Sup., CXXIX.""

Di seguito si riporta una breve bibliografia a cui si è fatto riferimento nella parte del degrado e dell'intervento.

G. CANEVA e G. DE MARCO, *Il controllo della vegetazione nelle zone archeologiche e monumentali*, in *Manutenzione e conservazione del costruito tra tradizione ed innovazione- Bressanone*, 1986, pp. 553-569.
AA.VV., *Le Pietre nell'Architettura: Struttura e Superfici*, Atti del Convegno di Bressanone 25-28 Giugno 1991.
AA.VV., *Le Superfici dell'Architettura: il cotto. Caratterizzazioni e trattamenti*, Atti del Convegno di Bressanone, 30/6 – 3/7, 1992.
L. MASETTI BITELLI (a cura di), *Archeologia, recupero e conservazione*, Fiesole, 1993.
G. GOVI, M. SPERANZA, A. CIARALLO, *Archeologia e botanica*. Ministero per i Beni Culturali e Ambientali, Soprintendenza archeologica di Pompei, 2000.
M.R. DI FABIO a cura di*, Le Mura di Ferrara. Storia di un restauro*, Minerva Edizioni, Ferrara, 2003.
G.CANEVA, *La biologia vegetale per i beni culturali, conoscenza e valorizzazione*. Vol II, Nardini Editore, 2001.
M.A. GIUSTI a cura di, Atti del convegno *Le Mura di Lucca*, Lucca 17-19 maggio 2001, Alinea, Firenze, 2005.

Istituto Lombardo (Rend. Scienze) 146, 65-92 (2012)

IL CANTIERE DELLA "CA' GRANDA". LA STORIA DEL MANUFATTO NELLA LETTURA DELL'ARCHITETTO-CONSERVATORE

REBECCA FANT (*)

Nota presentata dal m.e. Amedeo Bellini
(Adunanza del 22 Marzo 2012)

SUNTO. – La fase della conoscenza indiretta e diretta delle facciate si pone quale metodo e strumento finalizzato a porre sotto una nuova luce il complesso architettonico. Ciò permette di rileggere sincronicamente la documentazione scientifica già esistente e quella in fase di acquisizione, ma soprattutto il manufatto-documento *in itinere*. L'affinamento dello sguardo analitico e delle possibilità di indagine, unito ai potenziali rapporti tra le informazioni raccolte, attua un processo di continua ricerca e di possibili legami con le storie passate e future; storie materiali di cantiere non sempre documentate negli archivi, ma che di fatto hanno influenzato le scelte metodologiche d'intervento in funzione delle risposte puntuali delle superfici. Tale processo è stato perseguito come conseguenza della considerazione che conoscere l'oggetto, la sua realtà fisica, le sue storie, implementa e stratifica proprio nel momento dell'intervento i processi conoscitivi, creando nuove relazioni e nuove interpretazioni, massimizzando le permanenze, così che le letture dei protagonisti coinvolti possano diventare strumenti di avvicinamento al monumento anche per il pubblico dei "non esperti"; è un nuovo modo di costruire (nel senso di stratificare, interpretare e aggiungere) per e con la collettività, un moderno modo di condividere e sentire la "cosa pubblica".

***

(*) Architetto specialista in Restauro dei Monumenti. Docente a contratto del Politecnico di Milano, Italy.
E-mail: fant@rebeccafantarchitetto.it

ABSTRACT. – Knowledge is a methodological tool aimed to enlighten an architectural complex. During the preliminary survey a synchronous reading of the existing scientific documentation and of the new data coming from the examination of the site, is allowed. The links among the analytical view from one side, and the collected data, from the other, are the base for the continuity between past and future. Material histories, not always documented in archives, influence methodological choices as to surfaces features. This process is considered as a step in the knowledge process regarding the building, its physical truth, its historical burden; the intervention works implements and stratifies this process, creating new relations and meanings. Doing this, the knowledge offered by the different professionals involved constitutes an approaching tool even for the non-experts. It is a new way of building with the meaning of stratifying, adding, interpreting with and for the community; a modern mode of sharing and feeling about the public affairs.

1. INTRODUZIONE

L'intervento di restauro conservativo delle facciate quattro-seicentesche della "Ca Granda" prospicienti le vie Festa del Perdono e San Nazaro è stato un appalto pubblico affidato con procedura di sponsorizzazione, della durata di circa quattro anni (dal 2009 e il 2013) ed eseguito con una ripartizione in sette lotti d'intervento[1] (*Fig. 1*).

La realtà complessa e polifonica della fabbrica, il tipo di appalto (*Tab. 1*) e l'approccio metodologico applicato, hanno indotto fin da subito a porre un'attenzione particolare alle superfici che comprendesse differenti aspetti, apparentemente non strettamente legati.

A partire dal "progetto definitivo" posto a base di gara, si è cercato di verificare le scelte d'intervento previste attraverso riscontri diretti ovvero una conoscenza materica delle facciate e indiretti mediante una lettura approfondita della bibliografia.

Lo studio del materiale e del suo degrado è stato sviluppato nel 2009 con il cantiere sperimentale, attraverso le attività di diagnostica e campionatura dell'intervento,[2] propedeutiche alla stesura del progetto esecutivo. In tale fase si sono preparate le superfici e prelevati dei cam-

[1] A seguito di bando e relativa gara, l'Università degli Studi di Milano ha stipulato una convenzione nel 2008 con l'Associazione Temporanea di Imprese *T.M.C. Pubblicità s.r.l./Gasparoli.*

[2] Questa prima fase ha visto coinvolti in particolare la sottoscritta in veste di progettista esecutivo e direttore dei lavori, l'Impresa Gasparoli e l'Istituto per la Conservazione e la Valorizzazione dei Beni Culturali (ICVBC) – Area della Ricerca

pioni che poi sono stati studiati in laboratorio dall'Istituto per la Conservazione e la Valorizzazione dei Beni Culturali di Milano.[3]

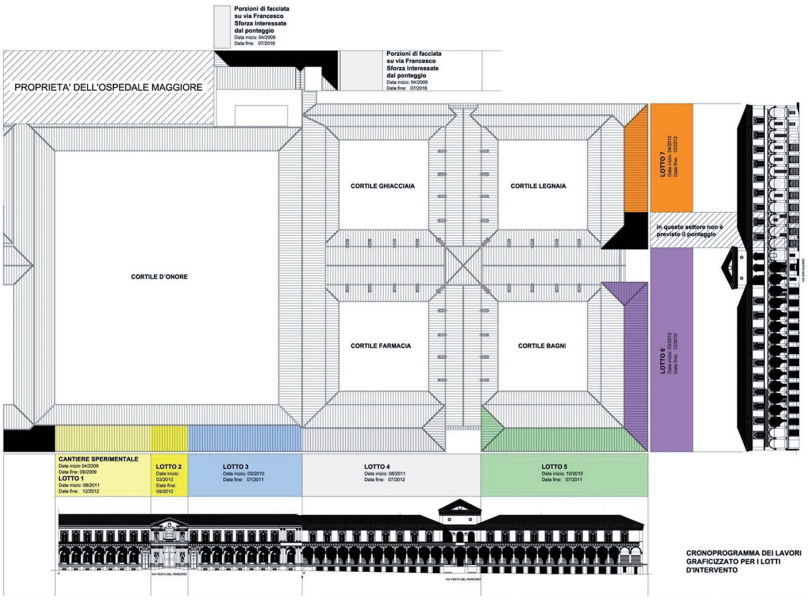


Fig. 1 – Schema dei lotti d'intervento.

Parallelamente, pur leggendo l'intervento derivante dalle condizioni reali del manufatto e quindi svincolato da giudizi di tipo storico-estetico o critico, si è avviato un percorso di conoscenza che ha consentito di rileggere la storia della costruzione, anche se molte domande non hanno ancora trovato una risposta certa.

Le nuove osservazioni derivano dai seguenti fortunati aspetti:

CNR sezione di Milano "Gino Bozza", Milano Bicocca, con Responsabile scientifico Dott. Marco Realini, presenti Dott. Roberto Bugini, Dott. Antonio Sansonetti, Dott.ssa Chiara Colombo e loro collaboratori. Invece la sperimentazione di bio pulitura con batteri solfato riduttori è stata condotta dal Centro interdipartimentale di Ricerca e Servizi per i Beni Culturali (PhD DiSTAM) - University of Milan, referenti scientifiche la Prof.ssa Claudia Sorlini e la Dott.ssa Francesca Cappitelli.

[3] Per quanto riguarda gli aspetti della diagnostica di cantiere e di laboratorio, si rimanda al lavoro di Antonio Sansonetti "Il cantiere della Ca' Granda: la lettura dell'esperto scientifico".

– i tempi lunghi del cantiere e l'estensione delle superfici hanno portano a prendere confidenza e "familiarità" con il manufatto architettonico, permettendo una continua riflessione dettata dal processo di scoperta/acquisizione di dati e dalla loro provvisoria elaborazione;
– il continuo accostamento e la rilettura dei dati stessi hanno consentito di aprire connessioni tra tematiche differenti, disvelando la creatività del pensiero e delle molteplici connessioni che potenzialmente ogni manufatto possiede per chi lo sa ascoltare o interpretare;
– il confronto pluri-disciplinare, concretizzatosi attraverso le effettive collaborazioni tra figure professionali diverse, nel rispetto dei ruoli, ha concatenato le esperienze passate e presenti relative alla Ca' Granda, quali singolarità che si sono espresse nei comuni obiettivi finali di carattere conservativo e di valorizzazione del bene.

Proprietà: Responsabile Unico del Procedimento:	**Università degli Studi di Milano** Dott. Ing. **Michele Tagliaferri** dal 2011 Arch. **Peppino D'Andrea**
Alta sorveglianza:	**Soprintendenza per i Beni Architettonici e per il Paesaggio di Milano** Arch. **Libero Corrieri** dal dicembre 2011 Arch. **G.B. Sannazzaro**
Comitato Tecnico Scientifico:	Arch. **Libero Corrieri –** Arch. **Aldo Di Silvestro** Ing. **Michele Tagliaferri** dal 2011 Arch. **Peppino D'Andrea**
Committente:	**TMC Pubblicità s.r.l.** – Milano
Progetto esecutivo e Direzione Lavori:	Arch. **Rebecca Fant** – Milano Collaboratrici: Arch. Stefania Bossi – Arch. Michela Catalano Per le prestazioni di Misura e contabilità: Ing. **Paolo Elli**
Consulenti scientifici:	Prof. Arch. **Amedeo Bellini -** Prof. Ing. Arch. **Stefano Della Torre**
Consulente tecnico-artistico del committente:	Arch. **Carlo Valentini**
Coordinatore per la sicurezza in fase di progettazione e di esecuzione:	Arch. **Gabriele Gotti**
Impresa esecutrice: Responsabile di cantiere:	**Gasparoli S.r.l.** – Gallarate (Va) Dott. **Marco Gasparoli**
Diagnostica:	**Istituto per la Conservazione e la Valorizzazione dei Beni Culturali Area della Ricerca CNR - Sezione di Milano "Gino Bozza"** Responsabile scientifico: Dott. Marco Realini Ricercatori: Dott. Roberto Bugini, Dott.ssa Chiara Colombo, Dott. Antonio Sansonetti
Sperimentazione di bio pulitura:	**Centro interdipartimentale di Ricerca e Servizi per i Beni Culturali (PhD DiSTAM) - University of Milan** Responsabili scientifici: Prof. Claudia Sorlini e Dott.ssa Francesca Cappitelli

Tab. 1 – Filiera di cantiere.

L'attività di indagine e registrazione dei dati raccolti ha consentito di rivelare materiali, tecniche, modi di costruire e di restaurare che riportano alle differenti culture e mentalità, esemplificate spesso nel singolo caso e in relazione con l'insieme.

Tale attività di studio, spesso disattesa nei cantieri di restauro per

mancanza di tempo e risorse economiche, in realtà trova riscontro a livello legislativo; nella legge di tutela vigente[4] infatti l'attività di studio è definita come fulcro di ogni indagine e intervento sugli edifici storici,[5] così che l'obiettivo ultimo della conservazione dovrebbe essere costruire nuova conoscenza, in un processo di riconoscimento dei valori che porta alla cura e all'attenzione dell'autenticità materiale del bene.

L'occhio dell'architetto-conservatore si caratterizza quindi anche per l'approfondimento o la reinterpretazione "provvisoria" delle conoscenze storiche che origina dalle nuove domande che si pongono al manufatto e dalle relazioni che si trovano tra i dati di varia natura a disposizione (tecnico-scientifici e storici).

Da ciò deriva che una delle tematiche più interessanti e attuali della disciplina riguarda proprio la registrazione, trasmissione e gestione delle informazioni prima-durante e dopo il cantiere, le tipologie di banche dati, nel duplice aspetto di trasmissione di informazioni e letture possibili e delle attività conservative eseguite o previste, in un'ottica di processo di lungo periodo.[6]

L'indispensabile confronto tra i dati d'archivio e bibliografici, le indagini *in situ* e quelle di laboratorio presuppongono delle competenze specialistiche e multi-disciplinari che vedono la conoscenza come attività plurima strettamente legata alla conservazione: di mantenimento e custodia nel tempo di quanto è giunto sino a noi, che concilia la preservazione di tutte le possibili letture o interpretazioni dei suoi valori con l'utilizzo rispettoso del manufatto architettonico.

Effettivamente il cantiere, attraverso le attività di "scavo" e le indagini si configura come il luogo privilegiato per ascoltare nuovamen-

---

4 D.Lgs. 42/2004, Codice dei beni culturali e del paesaggio, detto Codice Urbani.

5 Art. 29, comma 1 del D.Lgs. 42/2004.

6 Tale visione si colloca nell'ambito della manutenzione dei beni culturali esistenti, secondo l'orientamento più aggiornato del dibattito in corso. Mi riferisco al recepimento di un concetto di manutenzione che da un approccio tecnologico-industriale si è trasformato in nuovo processo "dedicato" al patrimonio edilizio esistente. Le Linee Guida promosse dalla Regione Lombardia nel 2003 sono infatti considerate lo spartiacque di tale sviluppo culturale che si lega anche all'obbligo di redigere, da parte del progettista prima e del direttore dei lavori poi, il "piano di manutenzione", in stretta relazione con il progetto e il "consuntivo scientifico" che dovrebbe definire i "risultati culturali e scientifici raggiunti" nel restauro. DPR 207/2010, art. 38 Piano di manutenzione dell'opera e art. 250 consuntivo scientifico.

te e rileggere il monumento che costituisce la testimonianza autentica di una complessità stratigrafica e interpretativa spesso poco "definita" e che, come per la Ca' Granda, ha reso difficoltoso e a volte impossibile separare l'attribuzione a un solo periodo storico rispetto alle vicende che si sono succedute e stratificate.

Ripensare alla Ca' Granda in questi termini ha trovato riscontro nelle celebrazioni del 90° anniversario della fondazione dell'Università degli Studi di Milano (novembre 2014), avvenuta con l'inaugurazione di un percorso di valorizzazione e di fruizione del complesso monumentale che ospita sia la Fondazione IRCCS legata all'Ospedale Maggiore Policlinico sia l'Università: il progetto *La Ca' Granda dei Milanesi. Itinerario interdisciplinare nel fulcro di una metropoli multiculturale* si è prefissato l'obiettivo di educare, guidare e assistere il patrimonio storico-scientifico che il complesso custodisce, con opportuni strumenti divulgativi e multimediali, visite guidate, eventi.[7]

2. Rilettura delle vicende storiche

La fabbrica è costituita dalle fasi costruttive quattrocentesca, seicentesca e sette-ottocentesca, alle quali si devono sovrapporre tutte le aggiunte, modifiche e stratificazioni derivanti da necessità utilitaristiche, adeguamenti igienico sanitari, manutenzioni e restauri che si sono avvicendati sin dal Cinquecento fino ai giorni nostri. Quindi lo stato attuale rivela tracce e permanenze di storie materiali di cantiere, non sempre documentate negli archivi, ma che di fatto si sono rivelati dati utili da registrare perché hanno influenzato le scelte metodologiche d'intervento, in funzione delle risposte puntuali ai trattamenti delle superfici e dei materiali presenti.

Per contro, queste differenti risposte legate alla consistenza fisico-materica, come il differente comportamento dei materiali rispetto al degrado, sono stati indizi utili per la ricerca delle anomalie, delle irregolarità, apparentemente insignificanti, ma che ci hanno indirizzato a

[7] Progetto a cura della Prof.ssa Francesca Vaglienti, al quale si è partecipato in qualità di consulente per gli aspetti architettonici e di restauro. F. VAGLIENTI (a cura di), *La Ca' Granda dei Milanesi. Itinerario interdisciplinare nel fulcro di una metropoli multiculturale, edizioni Nexo, Milano, 2014.*

una lettura più attenta alla centralità del singolo elemento e della singola formella.

Il taglio scelto per questo contributo esclude di occuparci dettagliatamente degli eventi del complesso della Ca' Granda, in quanto la letteratura è abbondante e affronta più tematiche: da studi di inquadramento più generale ad approfondimenti specialistici[8] [2]. Tuttavia è indispensabile ricordare brevemente le principali vicende storiche che si sono succedute e che hanno portato all'attuale stato di fatto della fabbrica e in particolare delle facciate così dette quattro-seicentesche.

L'attuale sede dell'Università degli Studi di Milano ha origine dalla volontà di Francesco I Sforza di fondare una struttura ospedaliera in grado di rispondere alle esigenze del ducato che fino a quel momento erano state assolte da più complessi cittadini. Lo scopo sociale e politico, nonché l'alto valore simbolico e rappresentativo del *glorioso albergo de' poveri di Cristo* sono ben illustrati nel *Trattato di Architettura* del Filarete,[9] dove lo stesso progetto viene inquadrato all'interno di Sforzinda, "città ideale" ed evidente tributo al suo committente e protettore (*Fig. 2*). Infatti è noto che fu proprio lo Sforza a volere l'Averlino con un ruolo simile a quello attuale di progettista e direttore dei lavori,[10] ruolo che svolse dal 1456 al 1465, quando lasciò Milano.

Dal 1459 al 1465 avvenne l'edificazione della prima crociera o infermeria degli uomini, costituita dai corpi verso le facciate di via Festa del Perdono e di via San Nazaro (porzione sinistra del complesso), compresi i portici fino all'altezza della fascia marcapiano.

---

[8] Per approfondimenti di carattere architettonico si rimanda alla bibliografia allegata, seppure da considerare non esaustiva; invece per una bibliografia specifica e pressoché completa fino alla data di edizione si veda la *Bibliografia specifica* in FRANCHINI, 1995, da p. 174.

[9] Il *Trattato di architettura* di Antonio Averlino, detto il Filarete, fu steso presumibilmente tra il 1461 e il 1464, ma non venne pubblicato. Il complesso ospedaliero viene descritto all'interno del trattato nell'XI libro. Per approfondimenti si rimanda in particolare all'edizione critica di FINOLI-GRASSI, 1972.

[10] Si ricorda che il duca dovette faticare non poco per ottenere la nomina del Filarete, da parte del Capitolo dell'Ospedale, a: "Magister Antonius de Florentia designaverit formam et effigiem dicti hospitalis", e così ricevette ricevette il titolo di "Architectum fabricatorem directorem et ingenierium tanti operis" ovvero dell'Ospedale Maggiore di Milano, secondo quanto trascritto nell'Ordinazioni Capitolari, II, c. 104, documento del 29 febbraio 1460, riportato da Lucio Franchini in FRANCHINI, 1995, p. 35 e relativa nota 121.

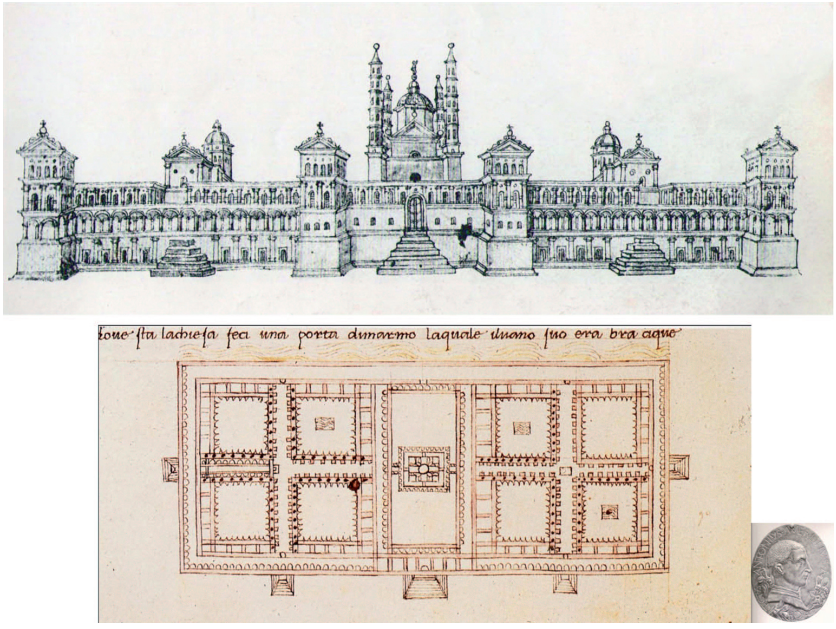


Fig. 2 – Illustrazioni dal Trattato d'architettura *del Filarete tratte da Grassi, 1972, ill. nn. 1, 8 e 15.*

Il primo cortile, detto "della Farmacia",[11] fu ultimato nel 1467 da Guiniforte Solari che, succeduto al Filarete, diresse anche la costruzione del cortile "dei Bagni", nonché il secondo ordine della facciata principale e la prima parte verso l'odierna via Francesco Sforza. Alla sua morte i lavori furono diretti dal figlio Pietro Antonio e da Ambrogio da Rosate che proseguirono con la costruzione dei cortili "della Ghiacciaia" (già detto "della Spezieria") e "della Legnaia" (già detto "della Cucina"). Infine, fu edificato il lato porticato verso l'attuale cortile centrale, attribuito a Giovanni Antonio Amadeo che dal 1495 assunse la direzione del cantiere. Nel Cinquecento venne chiuso il portico, prima con materiale ligneo (1529) e poi, nel 1597, con una cancellata metallica ancora presente; mentre nel 1600 Dionigi Campazzo chiuse la terrazza in corrispondenza dell'ingresso principale della crociera.

[11] Le vicende sono trattate in più testi bibliografici, per semplicità si rimanda a CARLESSI-KLUZER, 2011, p. 29.

Solamente con il lascito testamentario di Giovanni Pietro Carcano, del 1621, si diede avvio alla sistematica e poderosa ripresa del complesso con l'ideazione e la realizzazione del corpo centrale, detto richiniano, dal nome di Francesco Maria Richini che ne è considerato l'ideatore principale. Ricordiamo che, a tale proposito, i lavori consentirono "di dare da un lato un più ordinato e funzionale assetto al complesso e, dall'altro, di conferire all'edificio i caratteri di decoro e monumentalità che avevano contraddistinto il progetto primitivo"[12] (*Figg. 3, 4*).

Infatti, a partire dal 1625 il Capitolo diede incarico di proseguire i lavori a Giovanni Battista Pessina, ingegnere dell'Ospedale, coadiuvato dagli architetti Francesco Maria Richini e Fabio Mangone e dal pittore Giovanni Battista Crespi. Il cantiere venne avviato nel 1626 e si concluderà nel 1649, a partire dal corpo d'ingresso, proseguendo con il lato destro, poi con quello verso il naviglio e con il lato dell'Amadeo, del quale si dovette demolire il portico recuperando l'apparato decorativo.[13]

Fig. 3 – Veduta della facciata di via Festa del Perdono al settecento (da Lo stato presente di tutti i paesi e popoli del mondo naturale, politico, morale … *Venezia, 1715) tratto da Grassi, 1972, ill. n. 321.*

12 CARLESSI-KLUZER, 2011, p. 30.

13 CARLESSI-KLUZER, 2011, p. 35. Dal 1626 al 1634 l'ingegnere Giovanni Battista Pessina fu a capo dei lavori in qualità di Ingegnere dell'Ospedale, a seguire dal 1634 al 1644 con Giovanni Angelo Crivelli, e dal 1644 al 1649 con Carlo Buzzi.

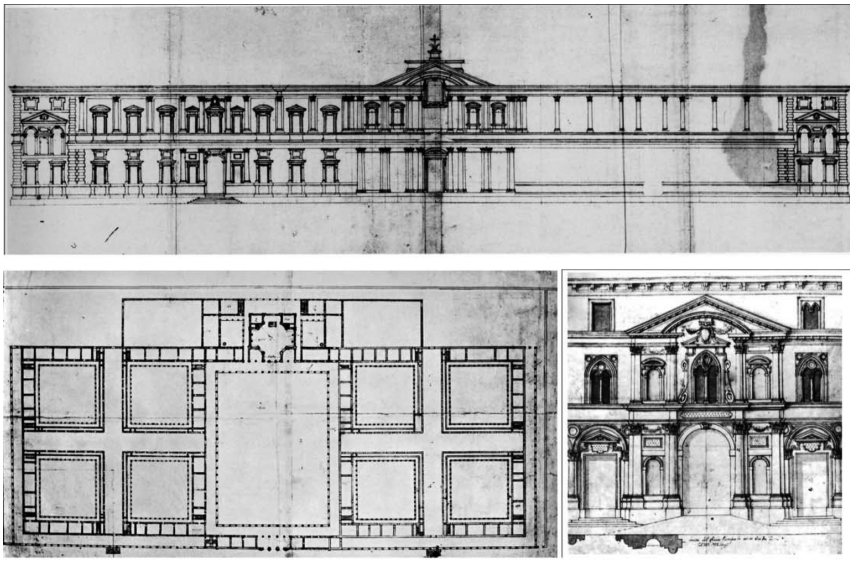


Fig. 4 – Illustrazioni dell'ampliamento dell'Ospedale (Raccolta Bianconi, Tomo III, p. 2, p. 8) e disegno autografo del Richini riferito alla parte centrale della facciata, verso l'antica via dell'Ospedale (biblioteca Ambrosiana F. 251, 207 riprodotto in "Arte Lombarda", n. 37, Milano, 1972) tratte da AA.VV., 1993, 66-67-68.

Oltre alla costruzione del cortile centrale e della corrispondente porzione di facciata verso via Festa del Perdono, i lavori proseguirono nei cortili sforzeschi "della Ghiacciaia", "della Legnaia" e "dei Bagni" (1636). In questa fase seicentesca abbiamo la presenza in cantiere di molti scultori operanti in Milano che lavorarono all'apparato scultoreo decorativo come Lasagna, Bianco, Rusca, Giovenzano e Giussano (*Fig. 5*).

Anche il progetto del Richini, come già quello del Filarete, non verrà completato, ma nel 1649 si iniziarono le fondamenta della seconda crociera, quella detta "delle donne" e si proseguì la realizzazione dei bracci per tutta la seconda metà del secolo. Di questo quadrilatero barocco, solamente il cortile "della Segreteria" detto anche "delle Balie" venne realizzato porticato con logge soprastanti; mentre a partire dal 1695, probabilmente su progetto dell'ingegnere Attilio Arrigoni, si andava definendo il corpo centrale verso via Francesco Sforza con l'abside della chiesa (la terza in ordine di edificazione/sostituzione).

Sempre a causa delle gravi condizioni economiche, nel 1791, l'ingegnere Pietro Castelli approntò un progetto di variante al progetto

richiniano per concludere la crociera barocca denominata "ala Macchi" o "ala Macchia", che risultasse funzionale ed evitasse inutili sprechi economici per aspetti formali, ritenuti non sostanziali (*Fig. 6*).

Fig. 5 – Veduta della fronte verso via Festa del Perdono, sec. XIX. Disegno di L. Rupp, Raccolta Bertarelli. Tratto da L. Grassi, Lo "Spedale dei Poveri" del Filerete. Storia e Restauro*, Milano, 1972, ill. n. 58 e relativa didascalia.*

Fig. 6 – Facciata verso via Festa del Perdono: tratto quattrocentesco, seicentesco e ottocentsco, dopo di restauro conservativo delle facciate.

In tale occasione,[14] in considerazione dello stato di conservazione delle terrecotte esistenti, a partire dal 1815 Luigi Cagnola fu delegato a sovrintendere i lavori di restauro "alla facciata di codesto Ospedale Maggiore nel modo, e con sagome di lavori eguali ai già esistenti ornati".[15]

A meno di eventuali invenzioni degli autori, in una veduta del 1810 c.a (*Fig. 7*). si nota il dettaglio della prima bifora del fronte verso via Francesco Sforza, poi rinvenuta durante i restauri della Grassi, ma soprattutto la fascia marcapiano è ancora simile a quella presente oggi sulla facciata verso via Festa del Perdono; nella veduta del 1840 (*Fig. 8*), invece, risulta cambiato il dettaglio della bifora, sostituito con due aperture rettangolari (presenti fino al restauro della Grassi) e la fascia marcapiano risulta intonacata.

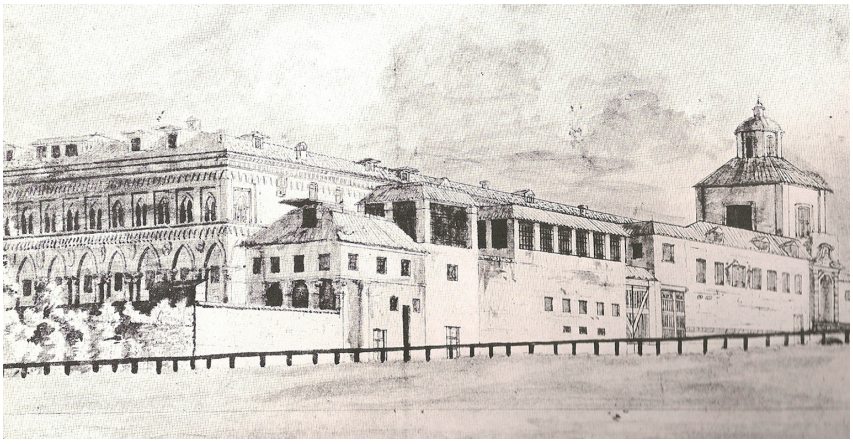


Fig. 7 – Veduta dal Naviglio verso la via Francesco Sforza. Disegno a penna e seppia, c. 1810, Raccolta Bertarelli. Tratto da L. Grassi, Lo "Spedale dei Poveri" del Filerete. Storia e Restauro*, Milano, 1972, ill. n. 123 e relativa didascalia.*

Nel periodo compreso tra il 1674 e il 1731 si era già chiuso il portico esterno verso via San Nazaro e nel 1694 il portico verso via Festa del Perdono, questa volta però in muratura; mentre negli anni tra il

14 Si tratta dell'approvazione del progetto del Castelli da parte della commissione di Pubblico Ornato che vede il Cagnola presidente di commissione.

15 FRANCHINI, 1995, p. 149 e relativa nota 45.

1785 e il 1804 risale la costruzione dei sopralzi della crociera "delle donne" e l'ammodernamento di dieci bifore verso San Nazaro e altre otto verso il Naviglio. Vale la pena segnalare che lungo il secondo tratto della facciata verso via S. Nazaro si è rilevata la data 1792 incisa in una formella della cornice superiore del fregio marcapiano (*Fig. 9*).

Fig. 8 – L'Ospedale, visto dal ponte di Porta Romana, in un dipinto di anonimo. 1840 c.a, Museo di Milano. Tratto da L. Grassi, Lo "Spedale dei Poveri" del Filerete. Storia e Restauro*, Milano, 1972, ill. n. 125 e relativa didascalia.*

Fig. 9 – Fotografia della formella superiore della fascia marcapiano in corrispondenza del davanzale finestre con incisa la data 1792, facciata verso via San Nazaro.

Ulteriori interventi compresi tra il 1841 e il 1857 sono stati la realizzazione della doppia crociera anche per gli uomini (nel 1856-57), a cura dell'ingegnere Antonio Tarra, e riparazioni varie alle facciate, in particolare verso via Festa del Perdono e verso il naviglio (attuale via Francesco Sforza). A tale riguardo nella relazione del giugno 1857 l'ingegnere Tarra dichiara che "in quanto può esservi di vero ristauro, nulla venne eseguito, consistendo il fatto in questi ultimi tempi nell'aver rimessi alcuni pezzi di antico ornamento alle fronti dello Stabilimento verso Strada a ripristino dello Stato di prima".[16]

Tra gli interventi di restauro della seconda metà dell'Ottocento segnaliamo quelli sulla "parte ornamentale della facciata verso strada e verso il piazzale del Cascinotto" ovvero la facciata quattrocentesca verso via Festa del Perdono, risalenti agli anni 1866-71 e 1879. Si tratta molto probabilmente degli interventi del Pandiani di cui si lamentarono diversi architetti dell'epoca tra i quali Mongeri, Brocca e Paravicini. Il restauro del Pandiani infatti aveva inserito ventotto medaglioni nel fronte principale sforzesco, con busti barocchi rinvenuti presso i depositi, a discapito di una coerenza stilistica, ma soprattutto aveva sostituito le formelle in terracotta degradate o mancanti con nuove formelle realizzate in malta di cemento, mimetizzandone il colore con una finitura rossa (*Fig. 10*).

Nel medesimo intervento erano anche state "raddrizzate" e in parte sostituite alcune colonne e rifatto il basamento del portico.[17]

Fig. 10 – Fotografia della formella in terracotta realizzata in malta di tipo cemento decorativo con finitura rossa.

16 FRANCHINI, 1995, p. 150 e relativa nota 46.
17 FRANCHINI, 1995, p. 151-52.

Tito Vespasiano Paravicini descrivendo le particolarità del monumento che "formano il carattere speciale dell'edificio, la sua propria fisionomia", denuncia la "deprecata sostituzione dei materiali nell'Ospedale Maggiore".[18] Tale critica è rimarcata anche in uno scritto del 1881, testo di una conferenza tenuta al Collegio degli Ingegneri ed Architetti in Milano e pubblicato ne *Il Politecnico,* dove viene segnalata la "sostituzione di ben cinque arcate dell'Ospedale Maggiore di Milano";[19] si tratta molto probabilmente dell'intervento risalente al 1868-1871 del Pandiani.[20]

Dopo un lungo dibattito sulle sorti del complesso ospedaliero della Ca' Granda, ormai considerato obsoleto e non più funzionale ad assolvere il ruolo di Ospedale Maggiore, nel 1935 venne ceduto in gran parte al Comune, per farne la sede prestigiosa dell'Università degli Studi, nata nel 1924. L'amministrazione pubblica incaricò fin da subito Ambrogio Annoni della progettazione del restauro.[21] Ma il conflitto

---

18 BELLINI, 2000, p. 52 e relativa NOTA 38. La frase, riportata da Bellini, è tratta da un inedito del Paravicini dal titolo *Considerazioni ed appunti per ristauro dei monumenti architettonici principalmente della Lombardia*, proveniente dal Fondo Paravicini presso la Biblioteca Ambrosiana di Milano. Preme riportare quanto osservato da Bellini riguardo all'atteggiamento rispettoso del Paravicini di tutto quanto "si allontana dalla regola, ma che per questo non deve essere soggetto a censura, ma anzi piuttosto considerato come caratteristica specifica dell'opera" p. 146. Tale atteggiamento per l'epoca è un'eccezione che ci fa mettere in guardia sulla prassi operativa dei restauri ottocenteschi.

19 BELLINI, 2000, p. 56 e relative NOTE 48 e 49. Bellini chiarisce che pur definendolo "lato attribuito a Bramante" in realtà l'autore si riferisce a quello richiniano". Di tali restauri si trovano dettagli anche in FRANCHINI, 1995, p. 151.

20 Sulla prassi comune di sostituire gli elementi degradati si veda anche il caso delle formelle della Certosa di Pavia, realizzate nel 1873 in occasione di una programmazione di intervento. Nel caso specifico il materiale non era stato utilizzato perché di dimensioni troppo piccole rispetto all'esistente, non avendo considerato tale fenomeno durante l'asciugatura del materiale. Gli elementi si trovano tutt'ora nei depositi della fabbrica. BENTIVOGLIO, 2008, p. 287.

21 Annoni aveva ricevuto l'incarico di occuparsi del progetto di massima del restauro della parte sforzesca, compreso il rilievo e il progetto "organico di restauro", affiancato da Pier Luigi Magistretti e Piero Portaluppi; mentre per quanto riguarda il rilievo grafico e fotografico collaborarono Liliana Grassi, Livio e Pier Giacomo Castiglioni, ed Egizio Nichelli. Ancora, dopo il '49 nel Comitato Tecnico dell'Università saranno presenti l'ingegnere Americo Belloni, gli architetti Liliana Grassi e Adalberto Borromeo. FRANCHINI, 1995, pp. 158-159. Per una trattazione dettagliata degli interventi della seconda metà del Novecento si rimanda in particolare ai testi di GRASSI, 1958 e 1972; VITAGLIANO, 2006 e 2008; CAVALLERI, 2008; CRIPPA, 1986 e 2008.

bellico interruppe il programma e dopo di allora tutto non fu più come prima. I bombardamenti avvennero tra il 1942 e il 1943 e "ridussero l'edificio in più parti allo stato di rudere e solo la forza psicologica reattiva alla irrazionalità della guerra poté sostenere il coraggio, oggi forse inimmaginabile, di affrontare una operazione di restauro di tali proporzioni, di raccogliere il recuperabile, ricomporlo, integrare quanto perduto dalla piccola alla grande lacuna. Furono centrati e distrutti il cortiletto della Ghiacciaia e la relativa fronte verso via Francesco Sforza, i tre lati adiacenti del cortilone, la loggetta del Campazzo e parte della facciata di via Festa del Perdono, mentre lo scoperchiamento e altri gravi danni furono inferti a entrambe le crociere."[22]

I restauri post-bellici, che proseguirono per quasi quarant'anni, iniziarono nel 1946 privilegiando la ristrutturazione dell'ala sforzesca e di quella centrale richiniana (*Fig. 11*).

Metodologicamente l'intervento di restauro e riuso dell'antico Ospedale Maggiore di Milano per l'Università degli Studi" superò la tendenza all'unitarietà formale e al completamento; l'antico venne valorizzato accostandolo al contemporaneo.

Fig. 11 – Fotografia storica della facciata di via Festa del Perdono dopo i bombardamenti.

[22] FRANCHINI, 1995, pp. 159.

In tale senso anche le facciate di via Festa del Perdono, via San Nazaro e via Francesco Sforza rappresentano la maturazione del pensiero teorico-metodologico dell'intervento sull'esistente: il passaggio graduale dalla riproduzione degli elementi in terracotta, dei quali possediamo ancora molti esemplari nei depositi dell'Università, alla scelta di mantenere le lacune, come si può osservare nella prima bifora d'angolo sulla facciata di via Francesco Sforza (*Fig. 12*), già menzionata.

Risulta quindi evidente che il complesso della Ca' Granda, di fatto è un palinsesto, definito dalla stratificazione di successivi ampliamenti, trasformazioni, sostituzioni, liberazioni e forse proprio questo è il segreto del suo fascino.

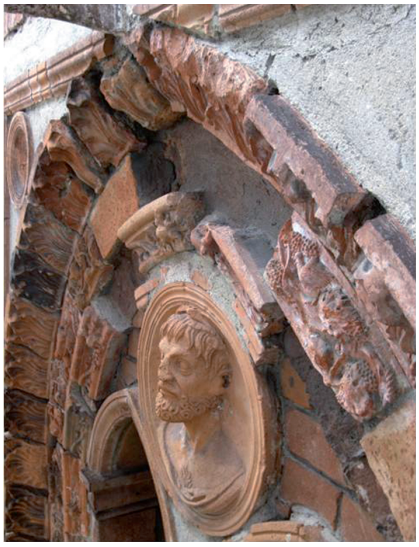


Fig. 12 – Dettaglio fotografico della prima bifora a sinistra verso via Francesco Sforza. Si tratta di una finestra scoperta durante i lavori di restauro post bellici, a cura di Liliana Grassi.

Quello che emerge dallo stato di fatto è una semplicità d'impianto planimetrico e apparentemente di facciata che si articola a partire dal doppio impianto a croce greca, che costituisce i chiostri interni, separato dal cortile centrale, sorta di raccordo tra le due ali con un perimetro in parte porticato. Il senso di decoro e di ordine della partizione nasce principalmente dalla proporzionalità, dalle similitudini euritmiche dell'apparato fittile, con la ripetitività di schemi e morfologie decorative

presenti negli archi a tutto sesto del portico, nelle ghiere e nelle cornici delle bifore con motivi a stampo (*Fig. 13*).

Non possiamo dimenticare gli autori originari della decorazione stessa, in particolare già nel 1461 risulta costruito il portico fino al "frixo posito super portichum", completato nel 1463 con l'opera appunto di Rinaldo de Stauris da Cremona che decorava allora il cortiletto dell'ospedale di Cremona e poi quello di Pavia.[23] Sappiamo anche che il Filarete forniva materiale lavorato e pezzi decorativi, così come i Solari; in particolare a Francesco si deve la realizzazione, nel 1467, del piano superiore del fronte verso via Festa del Perdono, mentre il fronte verso S. Nazaro risale al 1481, anche se i lavori proseguiranno fino al 1493.[24]

Fig. 13 – Dettaglio fotografico dell'attacco tra la parte quattrocentesca e quella seicentesca della facciata di via Festa del Perdono.

23 FRANCHINI, 1995, pp. 140; MAGGI-NASONI, 1980.
24 FRANCHINI, 1995, pp. 142; MAGGI-NASONI, 1980.

Tutto l'apparato architettonico-decorativo in terracotta che costituisce le ghiere lisce, le liste marcapiano, le cornici di davanzale, le cornici sottogronda, la serie di dentelli a dente di sega, accostate nei cornicioni solariani della Ca' Granda, contribuisce a caratterizzare fortemente le facciate (*Fig. 14*).

Fig. 14 – Schema con la localizzazione e particolari delle cornici di bifora quattrocentesche; sono evidenti formelle non intere.
Tratto dalle tavole della tesi PORTA-TODESCO, 2009-10.

Tale aspetto è riconosciuto anche nel Seicento quando, per quanto riguarda il tratto della facciata seicentesca verso via Festa del Perdono, le disposizioni dell'Ospedale prevedevano che fosse allineato con la parte già in essere e che ne seguisse lo schema e il ritmo, con la costruzione di un piano attico: "il terzo ordine in facciata"[25] demolito negli interventi post bellici da Annoni per essere stato interpretato come mera aggiunta utilitaria. Nelle differenze delle due porzioni emerge la particolarità delle formelle in terracotta decorate con motivi del repertorio classico dei putti vendemmianti, fin qui definite a stampo per semplicità, ma in realtà nel Seicento realizzate con una tecnica mista che comprende almeno in parte una modellazione manuale e ciò le rende pezzi particolari e unici; mentre le foglie d'acanto, i motti e i tortiglioni risultano effettivamente eseguiti a stampo (*Figg. 15, 16, 17*).

25 FRANCHINI, 1995, pp. 143.

Gli interventi che negli anni hanno interessato le superfici hanno compreso la sostituzione di singoli elementi ceramici, sia nel paramento murario che, ancor più, nelle parti di ornato. Si ritrovano così numerose formelle realizzate con materiali ceramici analoghi dal punto di vista composizionale rispetto ai materiali originari, ma differenti ad occhio nudo per colore (rosso-arancio, rosso bruno, crema, variegato), inclusioni, compattezza e risposta agli agenti atmosferici; oltre alle sostituzioni ottocentesche effettuate con malte anche cementizie. La sostituzione delle formelle del Quattrocento, effettuata tra il Sette e il Novecento, è probabilmente la causa della perdita quasi completa di un tipo di putto con il volto orientato verso sinistra (presente in sole cinque formelle), volto orientato al contrario rispetto al tipo più frequente[26] (*Fig. 18*). Questa doppia tipologia del putto vendemmiante è presente anche nel chiostro piccolo della Certosa di Pavia. Si ricorda che al 1464 risale l'inizio della decorazione in terracotta del chiostro piccolo della Certosa di Pavia, completata nel penultimo decennio; con documentata anche a Pavia la presenza di Rinaldo de Stauris (1464-66) e Francesco Solari (1462-67).[27]

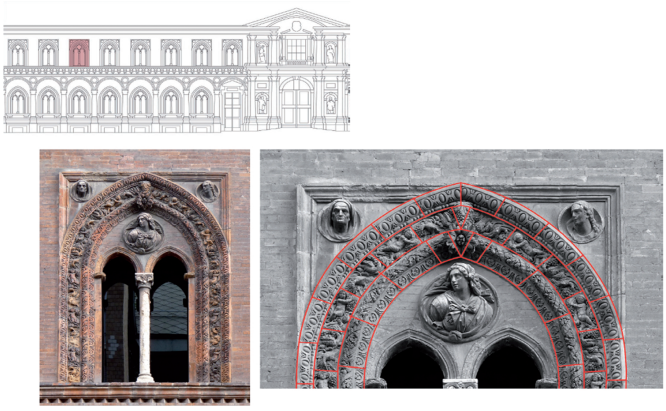


Fig. 15 – Schema con la localizzazione e particolari delle cornici di bifora seicentesca; sono evidenziate in rosso le formelle, tutte intere e posizionate rispetto a un asse centrale. Tratto dalle tavole della tesi PORTA-TODESCO, 2009-10.

[26] Il prof. Charles Morschech, che ringrazio pubblicamente, in visita al cantiere nel 2011, ha individuato per primo questa rara tipologia di formella.

[27] Per approfondimenti comparativi tra le formelle del cantiere di Milano e quello di Pavia si rimanda alla tesi G. PORTA, M. TODESCO, 2010, capitolo 9 *Gli apparati decorativi in terracotta delle facciate quattrocentesca e seicentesca*, pp. 63-35.

Fig. 16 – Comparazione tra la fascia marcapiano quattrocentesca e quella seicentesca. Tratto dalle tavole della tesi PORTA-TODESCO, 2009-10.

Fig. 17 – Comparazione di formelle in terracotta con i putti vendemmianti, parte seicentesca della facciata principale verso via Festa del Perdono.

Fig. 18 – Dettaglio fotografico di una formella della parte quattrocentesca con putto vendemmiante che presenta il volto rivolto leggermente a sinistra.

Lo stato di conservazione delle formelle era caratterizzato prevalentemente dalla presenza di depositi superficiali di colore scuro che spesso nascondevano il colore rosso delle superfici; dove invece era presente l'azione del dilavamento, il degrado si presentava con esfoliazioni, scagliature, fino a mancanza di porzioni che non permettevano più la leggibilità del soggetto modellato.

Tale casistica sottolinea come il problema dell'individuazione delle sostituzioni degli elementi non è un problema semplice. Da una osservazione ravvicinata emerge non tanto la difficoltà di lettura del dato singolo, quanto l'attribuzione certa della singola formella ad una specifica fase operativa, a causa della sovrapposizione di operazioni dettate da logiche e pratiche del fare in cantiere, mai documentate, di sostituzioni con possibile materiale di recupero, in analogia o similitudine.

L'analisi architettonica, le fasi costruttive, il repertorio di forme legato agli aspetti artigianali della produzione: serialità, unicità dei pezzi modellati o di statuaria si legano indissolubilmente alle storie dei restauri e delle manutenzioni. Merita segnalare l'unico caso riscontrato con uno strato finale di vetrina (*Fig. 19*).

Fig. 19 – Formella in terracotta invetriata presente nella parte seicentesca della facciata di via Festa del Perdono.
Si tratta dell'unico caso riscontrato di finitura tipo vetrina al piombo.

Inoltre, l'usanza nelle prassi manutentive o di restauro, di proteggere e conservare gli elementi in terracotta attraverso trattamenti di finitura, in parte nasconde e rende difficile il riconoscimento di altri indizi.[28] Ricordiamo infatti che, soprattutto tra la seconda metà dell'Ottocento e la prima metà del Novecento, si presenteranno esemplificazioni di prescrizioni atte al restauro degli edifici storici di epoca medievale-rinascimentale, nonché accorgimenti per rendere durevoli edifici in cotto di nuova costruzione.

Sulle formelle in terracotta ornamentale, sia quattrocentesche sia seicentesche, sono stati rilevati lacerti di quelle finiture di colore rosso, menzionate in letteratura, applicate al fine di omogeneizzare le tonalità del corpo ceramico e/o con scopi di protezione.[29] In molti casi il degrado del corpo ceramico non ha consentito la conservazione delle finiture su tutte le superfici: e inoltre l'estensione e l'omogeneità delle finiture osservate era, prima dell'intervento, di difficile valutazione a causa della presenza degli estesi e coprenti depositi grigi dei depositi[30] (*Fig. 20*).

Lo studio analitico di tali finiture è stato eseguito attraverso una campagna di tipo non invasivo, mediante fluorescenza a raggi X portatile in dispersione di energia (EDXRF), che ha portato a prelevare solamente 12 campioni studiati a mezzo di stratigrafie, con tecniche di labo-

---

[28] Convegno di Bressanone del 1992. *Le superfici dell'architettura: il cotto. Caratterizzazione e trattamenti,* a cura di G. BISCONTIN – D. MIETTO. Dal XIX sec. si presenteranno esemplificazioni di prescrizioni atte al restauro degli edifici storici di epoca medievale-rinascimentale, nonché accorgimenti per rendere durevoli e mantenere edifici in cotto di nuova costruzione, soprattutto tra la seconda metà dell'Ottocento e la prima metà del Novecento. Il caso di Bologna è tra i più significativi; il Rubbiani, infatti, ricorda che nei secoli XIV-XV-XVI era usanza dare una tinta rossa ai cotti (*La Certosa di Pavia e il suo museo. Ultimi restauri e nuovi studi,* da pag. 276), ripresa poi da Gino Chierici, Soprintendente a Milano dal 1935 al 1945. Tale prassi per la Certosa di Pavia è rimasta in voga fino al 1993.

[29] Si ricorda il caso dei cotti del chiostro piccolo della Certosa di Pavia, coevi per origine ai nostri quattrocenteschi. In particolare, per quanto riguarda la questione delle finiture e dei trattamenti si rimanda ai contributi della sezione *Diagnostica e restauri: rendiconti, metodologie e prospettive d'intervento sul Chiostro piccolo e sulle vetrate* in *La Certosa di Pavia e il suo museo. Ultimi restauri e nuovi studi,* da pag. 273.

[30] Anche a seguito della pulitura degli elementi, sempre il prof. Charles Morschech, nella medesima visita al cantiere del 2011, ha individuato l'esistenza di almeno due mani differenti per i medaglioni centrali delle bifore quattrocentesche, attribuendo a Francesco Solari quello della Fig. 20, dato che troverebbe riscontro anche nei documenti di archivio.

ratorio. Le indagini hanno rilevato che si tratta di due tipologie differenti di finitura rossa, anche se non distinguibili a occhio nudo e probabilmente frutto di due fasi distinte.[31]

Risulta quindi evidente che le finiture rosse rappresentano un aspetto caratterizzante le vicende costruttivo-manutentive della fabbrica, per altro non documentate nelle fonti d'archivio. Per questi motivi è risultata cruciale la scelta di conservarle integralmente, rendendo complessa la scelta di un metodo di pulitura in grado di garantire la massima permanenza di queste fragili pellicole, solubili all'acqua. Si segnala però che sulle finiture della Ca' Granda non risulta alcun rimando negli studi bibliografici più recenti,[32] come poco si trova in generale sugli intonaci[33] (*Fig. 21*).

Fig. 20 – Medaglione centrale della decima bifora, parte quattrocentesca della facciata verso via Festa del Perdono, possibile attribuzione a Francesco Solari. Prima e dopo l'intervento di pulitura.

Ai fini del ragionamento, preme ricordare che anche sulle terrecotte del chiostro piccolo della Certosa di Pavia erano evidenti tracce di finiture più antiche, nascoste durante l'intervento di restauro ottocentesco

31 Si rimanda al lavoro di Antonio Sansonetti "Il cantiere della Ca' Granda. La lettura dell'esperto scientifico".

32 A cura dell'autrice e altri, 2013.

33 Per quanto riguarda la decorazione della "porta e la colonnetta di una 'fenestra magna'", nei documenti di archivio risulta che nel 1459 fu dato incarico a Pietro Ambrogio de' Monti di eseguirla. (V. Biagetti, L'Ospedale Maggiore di Milano, Milano, 1937, pp. 54-55, nota 13 di FRANCHINI, 1995).

(1882-1884) in cui si è "creduto far cosa bella tinteggiare di rosso tutte queste terrecotte per rendere loro quella uniformità di colore che piace al volgo […]".[34] Paravicini infatti con queste parole contesterà anche in questo caso "i restauri e le inutili operazioni di rinnovamento".[35]

Fig. 21 – Portico della parte quattrocentesca verso via Festa del Perdono con lacerti di decorazioni pittoriche in corrispondenza dell'ingresso e intradosso degli archi (affreschi) e delle volte (tinteggiature).

3. Conclusioni

"L'abitudine a ricercare una visione totalizzante e generale dell'oggetto per coglierne immediatamente gli aspetti 'significanti' e arrivare quindi alla sua comprensione ci porta troppo spesso a trascurare i dettagli".[36] Questa frase di Maggi-Nasoni ben interpreta l'occhio privilegiato nell'intervento conservativo eseguito basandosi sulle osservazioni derivanti dal sistema di riferimento, dalle specifiche competenze che diventano raccolta di indicazioni per lavori specialistici futuri. Mi piace

34 BELLINI, 2000, p. 147 e ancora p. 201 dove in un rapporto alla SPAB, successivo al luglio 1881 si legge: "…Tinteggiature uniformi rosse sulle terrecotte che ne tolgono la loro naturale varietà di colore ed entrando fra i minimi particolari decorativi tolgono nettezza alle parti e ne rendono saturo il disegno …".

35 BELLINI, 2000, p. 61 e relativa NOTA 55.

36 MAGGI – NASONI, 1983, p. 15.

pensare che offrire aperture di senso può essere fatto anche attraverso il porre domande o portare osservazioni provvisorie e non risolutive, ponti tra l'opera conservata nella sua unicità e autenticità materiale e chi vuole ascoltarla.

RIFERIMENTI BIBLIOGRAFICI

[1] F. VAGLIENTI a cura di, *La Ca' Granda dei Milanesi. Itinerario interdisciplinare nel fulcro di una metropoli multiculturale, Nexo, Milano, 2014.*

[2] Per le vicende storiche e la storia dei restauri

P. CANETTA, *Cronologia dell'Ospedale Maggiore di Milano*, Milano, 1884.

P. PECCHIAI, *Guida dell'Ospedale Maggiore di Milano e degli Istituti annessi*, Milano, 1926.

L. GRASSI, *La Ca' Granda. Storia e restauro*, Cordani, Milano, 1958.

L. GRASSI, *Lo "Spedale di poveri" del Filerete. Storia e restauro*, Cordani, Milano, 1972.

L. GRASSI, *"…il glorioso albergo de' poveri di Dio, sotto Francesco Sforza, duca quarto di Milano"*, in *La Ca' Granda. Cinque secoli di storia e d'arte dell'ospedale Maggiore di Milano*, catalogo della mostra, Electa, Milano, 1981.

L. GRASSI, *Note sull'architettura del ducato sforzesco, in Gli Sforza a Milano e in Lombardia e i loro rapporti con gli stati italiani ed europei (1450-1530)*, Milano, 1982, pp. 449-468.

A. BELLINI, *Liliana Grassi*: *Un ricordo dopo dieci anni*, in "TeMA", N. 4, 1985, Franco Angeli, Milano, 1996, pp. 50-51.

M.A. CRIPPA a cura di, *Liliana Grassi Architetto il pensiero i restauri i progetti*, , Assicurazioni Generali, Catalogo della mostra, Milano, Stampa Zappa, Bollate (Mi),1986.

L. FRANCHINI, *"Ospedali lombardi del quattrocento. Fondazione, trasformazioni, restauri"*, New Press, Como, 1995.

M. CARLESSI, P. M. GALIMBERTI, A. KLUZER a cura di, *Il cuore dell'antico Ospedale Maggiore di Milano. I luoghi dell'Archivio e la Chiesa della beata Vergine Annunciata,* Silvana editoriale, Cinisello Balsamo (Mi), 2011.

M. CARLESSI, A. KLUZER a cura di, *Il cuore della Ca' Granda. Ricordi, scoperte e nuovi temi di storia e restauro nell'Ospedale Maggiore di Milano*, "Cheiron. Materiali e strumenti di aggiornamento storiografico", Bulzoni Editore, anno XXX-59, I semestre, 2013.

In particolare per le vicende del Novecento e i restauri

AA.VV*., La Ca' Granda di Milano. L'intervento conservativo sul cortile richiniano*, Progetto Monumenti Snam, Amilcare Pizzi, Milano, 1993.

F. SERRA, *I restauri alla Ca' Granda dal 1850 al 1969*, in *Milano Restaurata: il Monumento e il suo doppio*, G. GUARISCO a cura di, Alinea, Firenze, 1995, pp. 114-119.

G. VITAGLIANO, *Storia, restauro e progetto nell'attività di Liliana grassi. Un'operosità teoricamente fondata*, "Palladio", Anno XIX n. 38, 2006, pp. 101-128.

M.A. CRIPPA, E. SORBO, *Liliana Grassi. Il restauro ed il recupero creativo della memoria storica, Monsignori editori,* Roma, 2007.

G. VITAGLIANO, *Una storia del restauro in corpore vili. Gli interventi all'Ospedale Maggiore di Milano nella seconda metà del Novecento*, in AA.VV., *Restauro, monumenti e città. Teorie ed esperienze del Novecento in Italia,* Electa, Napoli, 2008, pp. 144-199.

M. BORIANI, *La teoria e la prassi del restauro italiano nel secondo dopoguerra*, in *Progettare per il costruito. Dibattito teorico e progetti in Italia nella seconda metà del XX secolo*, in M. BORIANI a cura di, Città Studi edizioni, De Agostini, Novara, 2008, pp. 3-28.

F.B. CAVALLERI, *Dai principi teorici alle realizzazioni. Tre livelli di riflessione*, in *Progettare per il costruito. Dibattito teorico e progetti in Italia nella seconda metà del XX secolo*, in M. BORIANI a cura di, Città Studi edizioni, De Agostani, Novara, 2008, pp. 61-69.

F.B. CAVALLERI, *Ospedale Maggiore, Milano: sede dell'Università Statale*, in *Progettare per il costruito. Dibattito teorico e progetti in Italia nella seconda metà del XX secolo*, M. BORIANI a cura di, Città Studi edizioni, De Agostani, Novara, 2008, Scheda due, pp. 131-138.

Per i repertori decorativi

L. MAGGI, M.C. NASONI, *Per l'analisi del repertorio decorativo tardoquattrocentesco a Milano: l'Ospedale Maggiore*, in Università di Pavia – Istituto di Storia dell'arte, *La scultura decorativa del Primo Rinascimento*, Atti del I Convegno internazionale di studi, Pavia 16-18 settembre 1980, Roma, 1983, pp. 15-27.

I. BALESTRERI, *Lo studio della decorazione in terracotta in area lombarda: un intreccio di problemi per il primo rinascimento*, in M.C. LOI, L. PATETTA, *Tradizioni e regionalismi nel primo rinascimento italiano*, Edizioni Unicopi, Milano, 2005, pp. 59-69.

M.C. LOI, *La decorazione in terracotta nei grandi cantieri lombardi*, in M.C. LOI, L. PATETTA, *Tradizioni e regionalismi nel primo rinascimento italiano*, Edizioni Unicopi, Milano, 2005, pp. 47-57.

Per finiture e trattamenti del cotto

G. BISCONTIN, D. MIETTO a cura di, *Le superfici dell'architettura: il cotto. Caratterizzazione e trattamenti*, *A*tti del Convegno di Bressanone, Padova, 1992.

A. BELLINI, *Tito Vespasiano Paravicini*, Guerini Studio, Milano, 2000.

S. ARDIZIO, F. ARIINI, *La conservazione del cotto all'aperto e in ambiente confinato: due casi-studio milanesi*, Tesi di Laurea, relatore S. DELLA TORRE, co-relatori M. REALINI, C. COLOMBO, Politecnico di Milano, Facoltà di Architettura e Società, A.A. 2002-2003.

B. BENTIVOGLIO RAVASIO, L. LODI, M. MAPELLI a cura di, *La Certosa di Pavia. Ultimi restauri e nuovi studi*, Atti del Convegno, Ministero per i Beni e le Attività Culturali. Direzione Regionale per i Beni Culturali e Paesaggistici della Lombardia, Milano, 2008.

G. PORTA, M. TODESCO, *La valutazione degli interventi conservativi sulla facciata seicentesca della Ca' Granda. L'esperienza dell'intervento di restauro nel cantiere sperimentale del I lotto*, Tesi di Laurea, relatore S. DELLA TORRE, Co-relatori R. FANT, A. SANSONETTI, Politecnico di Milano, Facoltà di Architettura e Società, Corso di Laurea in Scienze dell'Architettura, A.A. 2009-2010.

R. FANT, C. COLOMBO, A. SANSONETTI, *L'apparato decorativo in terracotta della Ca' Granda: indagini conoscitive e intervento di restauro,* in M.G. ALBERTI OTTOLENGHI, L. BASSO a cura di, *Terrecotte nel Ducato di Milano. Artisti e cantieri del primo Rinascimento,* Atti del Convegno 17-18 ottobre 2011, Milano e Certosa di Pavia, edizioni ET, Milano, 2013, pp. 253-270.

Istituto Lombardo (Rend. Scienze) 146, 93-112 (2012)

IL CANTIERE DELLA "CA' GRANDA". LA LETTURA DELL'ESPERTO SCIENTIFICO

ANTONIO SANSONETTI (*)

Nota presentata dal m.e. Amedeo Bellini
(Adunanza del 22 marzo 2012)

SUNTO. – Il lungo intervento di conservazione condotto sui fronti esterni della "Ca' Granda" in Milano ha consentito di condurre un cantiere sperimentale, durante il quale è stato possibile valutare gli effetti di alcuni metodi di pulitura e di applicazioni di trattamenti protettivi e consolidanti. Questo lavoro è stato preceduto da una approfondita caratterizzazione delle superfici in terracotta che ha messo in evidenza l'esistenza di due diverse finiture rosse, al di sotto dei depositi concrezionati; le finiture sono composte da gesso e colla di coniglio la prima, olio siccativo e biacca la seconda, e sono entrambe pigmentate con ocre. Mettendo in fila l'insieme delle osservazioni, macroscopiche, microscopiche e composizionali, si è pervenuti, alla scelta del metodo di pulitura che utilizza una sorgente impulsata laser Nd:YAG con emissione a 1064 nm. In questo modo è assicurata la maggiore conservazione dei materiali di finitura e dei microframmenti in corso di distacco a causa dei fenomeni di degrado, in alcuni casi particolarmente avanzati.

***

ABSTRACT. – The conservation works on the external facades of the so-called "Ca' Granda" allowed to carry out a pilot yard, during which several cleaning methods have been tested; moreover it was possible to apply water-repellent and consolidation treatments, in order to monitor their performances. This survey was preceded by a complete characterization as regards Terracotta surfaces, which highlighted the presence of two different red finishes: the first was composed by gypsum and rabbit glue and second by siccative oil and lead white; both were pigmented with red ochre. Collecting all the observations, by a macroscopic and microscopic point of view, the choice of laser cleaning proved to be the most suitable. It used a Nd:YAG pulsed source, emitting at 1064 nm. Doing so, the assurance of the most complete conservation has been obtained, taking into account the particularly worrying state of conservation of the substrate.

(*) Istituto per la Conservazione e Valorizzazione dei Beni Culturali. Consiglio Nazionale delle Ricerche.
E-mail: sansonetti@icvbc.cnr.it

1. Introduzione

La Conservazione dei Beni Culturali diviene progressivamente, pur con molte difficoltà, un campo di azione multidisciplinare. Il riferimento alle difficoltà nasce da considerazioni attinenti i diversi linguaggi e metodi di lavoro che si vedono operare nel campo. E' in questo ambito che le scienze sperimentali come la Fisica, la Chimica, e la Biologia hanno collaborato a quell'obiettivo comune cui tende il processo della Conservazione di un Bene Culturale. Una professionalità con competenze scientifiche può quindi contribuire grandemente a questo processo: la conoscenza materica del Bene, lo studio del suo stato di conservazione, la scelta dei prodotti e dei metodi più appropriati per l'intervento, sono tutti percorsi che si avvalgono di tali competenze. Il caso del cantiere della Ca' Granda è risultato paradigmatico a questo proposito e ha permesso una buona confluenza delle diverse discipline chiamate alla collaborazione per una sua corretta conservazione. Le ragioni di tutto questo risiedono nei giusti tempi messi a disposizione per gli studi preliminari, nel favorire un continuo scambio tra l'apporto della diagnostica e le scelte progettuali, nei comuni intenti di tutte le parti coinvolte, in particolare modo per quanto concerne la divulgazione dei risultati conseguiti.

2. Il rilievo dello stato di conservazione

Un corretto rilievo dello stato di conservazione risulta utile nelle scelte progettuali di intervento, sia per quanto riguarda i metodi di pulitura, sia per ciò che concerne la scelta dei prodotti di conservazione e del metodo per la loro applicazione. Solo in questo modo il computo delle ore lavorative e dei costi dell'intervento potrà basarsi sui di un set di dati il più possibile oggettivo e consentire il rispetto dei tempi e del budget.

Lo stato di conservazione delle terrecotte ornamentali e dei mattoni delle facciate esterne della Ca' Granda era caratterizzato, prima dell'intervento, prevalentemente dalla presenza di abbondanti depositi superficiali di colore scuro che nascondevano, spesso quasi totalmente, il colore rosso delle superfici. I depositi sono definiti dalla Normativa UNI Beni Culturali 11182, come accumuli di materiali estranei, quali polvere, terriccio, guano, ecc. avente spessore variabile, generalmente scarsa coerenza e scarsa aderenza al materiale sottostante. Lo spessore dei depositi rag-

giungeva valori di circa 450 micron e risultava costituito da composti di zolfo e calcio oltre che da particelle carboniose e silicati (*Fig. 1*).

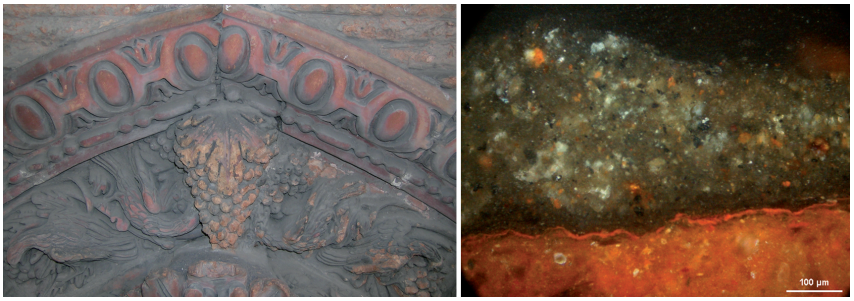


Fig. 1 – Deposito superficiale su terrecotte ornamentali. A destra: sezione lucida trasversale osservata al microscopio ottico in luce riflessa. Si osserva il corpo ceramico come substrato e i soprastanti strati di deposito.

In alcuni punti il deposito superficiale non era presente ed il corpo ceramico risultava a vista a causa dell'avanzare di esfoliazioni, e di scagliature, generalmente in corrispondenza di soluzioni di continuità del materiale originario.

In alcune porzioni dei paramenti era possibile individuare la perdita di frammenti di materiale ceramico; in particolare sono stati identificati fenomeni di disgregazione (*Fig. 2*) fino ad arrivare alla mancanza di elementi tridimensionali.

Fig. 2 – Disgregazione a carico della porzione di una cornice.

Gli interventi che negli anni interessarono le superfici compresero la sostituzione di singoli elementi ceramici, sia nel paramento murario che, ancor più, nelle parti di ornato. Si vedono quindi oggi numerose formelle realizzate con materiali ceramici analoghi dal punto di vista

composizionale rispetto ai materiali originari, ma anche sostituzioni effettuate con malte tradizionali o cementizie.

L'analisi architettonica, le fasi costruttive, il repertorio di forme legato agli aspetti artigianali della produzione, la sua serialità da un lato e l'unicità dei pezzi modellati o di statuaria dall'altro, si legano indissolubilmente alle storie dei restauri.

Inoltre, l'usanza di proteggere e conservare gli elementi in terracotta nelle prassi manutentive o di restauro, attraverso trattamenti di finitura, in parte nasconde e rende difficile il riconoscimento di altri indizi. Ricordiamo infatti che, soprattutto tra la seconda metà dell'Ottocento e la prima metà del Novecento, si presenteranno esemplificazioni di prescrizioni atte al restauro degli edifici storici di epoca medievale-rinascimentale, nonché accorgimenti per rendere durevoli e mantenere edifici in cotto di nuova costruzione.

3. Studio e caratterizzazione dei materiali costitutivi

3.1 Caratterizzazione dei materiali ceramici

Allo stato della ricerca i metodi derivati dalle scienze sperimentali, costituiti da strumenti e analisi precisi e sensibili, consentono una lettura approfondita, di taglio nuovo, ma non sono neutrali. L'unicità del dato, la sua fragilità, gli errori sistematici e casuali, nonché la frammentarietà e relatività delle informazioni ci devono mettere in guardia su ciò che viene definito "oggettività della diagnostica", che invece deve essere sempre interpretata e valutata nella sua complessità e specificità.

Lo studio che qui si presenta ha riguardato sia le decorazioni in terracotta ornamentale, che gli apparati in mattoni, e si è focalizzato su alcune superfici di finitura di colore rosso, ben individuabili ad una distanza ravvicinata. I dati raccolti dalla campagna diagnostica sono stati confrontati con quanto emerso da precedenti indagini condotte dall'allora Centro Gino Bozza nel 1993,[1] permettendo da un lato di concentrare l'attenzione sulle porzioni di facciata non ancora studiate, dall'altro di ispezionare le stesse superfici a distanza di anni.

[1] Il Centro CNR Gino Bozza ha dato origine, con i gemelli centri di Roma e Firenze, all'attuale ICVBC del CNR.

In *Fig. 3* è riportato uno schema relativo alle aree indagate sui singoli lotti oggetto di studio e le aree oggetto della precedente campagna diagnostica.

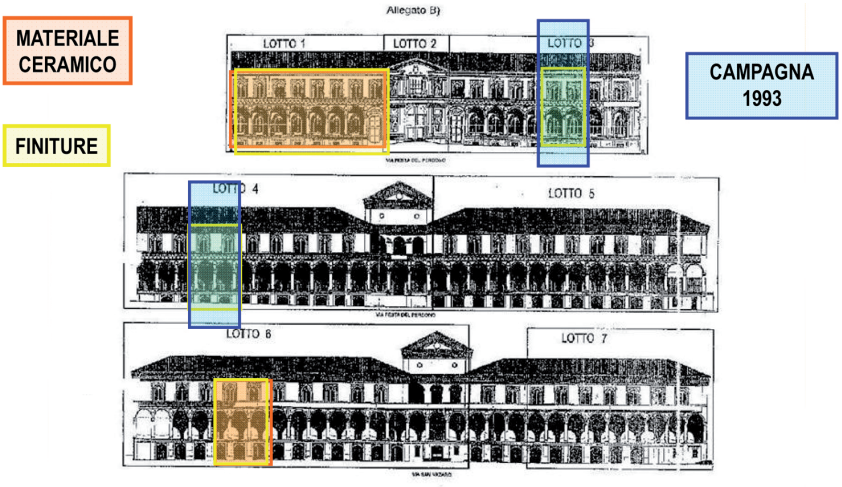


Fig. 3 – Schema delle aree oggetto di indagini diagnostiche.

Lo studio delle superfici di finitura è stato condotto sia per mezzo di metodi non invasivi (fluorescenza di raggi X portatile in dispersione di energia, EDXRFp), sia mediante tecniche di laboratorio più tradizionali.

Le analisi di laboratorio sono state condotte su circa cinquanta microframmenti prelevati su mattoni e terrecotte ornamentali delle aree di analisi indicate in *Fig. 3*. I campioni sono stati osservati allo stereomicroscopio per la programmazione del piano analitico (stereomicroscopio Leitz Wild M420 corredato di sistema per l'acquisizione di immagini in formato digitale/telecamera Nikon DS-5M/USB); le osservazioni sono state approfondite mediante microscopia ottica su sezioni sottili per la caratterizzazione del corpo ceramico (microscopio polarizzatore Nikon Eclipse E400 Pol), microscopia ottica su sezioni trasversali lucide (microscopio ottico Leitz Ortholux in luce riflessa) per lo studio delle superfici di finitura, diffrattometria a raggi X su polveri per la definizione mineralogica dei materiali e dei prodotti di degrado (Diffrattometro a raggi X Panalytical X'Pert PRO), fluorescenza a raggi X su polveri per la definizione della composizio-

ne elementare del corpo ceramico,[2] cromatografia ionica per la definizione del contenuto in sali solubili.[3]

Infine studi stratigrafici e morfologici sono stati condotti mediante analisi al microscopio elettronico a scansione corredato da spettrometro in dispersione di energia (SEM+EDS),[4] mentre la definizione della componente organica, ove presente, è stata indagata mediante analisi spettroscopica FTIR.[5]

Lo studio del materiale ceramico ha mostrato che l'impasto ha una analoga composizione per i mattoni pieni e per le formelle in terracotta. Esso presenta differenti aspetti macroscopici caratterizzati da una colorazione omogenea rossa, oppure da un colore crema; in alcuni casi e per quello che riguarda le terrecotte ornamentali, si osserva un aspetto che presenta le due colorazioni mescolate in una fiammatura.

---

[2] Le polveri (opportunamente seccate in stufa e calcinate) sono state miscelate con $Li_2B_4O_7$ in rapporto di peso costante 1:4. Le miscele ottenute sono state fuse in crogioli di platino utilizzando una perlatrice automatica con bruciatore a metano/ossigeno e quindi colati in un piattino di platino del diametro di 40 mm. I dischi vetrosi così ottenuti sono stati analizzati mediante la tecnica della spettrometria di fluorescenza a raggi X a dispersione di lunghezza d'onda (WD-XRF) utilizzando uno spettrometro a raggi X sequenziale Philips PW 2400 con tubo al Rodio calibrato con campioni standard internazionali certificati e campioni interlaboratori di rocce, minerali e materiali ceramici.

[3] Le analisi sono state eseguite mediante un cromatografo ionico Dionex DX 100 utilizzando la colonna Dionex Ion PacAS4A con una miscela $Na_2CO_3/NaHCO_3$ quale eluente per la separazione degli anioni e la colonna Dionex Ion Pac CS12 con una soluzione di acido metansolfonico quale eluente per la separazione dei cationi. I campioni in polvere (essiccati in stufa a 60°C) sono stati solubilizzati in acqua UHQ (Ultra High Quality avente resistenza non inferiore a 18MΩ) nel rapporto 100 mg di polvere in 100 ml di acqua. La soluzione è stata filtrata prima dell'introduzione in colonna.

[4] Le osservazioni morfologiche e le analisi spettrometriche sono state condotte mediante microscopio elettronico a scansione JEOL 5910 LV corredato di spettrometro a raggi X in dispersione di energia IXRF-2000. Gli studi morfologici sono stati condotti mediante osservazioni con elettroni retrodiffusi (BES) su sezioni trasversali lucide opportunamente rese conduttrici mediante deposizione di un film di grafite sulla superficie. La natura composizionale degli strati pigmentati è stata indagata mediante l'acquisizione di spettri e mappe EDS tra 0 e 20 keV sulle stesse sezioni trasversali lucide.

[5] Fourier Transform InfraRed Spectroscopy; le analisi sono state condotte mediante spettrofotometro Nicolet Nexus corredato di microscopio Continuμm. Gli spettri sono stati acquisiti in cella di diamante lavorando in trasmissione con detector MCT nell'intervallo di frequenze da 4000 a 650 cm^{-1}.

L'argilla impiegata per la realizzazione dei mattoni pieni e per la produzione delle formelle è di natura non carbonatica; dal punto di vista mineralogico si registra una preponderanza della componente silicatica in termini di quarzo e feldspati, sia sodici che potassici, la presenza di ematite, illite e clorite quali minerali argillosi, e di anfibolo. Il degrado delle malte adiacenti ha generato gesso come prodotto di neoformazione, che si è cristallizzato nella porosità del corpo ceramico. La temperatura di cottura può essere ipotizzata attorno ad un valore medio non superiore a circa 900°C.

I dati raccolti concordano con quanto rilevato durante la campagna di indagini del 1993, sia per quanto concerne la composizione del materiale, sia per le tecniche di lavorazione. Per quanto riguarda i dati ottenuti dall'analisi delle formelle, si può ipotizzare l'uso di impasti appositamente formulati per conferire una maggiore plasticità al materiale crudo. I differenti valori nei contenuti in allumina sembrerebbero indicare che l'argilla utilizzata è stata soggetta ad una depurazione durante i processi di preparazione del materiale (*Tab. 1*).

Tab. 1 – Composizioni elementari medie espresse in ossidi, in percentuale in peso, di mattoni e terrecotte.

Ossidi	**Lotto 4** Quattrocento (campagna 1993)		**Lotto 3** Seicento (campagna 1993)		**Lotto 6** Quattrocento (via San Nazzaro)		**Lotto 1** Seicento (via Festa del Perdono)	
	T (n=9)	M (n=11)	T (n=17)	M (n=11)	T (n=4)	M (n=4)	T (n=7)	M (n=5)
SiO_2	71.03	72.53	71.21	71.72	68.39	69.09	69.87	70.20
Al_2O_3	14.80	14.30	16.56	14.70	15.17	14.68	17.21	16.24
TiO_2	0.79	0.80	0.88	0.79	0.83	0.79	0.96	0.88
Fe_2O_3	5.33	5.01	4.65	5.20	4.67	4.90	4.04	4.94
MnO	0.13	0.10	0.05	0.12	0.08	0.09	0.04	0.07
MgO	2.23	1.88	1.15	1.03	1.87	1.85	1.35	1.46
CaO	1.92	1.67	1.38	1.75	2.43	1.60	2.62	2.04
Na_2O	1.40	1.50	1.79	1.82	1.63	1.67	1.55	1.49
K_2O	2.21	2.06	2.21	2.68	2.08	2.43	2.05	2.36
P_2O_5	0.16	0.15	0.12	0.19	0.11	0.14	0.09	0.16
P.F.	1.91	1.39	4.24	1.71	2.62	2.62	5.99	2.02

T=terrecotte ornamentali; M=mattoni (tra parentesi il numero di campioni analizzati). Analisi XRF di laboratorio; i dati sono stati normalizzati escludendo il valore di perdita al fuoco (P.F.).

3.2 I materiali di finitura

Sulle formelle in terracotta ornamentale, sia quattrocentesche che seicentesche, sono stati osservati visivamente residui di finiture di colore rosso; è ragionevole pensare che tali finiture siano state applicate al fine di omogeneizzare le tonalità del corpo ceramico e con scopi di protezione.[6] Il degrado del corpo ceramico non ha consentito la conservazione delle finiture su tutte le superfici. L'estensione e la omogeneità delle finiture osservate era, prima dell'intervento, di difficile valutazione a causa della presenza degli estesi e coprenti depositi grigi di cui si è già discusso. La presenza o meno della finitura può essere utilizzata come indizio per comprendere l'epoca di utilizzo dell'elemento che costituisce il substrato delle finiture stesse, pur con tutte le riserve cui si accennava sopra.

Lo studio analitico di tali finiture è stato eseguito dapprima mediante una campagna di tipo non invasivo mediante fluorescenza a raggi X portatile in dispersione di energia (EDXRF).[7] I punti di misura hanno rilevato, prevalentemente nel lato seicentesco, la presenza di piombo. I risultati di questo screening iniziale hanno permesso di ottimizzare il prelievo di campioni, riducendolo al minimo. Sui 12 campioni prelevati sono stati condotti studi stratigrafici mediante tecniche di laboratorio quali: osservazioni al microscopio ottico in luce polarizzata su sezioni trasversali lucide, microscopia elettronica a scansione corredata da microsonda elettronica in dispersione di energia (EDX), spettroscopia infrarossa (FTIR).

Le osservazioni microscopiche hanno rilevato strati di colore rosso, aventi spessore intorno ai 30-40 μm, aderenti al corpo ceramico e carat-

---

[6] Si ricorda il caso dei cotti del chiostro piccolo della Certosa di Pavia, coevi per origine ai nostri quattrocenteschi.

[7] E' stato utilizzato uno spettrometro XRF portatile (serie ALPHA 4000, Innov-X) con micro tubo a raggi X avente anodo al tantalio. L'analisi viene eseguita in due step con due energie diverse: 30KeV - 6,5 μA – filtro 2mm di alluminio per la determinazione degli elementi pesanti e 15KeV – 7 μA – filtro 0,1 mm di alluminio per la determinazione degli elementi leggeri. L'area irraggiata è di circa 155 mm^2. Lo spettrometro è equipaggiato con un rivelatore Si-PIN (FWHM <230 eV a 5.95 KeV linea Kα del Mn) raffreddato con una cella Peltier. Lo strumento viene appoggiato alla superficie da analizzare mantenendo costante la distanza del tubo e del rilevatore, posizionati ad un'angolazione fissa all'interno della testa dello strumento e protetti da una finestra di Kapton. La costanza delle condizioni geometriche è garanzia della riproducibilità delle misure e quindi della confrontabilità dei dati. Gli spettri XRF sono stati registrati in via digitale, tramite un palmare integrato e software dedicato.

terizzati occasionalmente dalla presenza di crettature perpendicolari all'andamento dello strato stesso; la loro composizione risulta a base di gesso ed ocra rossa, (spettro EDS *Fig. 4*). Gli spettri FTIR hanno consentito di individuare sostanze proteiche (probabilmente uovo o caseina) e olio (*Fig. 5*) Lo studio delle componenti organiche impiegate per la stesura delle superfici di finitura, è ancora oggetto di indagine. Si tratta di un approfondimento rispetto a quanto già definito nella campagna del 1993, dove l'olio era risultato il componente organico principale. Rimane comunque una ricerca complessa, non scevra da dubbi e perplessità, anche a causa della avvenuta mineralizzazione della maggior parte delle componenti organiche in ossalati di calcio. Sicuramente su superfici così porose ed esposte all'aperto, era prassi comune "trattare" le superfici in cotto con sostanze oleose o proteiche per "impermeabilizzarle". Così come l'applicazione di trattamenti era pensato per riaggregare il materiale, sorta di consolidamento e/o "ravvivare" il colore o omogeneizzare gli elementi cromaticamente disomogenei.

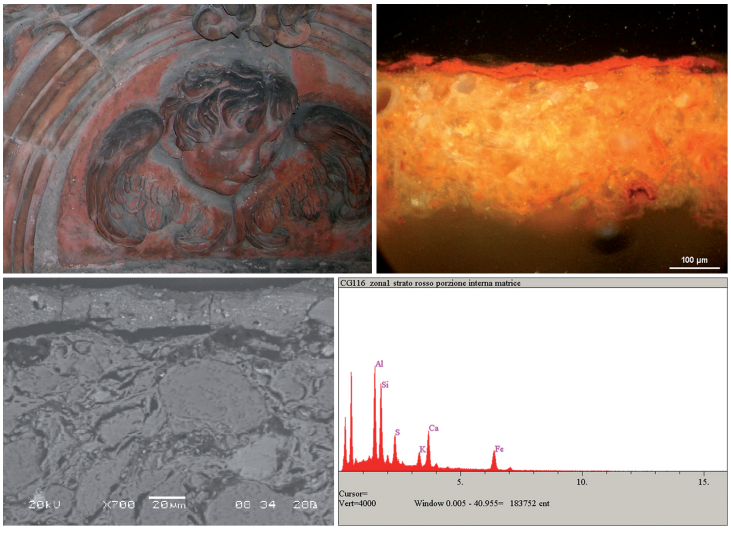


Fig. 4 – In alto a sinistra area di campionamento, a destra sezione lucida trasversale osservata al microscopio ottico in luce riflessa. In basso a sinistra sezione lucida trasversale osservata al microscopio elettronico a scansione in modalità low vacuum (28Pa). Lo strato esterno mostra discontinuità perpendicolari all'andamento dello strato di finitura rosso. A destra spettro EDS registrato in corrispondenza dello strato rosso. Si distinguono oltre ai picchi del corpo ceramico anche calcio, zolfo e ferro, propri dello strato rosso esterno.

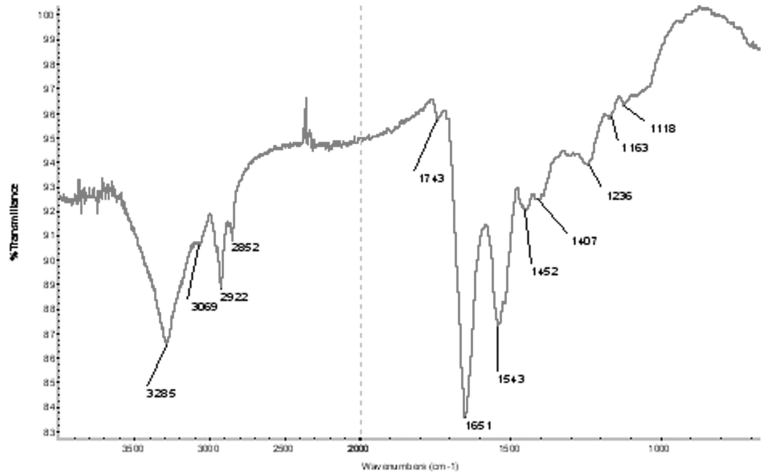


Fig. 5 – Spettro FTIR registrato in corrispondenza dello strato di finitura. Residuo solido dopo l'estrazione in esano – cella di diamante, si notano i segnali attribuibili ad una sostanza proteica (probabilmente uovo), o ad una miscela di caseina e olio.

L'altra tipologia di finiture rosse mostra uno strato di spessore piuttosto regolare pari a circa 30 micron; essa si compone di biacca ed olio siccativo, probabilmente olio di lino. La sostanza pigmentante è ematite Fe_2O_3 per entrambe le finiture. Nelle porzioni di campione indagato, lo stato di adesione tra finiture e corpo ceramico è buono.

Risulta quindi evidente che le finiture rosse rappresentano un aspetto caratterizzante le vicende costruttivo-manutentive della fabbrica, per altro non documentate nelle fonti d'archivio. Per questi motivi è risultata cruciale la loro conservazione, rendendo complessa la scelta di un metodo di pulitura in grado di garantire la massima permanenza di queste fragili testimonianze.

4. Valutazione degli interventi conservativi

4.1 Obiettivi del cantiere pilota

Esiste una esigenza riconosciuta di anticipare con un cantiere pilota le risultanze di un intervento conservativo, avvalendosi di una prima fase di conoscenza e di studio; in questa fase la valutazione degli effetti dei metodi e dei prodotti di conservazione, dovrebbe essere con-

dotta con indagini strumentali e di laboratorio. Tale esigenza, non ancora sufficientemente diffusa nella prassi del cantiere, ha trovato nel caso della Ca' Granda una felice e virtuosa realizzazione in una collaborazione sinergica, consentendo di testare gli interventi ed in particolare i metodi di pulitura, i prodotti per il consolidamento superficiale e per il trattamento di protezione idrorepellente.

La successione temporale dei lotti di intervento della Ca' Granda, che si sono sviluppati in un periodo di circa 4 anni, ha consentito di monitorare le prestazioni dei prodotti utilizzati nella conservazione in condizioni naturali di invecchiamento e per un tempo piuttosto lungo di circa due/tre anni. Questa possibilità è piuttosto infrequente e di norma ci si limita ad applicare i prodotti e ad ispezionare lo stato di conservazione nei tempi programmati.

Il progetto diagnostico è stato quindi svolto in piena e proficua sinergia con i progettisti dell'intervento, che hanno operato la scelta finale dei metodi in base alle risultanze del cantiere di studio.

Per quanto concerne la pulitura si ricorda brevemente che essa è una operazione chiamata a rimuovere quanto si identifica come estraneo rispetto alla superficie del manufatto e che inoltre si ritiene possa danneggiare il substrato e la sua buona conservazione. La pulitura ha anche lo scopo di rendere meglio leggibile la morfologia del manufatto stesso, i suoi effetti chiaroscurali e decorativi. Nel presente scritto si intende con consolidamento superficiale l'applicazione di un prodotto o di un formulato che abbia l'obiettivo di restituire la coesione perduta alla microstruttura dei materiali in opera. Questo effetto viene ottenuto applicando il prodotto consolidante con l'accortezza che penetri secondo quello che viene definito un gradiente di concentrazione. Infine l'intervento di protezione della superficie applica dei prodotti idrorepellenti in modo da limitare l'accesso di acqua liquida alla superficie stessa, con l'obiettivo di rallentare le cinetiche del degrado nel cui meccanismo l'acqua possiede un ruolo molto importante.

In corrispondenza del lotto 1 (corpo di fabbrica sinistro guardando il portale di ingresso seicentesco in Via Festa del Perdono – *Fig. 3*) è stato allestito il cantiere pilota. E' stata scelta una area del paramento in laterizio e un area adiacente delle cornici delle bifore decorate con formelle di terracotta su cui sono stati valutati cinque differenti metodi di pulitura, i prodotti conservativi e i loro metodi di applicazione. Le zone dei test sono state scelte in modo che avessero omogenee condizioni di conservazione, così da introdurre elementi di maggiore obietti-

vità nella valutazione delle prestazioni di metodi e prodotti differenti. Di seguito sono descritte le diverse metodologie di pulitura testate, i loro parametri, nonché i risultati ottenuti.

- laser EOS 1000; sorgente Nd:YAG; durata dell'impulso laser di circa 60-130 micro-secondi. fluenza operativa media 1.35 J/cm^2, frequenza di ripetizione 20 Hz; diametro dello spot utilizzato 3,5 mm.
- aero-abrasivatura Rotec; abrasivo Mikhart 130 costituito per oltre il 99% di $CaCO_3$; granulometria media 140 micron, pressione di esercizio 0.2 bar;
- lavaggio con acqua nebulizzata - 5 cicli da 5 minuti ciascuno;
- impacco di carbonato di ammonio (soluzione satura); tempo di applicazione 3 ore;
- impacco di EDTA (soluzione acquosa al 7%); tempo di applicazione 16 minuti.

Le superfici in terracotta presentano un rilievo accentuato che non consente l'esecuzione di indagini non invasive di cantiere per la valutazione dei metodi di pulitura, che necessitano di superfici piane per il posizionamento degli strumenti. Pertanto la valutazione è stata effettuata mediante l'esame in laboratorio di campioni appositamente prelevati dalle superfici, tramite osservazioni microscopiche e indagini spettroscopiche. Da ogni area di pulitura è stato prelevato un frammento per indagare in sezione trasversale le differenze tra superficie pulita e superficie sporca. Il deposito compatto sui supporti ceramici presi in esame è costituito essenzialmente da silicati, particelle carboniose e gesso.

La formella pulita mediante soluzione satura di carbonato di ammonio (*Fig. 6*), presenta vaste aree dove il risultato ottenuto deve essere considerato insoddisfacente, soprattutto in corrispondenza dei sottosquadri dove l'impacco ha probabilmente faticato ad aderire; in alcune delle stesse aree di sottosquadro si osservano sbiancamenti diffusi. La superficie pulita appare schiarita come se fosse interessata da precipitati in forma di minuti cristalli bianchi: si tratta di un accumulo di sali solubili di spessore variabile attorno ai 100 micron che sono precipitati immediatamente al di sotto della superficie esterna. Permangono numerose particelle scure, raccolte in fondo ad incavi presenti sulla superficie e residui di fibre appartenenti al materiale spessente utilizzato per l'impacco (*Fig. 7*).

In alcune aree dei campioni prelevati nella parte non ancora pulita si individuano lacerti della finitura rossa composta da biacca ed olio sic-

cativo. E' infatti ben visibile uno strato chiaro nella immagine ottenuta mediante elettroni retro diffusi; questa modalità di acquisizione dell'immagine consente di mettere in mettere in evidenza le parti dove sono localizzati gli elementi ad alto peso atomico, che vengono visualizzate come molto brillanti (*Fig. 8*). Al termine della pulitura la finitura rossa è presente solo in minime tracce; a tale proposito si può ipotizzare che la finitura a biacca ed olio sia stata rimossa da un meccanismo di idrolisi che ha interessato la componente oleosa che accompagna la biacca, il cui attacco da parte del carbonato d'ammonio, è stato favorito dai lunghi tempi di contatto (3 ore). Il metodo testato non si è quindi dimostrato selettivo ma, al contrario rimuove sia il deposito grigio costituito da gesso, che la finitura. Per questo motivo il metodo è stato scartato.

Fig. 6 – Sinistra: formella dopo pulitura con Carbonato d'Ammonio. Destra: formella dopo pulitura con Rotec.

Per quanto concerne la formella pulita con il sistema brevettato Rotec, ad una osservazione macroscopica esso sembra avere raggiunto lo scopo di rimuovere la maggior parte dei depositi compatti (*Fig. 6*); tale obiettivo non viene però realizzato completamente. Una ulteriore osservazione a livello macroscopico interessa il tono cromatico generale della formella pulita con Rotec, che risulta in parte desaturato. Se si approfondiscono le osservazioni e si indaga mediante stereo-microscopia, si evince che il deposito non è stato rimosso completamente, neppure da quelle aree che macroscopicamente sembravano avere raggiun-

to un livello di pulitura soddisfacente. Permangono infatti ancora numerosi residui di sporco ben visibili nel fondo di micro-solchi presenti sulla superficie ceramica.

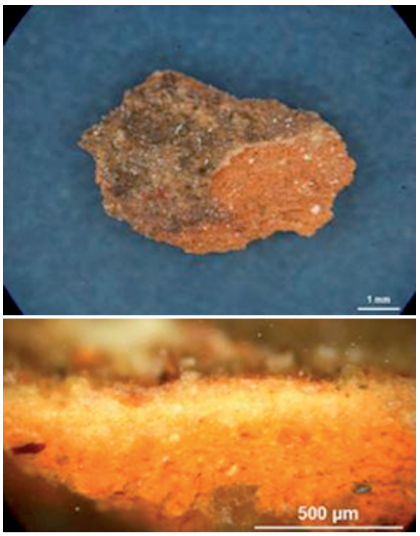


Fig. 7 – Sopra osservazione allo stereomicroscopio della superficie pulita con carbonato di ammonio dopo pulitura; sotto osservazione al microscopio ottico in luce riflessa di sezione lucida trasversale. Superficie pulita; si osserva la morfologia della formazione sub-superficiale.

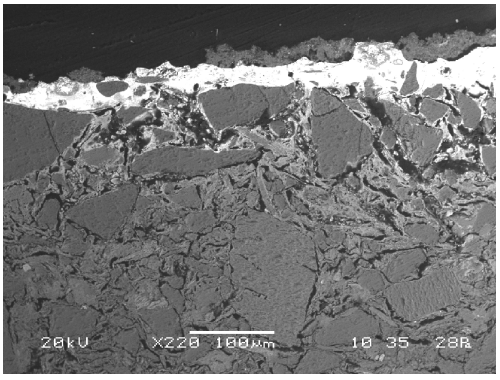


Fig. 8 – Osservazione in elettroni retrodiffusi (BSE). Area di indagine EDS. Sezione trasversale di un campione proveniente da una zona non ancora pulita; si osserva lo strato di finitura realizzato con un composto di piombo (strato chiaro) e, al di sopra, un deposito di gesso.

Il sistema ha rimosso il deposito superficiale, ma al tempo stesso non si osservano tracce della finitura, neanche in forma residuale, che unitamente a segni sulla superficie esterna, lasciano ipotizzare una abrasione spinta.

In *Fig. 9* è mostrato un dettaglio della formella pulita mediante laser Nd:YAG: si osserva la presenza di lacerti della finitura di colore rosso intenso al di sopra del corpo ceramico. Il laser consente una delimitazione molto precisa delle aree pulite, e conseguentemente è stato possibile prelevare dei frammenti che includono sia una parte pulita, che una non pulita. Nelle porzioni non pulite il deposito si presenta compatto, sottile e di colore grigio; sottostante ad esso è presente la finitura rossa. In corrispondenza delle aree pulite la finitura viene conservata. Le indagini hanno evidenziato che, in questo caso, la finitura rossa ha una composizione gessosa, contiene ossalato di calcio ed è pigmentata con ematite.

Fig. 9 – Sinistra: formella dopo pulitura con Laser.
Destra: formella dopo pulitura con EDTA.

La pulitura mediante EDTA non ha fornito risultati soddisfacenti riguardo alla rimozione del particellato (*Fig. 9*). L'EDTA è un complessante dell'atomo di calcio che mostra il massimo della sua efficacia quando agisce su di un substrato che contiene tale elemento (nel caso in questione gesso e/o carbonato di calcio); in assenza di una tale composizione il formulato si mostra quindi piuttosto inefficace. Per le

ragioni appena descritte ed essendo inoltre l'impacco costituito da un mezzo acquoso, esso ha rimosso la finitura di natura gessosa, parzialmente solubile in acqua, di cui infatti non si trova traccia dopo l'applicazione.

Lo spray di acqua nebulizzata non ha prodotto un risultato soddisfacente (*Fig. 10*) mostrando inoltre l'inconveniente dell'apparire di formazioni bianco/grigiastre localizzate. La finitura, ove presente, sembra conservarsi in parte. Si evince quindi che la nebulizzazione, protratta per un tempo di 25 minuti rimuove, almeno in parte la finitura gessosa, pur non ottenendo una superficie libera dai depositi neri. Vi sono infatti anche lacerti minimali di sporco grigio raccolto a piccole placche.

Fig. 10 – Formella dopo pulitura con nebulizzazione.

5. Protezione e consolidamento superficiale. Valutazione dei prodotti conservativi

I prodotti impiegati sia sui laterizi, sia sulle superfici in terracotta, sono stati scelti dopo un'ampia ricerca bibliografica e sulla base della precedente esperienza di test in laboratorio effettuata da ICVBC. La valutazione degli effetti di efficacia e di variazione cromatica è avvenuta mediante il metodo della spugna di contatto e mediante misure colorimetriche (Colorimetro a Riflettanza Minolta Chromameter 200). Il metodo della spugna di contatto, recentemente normato dall'UNI, si avvale di una spugna che viene imbibita di acqua, premuta contro la superficie assorbente in modo controllato e

riproducibile, e ripesata. Per differenza si calcola quanta acqua è stata assorbita sulla superficie di misura. Eseguendo l'applicazione della spugna prima e dopo i trattamenti conservativi, si misurano delle variazioni indotte dalla presenza del trattamento stesso e che riguardano le sue proprietà idrorepellenti.

Qui di seguito vengono brevemente illustrate le caratteristiche dei prodotti testati.

Consolidanti

1. RC 70 (Rhodia); Rhodia RC 70 è un silicato di etile che deve la sua azione alla idrolisi e successiva formazione di silice idrata che precipita all'interno della porosità del materiale lapideo; successivamente la silice idrata subisce un lento processo di disidratazione e polimerizzazione.
2. VP 5035 (Degussa); il prodotto è un silicato di etile modificato allo scopo di rendere il prodotto maggiormente elastico: una catena normal-butilica collega due gruppi etossi-silanici. L'inserimento della catena n-butilica permette inoltre una leggera azione idrorepellente.
3. ACRIL ME (CTS); si tratta di una resina acrilica, dispersa in acqua a pH 8.5 in microemulsione. Presenta una viscosità di massima di 500cps e un aspetto semi-trasparente.

Protettivi

1. RC 80 (Rhodia); è una miscela di silicato di etile e di poli-metil-silossano. Si tratta quindi di un consolidante con una leggera azione idrorepellente.
2. Wacker 290 (Wacker Chemie AG); è un silossano oligomerico.
3. Fluoline HY (CTS); si tratta di una soluzione di copolimeri fluorurati in miscela di solventi (acetato di butile, acetone).

Come si evince dai risultati riassunti in *Tab. 2*, tutti i prodotti testati abbattono l'assorbimento di acqua per capillarità in misura più che soddisfacente. L'efficacia di tutti i prodotti utilizzati viene quindi considerata ottima.

I risultati delle misure colorimetriche vengono espressi nel Sistema CIE 1931; questo sistema viene molto utilizzato quando è necessario valutare differenze di colore di una superficie opaca, tra prima e dopo un trattamento di finitura (*Tabb. 3 e 4*); da una analisi dei dati emerge che, per quanto concerne i protettivi, il minor impatto cromatico viene garantito dal prodotto Fluoline (ΔE=1,98); per quanto

concerne invece i consolidanti il prodotto che determina le minori variazioni cromatiche è ACRIL ME (ΔE=4,72).

Sulle terrecotte (*Tab. 4*), si considera accettabile un valore di ΔE inferiore a 3, soglia di una variazione cromatica identificabile ad occhio nudo; per quanto concerne i prodotti protettivi due di loro rispondono a tale requisito (W290 e Fluoline), mentre fa eccezione, seppur di poco, il prodotto RC80 (ΔE=4,24). Nel caso dei consolidanti tutti i prodotti impiegati hanno determinato variazioni che possono essere considerate accettabili.

Tab. 2 – Misure di assorbimento di acqua su laterizio.

	Prodotto	**H_2O ass. (g)**		**Riduzione**
		Prima	**Dopo**	**(%)**
Protettivo	RC80	3.16	0.01	99.8
	W290	3.16	0.01	99.8
	Fluoline	2.91	0.01	99.5
Consolidante	RC70	3.10	0.02	99.5
	WP5035	2.99	0.02	99.3
	ACRIL ME	3.23	0.02	99.5

Tab. 3 – Misure colorimetriche per la valutazione dei trattamenti protettivi e consolidanti su laterizio.

Prodotto	**T_0**			**T_1**						
	L	a	b	L	a	b	**Δ L**	**Δ a**	**Δ b**	**Δ E**
Protettivo										
RC80	48,19	10,37	11,45	38,11	13,62	14,59	-10,09	3,25	3,14	**11,05**
W290	49,96	8,74	9,35	41,94	12,70	15,09	-8,03	3,96	5,74	**10,63**
Fluoline	53,09	15,35	17,10	51,29	14,93	17,80	-1,80	-0,42	0,70	**1,98**
Consolidante										
RC70	56,11	10,82	12,48	45,96	14,99	16,19	-10,15	4,17	3,71	**11,58**
WP5035	51,27	14,72	15,78	39,92	18,41	19,19	-11,35	3,69	3,40	**12,41**
ACRIL ME	53,32	12,06	11,88	49,25	13,24	13,97	-4,07	1,18	2,10	**4,72**

Legenda: T_0=Misure effettuate prima dell'applicazione dei prodotti, T_1=Misure effettuate dopo l'applicazione dei prodotti.

Tab. 4 – Misure colorimetriche per la valutazione dei trattamenti protettivi e consolidanti su terracotta.

Prodotto	**T_0**			**T_1**						
	L	a	b	L	a	b	**Δ L**	**Δ a**	**Δ b**	**Δ E**
Protettivo										
RC80	51,62	23,02	27,08	47,48	23,88	26,86	-4,14	0,86	-0,22	**4,24**
W290	47,80	22,65	23,78	47,32	22,19	22,49	-0,48	-0,46	-1,29	**1,45**
Fluoline	45,81	18,75	19,56	43,90	20,01	21,33	-1,91	1,26	1,78	**2,90**
Consolidante										
RC70	41,99	22,01	22,31	43,35	21,65	21,34	1,36	-0,36	-0,97	**1,71**
WP5035	42,80	19,67	20,46	40,45	20,43	19,82	-2,35	0,76	-0,64	**2,55**
ACRIL ME	48,79	20,48	20,52	47,14	21,33	21,91	-1,65	0,85	1,40	**2,32**

Legenda: T_0=Misure effettuate prima dell'applicazione dei prodotti, T_1=Misure effettuate dopo l'applicazione dei prodotti.

6. Conclusioni

I dati raccolti durante il cantiere sperimentale hanno permesso di introdurre degli elementi di oggettività nella scelta del metodo di pulitura; infatti mettendo in fila l'insieme delle osservazioni, macroscopiche, microscopiche, e composizionali, si è pervenuti, attraverso pochi dubbi, alla scelta del metodo di pulitura che utilizza una sorgente impulsata laser Nd:YAG. In questo modo è assicurata la maggiore conservazione dei materiali di finitura e dei microframmenti in corso di distacco a causa dei fenomeni di degrado in alcuni casi particolarmente avanzati; ciò ha evitato l'intervento di pre-consolidamento che di norma complica di molto le operazioni di pulitura stesse. Se dal punto di vista del manufatto questa si è rivelata la scelta ottimale, dal punto di vista della logistica di cantiere, tale scelta ha comportato una revisione della tempistica e dei costi di intervento, l'uso di accorgimenti legati alla sicurezza dell'operatore, e un aggiornamento specialistico finalizzato all'ottimale utilizzo dello strumento.

7. Ringraziamenti

Gli autori desiderano ringraziare tutto il personale della sezione di Milano dell'Istituto per la Conservazione e Valorizzazione dei Beni Culturali del Consiglio Nazionale delle Ricerche, che ha partecipato allo studio oggetto di questo scritto.

BIBLIOGRAFIA

UNI 11182. 2006 Beni Culturali. Materiali Lapidei Naturali ed Artificiali. Descrizione della forma di alterazione. Termini e Definizioni.

UNI 11432:2011 Beni culturali - Materiali lapidei naturali ed artificiali - Misura della capacità di assorbimento di acqua mediante spugna di contatto.

A. Sansonetti, C. Colombo, C. Conti, S. Della Torre, R. Fant and J. Striova. Saving Finishings on Terracotta Tiles with Laser Cleaning, in: Book of Abstracts, IX Lacona Laser in Conservation of Artworks Int. Conference. London 7-10 October 2011.

R. Fant, M. Gasparoli, P. Gasparoli, A. Sansonetti, J. Striova. La salvaguardia di finiture rosse su terracotta tramite pulitura laser. Atti del 4° Convegno APLAR (Applicazioni laser nel restauro) – Roma 14-15 Giugno 2012.

Matematica Istituto Lombardo (Rend. Scienze) 146, 113-132 (2012)

FINITE-DIMENSIONAL GLOBAL ATTRACTOR FOR A NONLOCAL PHASE-FIELD SYSTEM

Nota del s.c. MAURIZIO GRASSELLI (*)

(Adunanza del 22 marzo 2012)

A mia madre

SUNTO. – Analizziamo un sistema di campo di fase in cui l'equazione di bilancio dell'energia è accoppiata linearmente con un'equazione differenziale ordinaria non-lineare e non-locale per il parametro d'ordine χ. Quest'ultima equazione è caratterizzata da una convoluzione spaziale che modella le interazioni a lungo raggio e da un potenziale di configurazione singolare che costringe χ ad assumere valori nell'intervallo $(-1, 1)$. Proviamo che il corrispondente sistema dinamico è dissipativo ovvero possiede un insieme assorbente limitato in un opportuno spazio delle fasi. Stabiliamo quindi l'esistenza di un attrattore globale di dimensione finita.

***

ABSTRACT. – We analyze a phase-field system where the energy balance equation is linearly coupled with a nonlinear and nonlocal ODE for the order parameter χ. The latter equation is characterized by a space convolution term which models long-range interactions and a singular configuration potential that forces χ to take values in the interval $(-1, 1)$. We prove that the corresponding dynamical system is dissipative, *i.e.*, it has a bounded absorbing set in a suitable phase space. Then we establish the existence of a finite-dimensional global attractor.

(*) Dipartimento di Matematica, Politecnico di Milano, Via E. Bonardi 9, 20133 Milano, Italy.
E-mail: maurizio.grasselli@polimi.it

1. Introduction

Consider a two-phase material (like, for instance, a mixture of ice and water) which occupies a bounded domain in $\Omega \subset \mathbb{R}^d$, $1 \leq d \leq 3$ and denote by ϑ its relative temperature with respect to a given (constant) critical one (*e.g.*, the one at which the two phases coexist). A well-known model which accounts for phase changes due to the variation of ϑ only was proposed and analyzed in [15] (see also [14] and references therein). This is based on the Ginzburg-Landau theory of phase transitions and it assumes as further variable an order parameter (or phase-field) χ which characterizes, for instance, the most energetic phase of the material (say water, in a water-ice system). Using a phenomenological argument, it is postulated that the evolution of χ is ruled by a *gradient flow* of the form

$$\chi_t = -\delta_\chi E(\vartheta, \chi), \tag{1}$$

δ_χ being the variational derivative with respect to χ of the *free energy* E defined by

$$E(\vartheta, \chi) = \int_\Omega \left(\frac{\nu}{2} |\nabla \chi(x)|^2 + W(\chi(x)) - \alpha \vartheta(x) \chi(x) \right) \mathrm{d}x, \tag{2}$$

where $\nu > 0$ and $\alpha \in \mathbb{R}$ are given constants, the latter one being related to the latent heat. Here W is the (density of) potential energy associated with the phase configuration which can be defined either on a finite interval (*e.g.* $(-1, 1)$) or on the whole real line. In the first case, W is called *singular* and the most typical form is the following

$$W(r) = (1+r)\ln(1+r) + (1-r)\ln(1-r) - \frac{\lambda}{2} r^2, \quad r \in (-1, 1), \tag{3}$$

where $\lambda \in \mathbb{R}$. Another important class of potentials are defined on $\mathbb{R}$ and are usually called *smooth*. A typical example is $W(r) = (r^2 - 1)^2$ which can also be viewed as an approximation of W of (3) on $[-1, 1]$.

It is worth observing that the term on the right hand side of (1) may be viewed as a sort of *generalized force* driving the evolution of χ, that is, the phase transformation. Combining now the balance equation for the internal energy density with (1) we find the evolution system

$$(\beta \vartheta + \alpha \chi)_t - \gamma \Delta \vartheta = f, \tag{4}$$

$$\zeta \chi_t - \nu \Delta \chi + W'(\chi) - \alpha \vartheta = 0, \tag{5}$$

in $\Omega \times (0, +\infty)$, where f represents a given (volumic) heat source and all the constants β, α, γ, ζ, ν are positive. Equations (4)–(5) are also known as Caginalp phase-field system. An important feature of this rather simple model is the fact that its (formal) asymptotic limits are well-known sharp interface problems (see, *e.g.*, [16, 17, 19]). We recall that phase-field systems are diffuse interface models in the sense that the sharp interface separating two different phases, which is usually very difficult to handle theoretically and numerically, is replaced by the level set of a suitable (smooth) order parameter.

From the mathematical viewpoint, system (4)–(5) endowed with initial conditions and various types of boundary conditions has been investigated by many authors (see, *e.g.*, [10, 12, 13, 20, 25, 26, 27, 30, 38, 39, 43, 44, 45, 51, 56, 57, 60] and references therein), even under more general assumptions (for instance, where α is a function depending on χ). Besides well-posedness results, an important issue is the longtime behavior of solutions. This behavior is usually non trivial since the set of stationary states is a *continuum* if the spatial dimension is greater than one. Thus the existence of a Lyapunov functional does not guarantee that a given trajectory converges to a single equilibrium. To prove that the Łojasiewicz-Simon approach has been employed (see, in particular, [20, 25, 26, 39, 60]). Regarding the global dynamics, system (4)–(5), endowed with appropriate boundary conditions, can be interpreted as a dynamical system in a suitable phase space. This system is dissipative, *i.e.*, there exists a bounded absorbing set in the phase space. In addition, it possesses a finite-dimensional global attractor as well as an exponential attractors (cf., for instance, [10, 12, 13, 20, 38, 44, 45, 51]).

On the other hand, the free energy E can be viewed as an approximation of a nonlocal expression of the following type (see [22] and references therein, cf. also [33, 34, 36, 37] for phase separation models)

$$E_{\mathtt{nl}}(\vartheta, \chi) = \int_\Omega \left(\int_\Omega k(x-y) \frac{|\chi(x)-\chi(y)|^2}{4} \, \mathrm{d}y + W(\chi(x)) - \alpha\vartheta(x) + \chi(x) \right) \mathrm{d}x, \tag{6}$$

where $k : \mathbb{R}^d \to \mathbb{R}$ is an interaction kernel satisfying $k(x) = k(-x)$ and

such that

$$\kappa(x) := \int_\Omega k(x-y)\ \mathrm{d}y \tag{7}$$

is bounded and nonnegative (see, for instance, [34, Rem. 2.2] for concrete examples). Indeed one can recover (2) by taking

$$\int_\Omega \frac{\nu}{2}|\nabla\chi(x)|^2\ \mathrm{d}x,$$

as a first approximation of

$$\int_\Omega\int_\Omega k(x-y)|\chi(x)-\chi(y)|^2\ \mathrm{d}y\,\mathrm{d}x,$$

where $\nu = 2\int_\Omega k(y)(y_i)^2 dy$ is supposed to be independent of coordinate i. In concrete examples k is localized in a neighborhood of 0 so that ν is related to the thickness of the interface between the components. It is interesting to point out that this kind of approximation was already introduced by van der Waals in his celebrated paper [54] and since then it was widely adopted in the mathematical literature on phase transitions. This might be due to the fact that most people are more used to deal with PDE rather than nonlocal operators.

The evolution system which derives from the nonlocal free energy (6) takes the form

$$\beta\vartheta_t + \alpha\chi_t - \Delta\vartheta = f, \tag{8}$$

$$\zeta\chi_t + \kappa\chi + W'(\chi) = \mathcal{J}[\chi] + \alpha\vartheta, \tag{9}$$

in $\Omega\times(0,+\infty)$, where

$$\mathcal{J}[v](x) = \int_\Omega k(x-y)v(y)\ \mathrm{d}y, \qquad x\in\Omega. \tag{10}$$

Here we have taken $\gamma = 1$ for simplicity. The well-posedness for this system was firstly established in [6] through a semigroup approach in

the case $\Omega = \mathbb{R}$ for a smooth W. In a similar setting, existence of traveling waves for small α was proven in [4] (see [3, 5, 7, 8, 11, 21, 32, 59] for various results on the nonlocal Allen-Cahn type equation, *i.e.*, (9) with $\alpha = 0$, cf. also the recent monograph [1] for closely related problems). Nucleation simulation by using nonlocal interactions has been studied in [55], while coarsening has been analyzed in [29]. Phase interface dynamics associated with (8)–(9) was analyzed [23, 24] (cf. also references therein and [22]) by formal asymptotics (see also [18] for higher-order approximations and [28] for a related model). The case of bounded multi-dimensional domains was examined in [9] and [31] (cf. also [47, 48, 49, 50, 58] for results on more refined models and [2] for the numerical analysis). Well-posedness issues were discussed in [9] when ϑ is subject to homogeneous Neumann boundary condition and W is smooth. In addition, some results concerned with the asymptotic behavior were demonstrated (*e.g.*, the existence of a bounded absorbing set in one spatial dimension). In [31], the potential W is smooth as well, but ϑ is subject to Dirichlet homogeneous. There the main goal was to establish the convergence of a given trajectory to a single stationary state an this fact was proven by means of a generalized version of the Łojasiewicz-Simon inequality, provided that W is real analytic. Note that, also in the nonlocal case, the set of stationary states can be a *continuum* in dimension greater than one, even though $z \mapsto W'(z) + \kappa z$ is invertible (see [31, Introduction], see also [7] for the nonlocal Allen-Cahn equation). A similar issue is analyzed in the more recent contribution [40], where equation (9) has an additional relaxation term of the form $\varepsilon\chi_{tt}$ and W is a singular potential (though not of type (3)).

However, none of the above results is concerned with the existence of a global attractor even though this is a rather natural feature of phase-field dynamics (cf. above for the local models). In this respect, it is interesting to report the related conclusion of the pioneering paper [6]:

Global information, for instance on the existence of an absorbing set and an inertial manifold …, would also be very useful.

A possible explanation might be that it is not evident how to show some precompactness for the variable χ (on this point see also [29]). In this note we establish something more than the existence of a bounded absorbing set, namely, the existence of a finite-dimensional global attractor. This is achieved by estimating the differences of two trajecto-

ries which originated from a suitable bounded absorbing set. Such a difference is controlled by a contractive part depending on the initial data plus a precompact term. For the latter it is essential (but *natural*) to assume some compactness on the linear nonlocal operator $\mathcal{J}$. For the sake of simplicity, we will take full advantage of the setting developed in [40]. Thus we will suppose that ϑ satisfies Dirichlet homogeneous boundary condition, while $W(r)$ will be a singular potential defined on $(-1,1)$ which goes to $+\infty$ as r approaches the pure states ± 1. This will allow us to easily deduce well-posedness since the model studied in [40] essentially reduces to ours when we take $\varepsilon = 0$. The drawback is that one cannot take as initial datum for χ a function which takes ± 1 values (*i.e.*, pure phases) in some subset of positive measure. In other (ancient) words *corpus inani distinctum, quoniam nec plenum naviter extat nec porro vacuum* ([52]). Nonetheless, with some extra-effort, our argument can also be applied to other types of potentials like, for instance, (3) (see Remarks 2.6 and 4.4 below) and to different boundary conditions for ϑ (*e.g.*, homogeneous Neumann).

The paper is organized as follows. In the next section we discuss well-posedness issues. Section 3 contains a basic uniform in time estimate which is exploited in Section 4 to show that our problem defines a dissipative dynamical system on a suitable phase space. Then we conclude with our main result: the existence of a finite-dimensional global attractor.

2. Well-posedness

Let us introduce some notation first. Set $H := L^2(\Omega)$ and $V := H^1_0(\Omega)$ and denote by $(\cdot,\cdot)$ the scalar product in H and by $\|\cdot\|$ the induced norm. In general, $\|\cdot\|_X$ will indicate the norm in a generic real Banach space X. Let V be endowed with the norm $\|\cdot\|_V := \|\nabla \cdot\|$. Let us identify H with its topological dual H' so that we have the continuous and compact inclusions $V \subset H \subset V'$. Moreover, we set $A := -\Delta : D(A) = H^2(\Omega) \cap V \to H$.

Regarding the potential W, we suppose

$$W \in C^2((-1,1);\mathbb{R}^+), \qquad \lim_{r \to 1^-, -1^+} W(r) = +\infty, \tag{11}$$

$$\exists\, \lambda \in \mathbb{R} : \quad W''(r) \geq -\lambda, \qquad \forall\, r \in (-1,1). \tag{12}$$

It is easy to see that (11)–(12) entail

$$W'(r)r \geq W(r) - \frac{\lambda r^2}{2} - W(0), \quad \forall\, r \in (-1,1). \tag{13}$$

Moreover, if $v \in L^\infty(\Omega)$ is any function such that $W(v) \in L^\infty(\Omega)$, then

$$\exists\, \delta = \delta\big(W, \|v\|_{L^\infty(\Omega)}\big) \in (0,1) \quad \text{such that} \quad -1+\delta \leq v(x) \leq 1-\delta \quad \text{for a.e. } x \in \Omega. \tag{14}$$

Consider now κ defined by (7) and the nonlocal operator $\mathcal{J}$ (cf. (10)). We require

$$k \in W^{1,1}(\mathbb{R}^d). \tag{15}$$

Observe that assumption (15) entails that $\mathcal{J} \in \mathcal{L}(H;H)$ is self-adjoint and compact. Also, $\mathcal{J}$ is compact from $L^\infty(\Omega)$ to $C(\overline{\Omega})$. Taking some constants equal to one, system (8)–(9) with ϑ subject to Dirichlet homogeneous boundary condition and initial conditions can now be written as follows

$$\vartheta_t + \alpha\chi_t + A\vartheta = f, \qquad \text{a.e. in } \Omega \times (0,T), \tag{16}$$

$$\chi_t + \kappa\chi + W'(\chi) = \mathcal{J}[\chi] + \alpha\vartheta, \qquad \text{a.e. in } \Omega \times (0,T), \tag{17}$$

$$\vartheta_{|t=0} = \vartheta_0, \quad \chi_{|t=0} = \chi_0 \qquad \text{a.e. in } \Omega, \tag{18}$$

where

$$\vartheta_0 \in V, \tag{19}$$

and

$$f \in H. \tag{20}$$

Moreover, we suppose that

$$\chi_0 \in L^\infty(\Omega) \quad \text{s.t. } \exists\, \delta_0 \in (0,1): \quad -1+\delta_0 \leq \chi_0 \leq 1-\delta_0 \quad \text{a.e. in } \Omega, \tag{21}$$

where the latter property is equivalent to say that $W(\chi_0) \in L^\infty(\Omega)$.

Well-posedness can be proven arguing as in the proof of [40, Thm.2.1] with $\varepsilon = 0$ (cf. also [40, Rem.2.7])

Theorem 2.1. *Let (11)–(12), (15), (19)–(21) hold. Then, for any given $T > 0$, there exists one and only one pair (ϑ, χ) such that*

$$\vartheta \in L^2(0,T;D(A)) \cap H^1(0,T;H) \cap C^0([0,T];V), \tag{22}$$

$$\chi,\ W(\chi) \in L^\infty(0,T;L^\infty(\Omega)), \quad \chi_t \in L^2(0,T;L^\infty(\Omega)), \tag{23}$$

which solves (16)–(18). Moreover, there exists $\delta = \delta(W, k_0, \kappa, \alpha, f, \delta_0) \in (0,1)$ such that

$$-1 + \delta \le \chi(t) \le 1 - \delta \quad \text{a.e. in } \Omega,$$

for almost any $t \in (0,T)$. Next, given two triplets $(\vartheta_{0,1}, \chi_{0,1})$, $(\vartheta_{0,2}, \chi_{0,2})$ of initial data satisfying conditions (19)–(21) (the latter w.r.t. possibly different constants $\delta_{0i} > 0$, $i = 1,2$) and denoting the corresponding solutions by (ϑ_1, χ_1), (ϑ_2, χ_2), respectively, we have the continuous dependence estimate

$$\begin{aligned} &\|(\vartheta_1 - \vartheta_2)(t)\| + \|\nabla(\vartheta_1 - \vartheta_2)\|_{L^2(0,t;H)} + \|(\chi_1 - \chi_2)(t)\| \\ &\quad \le \Lambda_0\big(\|\vartheta_{0,1} - \vartheta_{0,2}\| + \|\chi_{0,1} - \chi_{0,2}\|\big), \quad \forall\, t \in [0,T], \end{aligned} \tag{24}$$

where the positive constant Λ_0 depends on T, Ω, W, k_0, k_1, α, f, and on the initial data (in particular, on δ_{0i}, $i = 1,2$).

Remark 2.2. To establish existence instead of the fixed-point technique used in [40] one can use a vanishing viscosity argument like in [31].

We also have a higher-order control for the temperature difference, namely,

Corollary 2.3. *Let the assumptions of Theorem 2.1 hold. In addition to estimate (24) we have*

$$\begin{aligned} &\|\nabla(\vartheta_1 - \vartheta_2)(t)\| \\ &\quad + \|A(\vartheta_1 - \vartheta_2)\|_{L^2(0,t;H)} \le \Lambda_1\big(\|\vartheta_{0,1} - \vartheta_{0,2}\|_V + \|\chi_{0,1} - \chi_{0,2}\|\big), \end{aligned} \tag{25}$$

for all $t \in [0,T]$ where Λ_1 is a positive constant similar to Λ_0.

Proof. It suffices to write equation (16) for $\vartheta_1 - \vartheta_2$, multiplying by $A(\vartheta_1 - \vartheta_2)$ recalling (17) and using (24). ∎

Remark 2.4. Assumption (15) is satisfied by concrete examples of interaction kernels like the ones mentioned in [34, Rem. 2.2].

Remark 2.5. Assumptions (19) and (20) can be weakened as follows

$$\vartheta \in D(A^{\frac{\rho}{2}}), \quad f \in D(A^{(\rho-1)/2}), \tag{26}$$

where $\rho \in (\frac{3}{4}, 1)$. In this case, we recall that $D(A^{\varrho}) \subset H^{2\varrho}(\Omega) \hookrightarrow L^{\infty}(\Omega)$ since the spatial dimension is three at most. In this case estimate (25) becomes

$$\begin{aligned} &\|A^{\frac{\rho}{2}}(\vartheta_1 - \vartheta_2)(t)\| + \|A^{\frac{\rho+1}{2}}(\vartheta_1 - \vartheta_2)\|_{L^2(0,t;H)} \\ &\quad \leq \Lambda_1\big(\|A^{\frac{\rho}{2}}(\vartheta_{0,1} - \vartheta_{0,2})\| + \|\chi_{0,1} - \chi_{0,2}\|\big), \quad \forall\, t \in [0,T]. \end{aligned}$$

Remark 2.6. The second assumption (11) does not allow us to consider the presence of pure phases in the system (see also (21)). Note that, for instance, potentials like (3) should be bounded in $[-1, 1]$. This case can still be handled if we deal with the Allen-Cahn equation only and temperature is assumed to be given (see Remark 4.4 below). However, system (8)–(9) with more general potentials require further arguments and have been analyzed elsewhere (see [41]).

3. A DISSIPATIVE ESTIMATE

Here we establish some uniform in time estimates which are essentially contained in the proof of [40, Thm.2.1] taking $\varepsilon = 0$.

Theorem 3.1. *Let the assumptions of Theorem 2.1 hold. Then the unique solution (ϑ, χ) to (16)–(18) satisfies the following estimate*

$$\|\vartheta(t)\|_V^2 + \|W(\cdot, \chi(t))\|_{L^\infty(\Omega)} \leq C_0(1 + c_{\delta_0} + \|\vartheta_0\|_V^2)e^{-\varsigma t} + C_1, \tag{27}$$

for any $t \geq 0$. Here ς, C_0 and C_1 are positive constants which depend on Ω, λ, $W(0)$, k_1, α and f at most, while $c_{\delta_0} > 0$ depends on δ_0 and W.

Proof. Let us multiply equation (16) by $\vartheta(t) + \xi A\vartheta(t)$ for some given $\xi > 0$ to be chosen in the sequel. Integrating over Ω, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\|\vartheta\|^2 + \frac{\xi}{2}\|\nabla\vartheta\|^2\right) + \|\nabla\vartheta\|^2 + \xi\|A\vartheta\|^2 = (f - \alpha\chi_t, \vartheta + \xi A\vartheta).$$

Consider now equation (17). Multiplying it by $\chi_t(t) + \eta\chi(t)$, where $\eta > 0$ will be fixed in the sequel, and integrating over Ω we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\eta}{2}\|\chi\|^2 + \frac{1}{2}(\kappa\chi,\chi) + (W(\chi),1)\right) + \|\chi_t\|^2 + \eta(\kappa\chi,\chi) + \eta(W'(\chi),\chi) = (\mathcal{J}[\chi] + \alpha\vartheta, \chi_t + \eta\chi).$$

Adding the two identities we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E} + \|\nabla\vartheta\|^2 + \xi\|A\vartheta\|^2 + \|\chi_t\|^2 + \eta(\kappa\chi,\chi) + \eta(W'(\chi),\chi) = (f, \vartheta + \xi A\vartheta) - \xi\alpha(\chi_t, A\vartheta) + (\mathcal{J}[\chi], \chi_t + \eta\chi) + \eta\alpha(\vartheta,\chi),$$

where, for all $t \geq 0$,

$$\mathcal{E}(t) = \frac{1}{2}\|\vartheta(t)\|^2 + \frac{\xi}{2}\|\nabla\vartheta(t)\|^2 + \frac{\eta}{2}\|\chi(t)\|^2 + \frac{1}{2}(\kappa\chi(t),\chi(t)) + (W(\chi(t)),1).$$

Recalling (13), (15) and the fact that $|\chi| \leq 1$ almost everywhere in $\Omega \times (0,T)$, for any given $T > 0$, it is not difficult to choose ξ and η such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E} + c_1\mathcal{E} + c_2\left(\|A\vartheta\|^2 + \|\chi_t\|^2\right) \leq c_3 + c_4\|f\|^2, \tag{28}$$

where c_i, $i = 1,\ldots,4$, are positive constants. In particular, c_1 and c_2 depend on α, while c_3 depends on Ω, λ, $W(0)$ and k_1. Here we have also used the Young and Poincaré inequalities.

We now argue as in [40, Proof of Thm.2.1] and we test (17) by $\chi_t + \sigma\chi$, for some $\sigma > 0$ to be properly selected, but we do not integrate over Ω. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G} + (\chi_t)^2 + \sigma\kappa\chi^2 + \sigma W'(\chi)\chi = (\mathcal{J}[\chi] + \alpha\vartheta)(\chi_t + \sigma\chi), \tag{29}$$

where

$$\mathcal{G}(x,t) = \frac{\sigma}{2}(\chi(x,t))^2 + \frac{1}{2}\kappa(x)(\chi(x,t))^2 + W(\chi(x,t)), \quad \text{a.e. in } \Omega,\ t \geq 0.$$

Reasoning as before and using Young's inequality, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}(x,\cdot)+c_5\mathcal{G}(x,\cdot)+c_6(\chi_t(x,\cdot))^2 \leq c_7(1+|\vartheta(x,\cdot)|^2), \qquad \text{for a.a. } x\in\Omega,$$

where c_i, $i=5,\dots,7$ are positive constants depending at most on Ω, λ, $W(0)$, k_1 and α. Thanks to the continuous embedding $D(A)\subset L^\infty(\Omega)$, we deduce

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}(x,\cdot)+c_5\mathcal{G}(x,\cdot)+c_6(\chi_t(x,\cdot))^2 \leq tc_8(1+\|A\vartheta\|^2), \qquad \text{for a.a. } x\in\Omega. \tag{30}$$

If we multiply the above equation by $c_9=\frac{c_2}{2c_8}$ and we add it to (28), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{E}+c_9\mathcal{G}(x,\cdot))+c_1\mathcal{E}+c_9\mathcal{G}(x,\cdot)+\frac{c_2}{2}\|A\vartheta\|^2+c_{10}\|\chi_t\|^2 \leq c_{11}(1+\|f\|^2), \tag{31}$$

for almost any $x\in\Omega$.

Applying now Gronwall's inequality to (31) we get, for all $t\geq 0$,

$$\mathcal{E}(t)+c_9\mathcal{G}(x,t) \leq (\mathcal{E}(0)+c_9\mathcal{G}(x,0))\,e^{-\mu t} + \frac{2c_{11}}{\mu_1}\left(1+\|f\|^2\right), \quad \text{for a.a. } x\in\Omega,$$

where $\mu_1=\min\{1,c_1\}$, which yields (27). ∎

4. Existence of the global attractor

A consequence of inequality (27) is that $\mathcal{B}(R)=\{u\in V : \|u\|\leq R\}$ for a fixed $R>\sqrt{C_1}$ is absorbing for $\vartheta(t)$ as well as for $W(\chi(t))$. Therefore, consider, for instance,

$$X=\mathcal{B}(R)\times\{v\in L^\infty(\Omega) : |v|\leq 1-\delta_1, \text{ a.e. in } \Omega\}, \tag{32}$$

where $\delta_1\in(0,1)$ is such that

$$\{r\in(-1,1) : W(r)\leq R^2\}\subseteq[-1+\delta_1,1+\delta_1].$$

If, for each $(\vartheta_0, \chi_0) \in X$ and any $t \geq 0$, we define

$$S(t)(\vartheta_0, \chi_0) = (\vartheta(t), \chi(t)),$$

then, thanks to (27), there exists $t_0 > 0$ such that $S(t)X \subseteq X$ for all $t \geq t_0$. If we endow X with the $V \times H$-metric then we obtain a complete (bounded) metric space and $S(t)$ is strongly (Lipschitz) continuous semigroup on X owing to (24) and (25). We now prove the main result of this note, namely, the dynamical system $(X, S(t))$ has a finite-dimensional global attractor.

Theorem 4.1. *Let the assumptions of Theorem 2.1 hold. In addition, suppose that*

$$\lambda_0 := ess \inf_{x \in \Omega} \kappa(x) - \lambda > 0. \tag{33}$$

Then $(X, S(t))$ possesses a finite-dimensional connected global attractor.

Proof. Consider $(\vartheta_{0i}, \chi_{0i}) \in X$, $i = 1, 2$, set

$$(\vartheta(t), \chi(t)) = ((\vartheta_1 - \vartheta_2)(t), (\chi_1 - \chi_2)(t))$$

where $(\vartheta_i(t), \chi_i(t)) = S(t)(\vartheta_{0i}, \chi_{0i})$ for $t \geq t_0$, and observe that

$$\vartheta_t + \alpha\chi_t + A\vartheta = 0, \qquad \text{a.e. in } \Omega \times (t_0, +\infty), \tag{34}$$

$$\chi_t + \kappa\chi + W'(\chi_1(t)) - W'(\chi_2(t)) = \mathcal{J}[\chi] + \alpha\vartheta, \quad \text{a.e. in } \Omega \times (t_0, +\infty). \tag{35}$$

Let us multiply equation (34) by $A\vartheta(t)$. Integrating over Ω, we get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla\vartheta\|^2 + \|A\vartheta\|^2 = -(\alpha\chi_t, A\vartheta).$$

from which, using the Young and Poincaré inequalities, we derive the estimate

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla\vartheta\|^2 + c\|\nabla\vartheta\|^2 \leq c_\alpha\|\chi_t\|^2.$$

and, by comparison in (35), we deduce

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla\vartheta\|^2 + c\|\nabla\vartheta\|^2 \leq c(k_1, \delta_1, \alpha)\left(\|\chi\|^2 + \|\mathcal{J}[\chi]\|^2 + \|\vartheta\|^2\right), \tag{36}$$

for all $t \geq t_0$. On the other hand, multiplying (35) by $\chi(t)$ and integrating over Ω we find

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\chi\|^2 + (\kappa\chi, \chi) + (W'(\chi_1) - W'(\chi_2), \chi) = (\mathcal{J}[\chi], \chi) + \alpha(\vartheta, \chi),$$

and (33) entails

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\chi\|^2 + \lambda_0\|\chi\|^2 \leq (\mathcal{J}[\chi], \chi) + \alpha(\vartheta, \chi).$$

Then, Young's inequality gives

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\chi\|^2 + \frac{\lambda_0}{2}\|\chi\|^2 \leq c_{\lambda_0}\|\mathcal{J}[\chi]\|^2 + c_{\alpha,\lambda_0}\|\vartheta\|^2.$$

Since $\mathcal{J}$ is compact and self-adjoint we can find a finite-dimensional projector Π_{λ_0} such that

$$\|\mathcal{J}[v]\|^2 \leq \frac{\lambda_0}{4c_{\lambda_0}}\|v\|^2 + \|\Pi_{\lambda_0}[v]\|^2, \tag{37}$$

for all $v \in H$. As a consequence we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\chi\|^2 + \frac{\lambda_0}{4}\|\chi\|^2 \leq c_{\lambda_0}\|\Pi[\chi]\|^2 + c_{\alpha,\lambda_0}\|\vartheta\|^2. \tag{38}$$

Adding inequality(36) multiplied by $\mu_2 = \frac{\lambda_0}{8c(k_1,\delta_1,\alpha)}$ to (38) and using (37) yield

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\mu_2\|\nabla\vartheta\|^2 + \|\chi\|^2\right)\\ + c\mu_2\|\nabla\vartheta\|^2 + \frac{\lambda_0}{8}\|\chi\|^2 \leq c(\lambda_0, k_1, \delta_1, \alpha)\left(\|\Pi[\chi]\|^2 + \|\vartheta\|^2\right).\end{aligned} \tag{39}$$

Therefore, from (39), we deduce

$$\begin{aligned}\|\vartheta(t)\|_V^2 + \|\chi(t)\|^2 \leq c(\lambda_0, k_1, \delta_1, \alpha)e^{-\mu_3(t-t_0)}\left(\|\vartheta(t_0)\|_V^2 + \|\chi(t_0)\|^2\right)\\ + c(\lambda_0, k_1, \delta_1, \alpha)\int_{t_0}^{t}\left(\|\vartheta(\tau)\|^2 + \|\Pi_{\lambda_0}[\chi(\tau)]\|^2\right)d\tau,\end{aligned} \tag{40}$$

for all $t \in [t_0, T]$ and any fixed $T > t_0$. Here μ_3 is a positive constant depending on λ_0, k_1, δ_1, α.

We now introduce the following pseudometric in X

$$\mathrm{d}_T((\vartheta_{01}, \chi_{01}), (\vartheta_{02}, \chi_{02})) = \left(\int_{t_0}^{T} \left(\|(\vartheta_1 - \vartheta_2)(\tau)\|^2 + \|\Pi_{\lambda_0}[(\chi_1 - \chi_2)(\tau)]\|^2 \right) d\tau \right)^{1/2}$$

and we recall that a pseudometric is (pre)compact in X (with respect to the topology induced by the X-metric) if any bounded sequence in X contains a Cauchy subsequence with respect to d_T (see, for instance, [46, Def. 2.57]).

Let us prove that d_T is precompact in X. Let $\{(\vartheta_{0n}, \chi_{0n})\} \subset X$ (X is bounded) and set $(\vartheta_n(t), \chi_n(t)) = S(t)(\vartheta_{0n}, \chi_{0n})$. Thanks to (22), we have that $\{\vartheta_n\}$ is bounded in $L^2(t_0, T; D(A)) \cap H^1(t_0, T; H)$. Therefore it contains a subsequence which strongly converges in $L^2(t_0, T, V)$. On the other hand, we have that $\{\Pi_{\lambda_0}[\chi_n]\}$ is bounded in $L^\infty(t_0, T; H)$. Also, by comparison in (17), we deduce that $\{(\chi_n)_t\}$ is bounded in $L^\infty(t_0, T; H)$. Therefore $\{(\Pi_{\lambda_0}[\chi_n])_t\}$ is bounded in $L^\infty(t_0, T; H)$ as well. Then $\{\Pi_{\lambda_0}[\chi_n(\cdot)]\}$ contains a subsequence which strongly converges in $L^2(t_0, T; H)$. Summing up $\{(\vartheta_{0n}, \chi_{0n})\}$ contains a Cauchy subsequence with respect to d_T.

From (40), we deduce that there exists $t^* > t_0$ such that

$$\begin{aligned} &\|S(t^*)(\vartheta_{01}, \chi_{01}) - S(t^*)(\vartheta_{02}, \chi_{02})\|_X \\ &\quad \le \frac{1}{2} \|(\vartheta_{01} - \vartheta_{02}, \chi_{01} - \chi_{02})\|_X \\ &\qquad + C(\lambda_0, k_1, \delta_1, \alpha) \mathrm{d}_{t^*}((\vartheta_{01}, \chi_{01}), (\vartheta_{02}, \chi_{02})). \end{aligned}$$

Hence $S(t)$ has a (connected) global attractor (see [46, Thm. 2.56, Prop. 2.59]) of finite fractal (*i.e.*, box counting) dimension (cfr. [42, Thm. 2.8.1]). ∎

Remark 4.2. Assumption (33) seems unavoidable when one wants to investigate the long-time behavior of solutions (cf. [9, (A4)] and [31, (1.19)]). In particular, thanks to this assumption, if we take $f \equiv 0$ and

suppose W real analytic then, on account of [40, Thm. 2.6], we have that the ω-limit set of any pair (θ_0, χ_0) satisfying (19) and (21) reduces to a singleton $\{(0, \chi_\infty)\}$, where

$$\kappa\chi_\infty + W'(\chi_\infty) = \mathcal{J}[\chi_\infty], \qquad \text{a.e. in } \Omega.$$

Remark 4.3. On account of Remark 2.5, we could take a larger phase space by replacing V with $V_\rho = D(A^{\frac{\rho}{2}})$, $\rho \in (\frac{3}{4}, 1)$, in the definition of $\mathcal{B}(R)$ and endowing X with $V_\rho \times H$-norm.

Remark 4.4. Consider the following equation

$$\chi_t + \kappa\chi + W'(\chi) = \mathcal{J}[\chi] + g,$$

where $g \in L^\infty(\Omega \times (0, +\infty))$. In this case it is possible to show a (uniform) strict separation property even when W is a more general potential like, *e.g.*, (3). However, one should use a comparison argument like, *e.g.*, in [39]. Indeed, it is no longer sufficient to show the global uniform boundedness of $W(\chi)$ (cf. (27)). Then one can define a phase space given by the second component of X (see (32)) and, arguing as above for the χ component only, prove the existence of a finite-dimensional global attractor.

ACKNOWLEDGMENTS

The author thanks S. Frigeri, D. Pražák, and G. Schimperna for their helpful remarks.

REFERENCES

[1] F. Andreu-Vaillo, J.M. Mazón, J.D. Rossi, J. Julián Toledo-Melero, "Nonlocal diffusion problems", Mathematical Surveys and Monographs, 165. American Mathematical Society, Providence, RI; Real Sociedad Matemática Española, Madrid, 2010.

[2] S. Armstrong, S. Brown, J. Han, *Numerical analysis for a nonlocal phase field system*, Int. J. Numer. Anal. Model. Ser. B, 1 (2010), 1–19.

[3] P.W. Bates, S. Brown, J. Han, *Numerical analysis for a nonlocal Allen-Cahn equation*, Int. J. Numer. Anal. Model., 6 (2009), 33–49.

[4] P.W. Bates, F. Chen, *Traveling wave solutions for a nonlocal phase-field system*, Interfaces Free Bound., 4 (2002), 227–238.

[5] P.W. Bates, F. Chen, *Spectral analysis and multidimensional stability of traveling waves for nonlocal Allen-Cahn equation*, J. Math. Anal. Appl., 273 (2002), 45–57.

[6] P.W. Bates, F. Chen, J. Wang, *Global existence and uniqueness of solutions to a nonlocal phase-field system*, in: P.W. Bates, S.-N. Chow, K. Lu and X. Pan, Editors, US-Chinese Conference on Differential Equations and Applications, International Press, Cambridge, MA (1997), 14–21.

[7] P.W. Bates, A. Chmaj, *An integrodifferential model for phase transitions: stationary solutions in higher space dimensions*, J. Statist. Phys., 95 (1999), 1119–1139.

[8] P.W. Bates, P.C. Fife, X. Ren, X. Wang, *Traveling waves in a convolution model for phase transitions*, Arch. Rational Mech. Anal., 138 (1997), 105–136.

[9] P.W. Bates, J. Han, G. Zhao, *On a nonlocal phase-field system*, Nonlinear Anal., 64 (2006), 2251–2278.

[10] P.W. Bates, S. Zheng, *Inertial manifolds and inertial sets for the phase-field equations*, J. Dynamics Differential Equations, 4 (1992), 375–397.

[11] S.K. Bhowmik, D.B. Duncan, M. Grinfeld, G.J. Lord, *Finite to infinite steady state solutions, bifurcations of an integro-differential equation*, Discrete Contin. Dyn. Syst. Ser. B, 16 (2011), 57–71.

[12] D. Brochet, X. Chen, D. Hilhorst, *Finite dimensional exponential attractor for the phase field model*, Appl. Anal., 49 (1993), 197–212.

[13] D. Brochet, D. Hilhorst, *Universal attractor and inertial sets for the phase-field model*, Appl. Math. Lett., 4 (1991), 59–62.

[14] M. Brokate, J. Sprekels, "Hysteresis and Phase Transitions", Springer, New York, 1996.

[15] G. Caginalp, *An analysis of a phase field model of a free boundary*, Arch. Rational Mech. Anal., 92 (1986), 205–245.

[16] G. Caginalp, *Stefan and Hele-Shaw type models as asymptotic limits of the phase-field equations*, Phys. Rev. A, 39 (1989), 5887–5896.

[17] G. Caginalp, X. Chen, *Convergence to the phase field model to its sharp interface limits*, European J. Appl. Math., 9 (1998), 417–445.

[18] G. Caginalp, E. Esenturk, *A phase field model with non-local and anisotropic potential*, Discrete Contin. Dyn. Syst. Ser. S, 4 (2011) 311–350.

[19] G. Caginalp, P.C. Fife, *Dynamics of layered interfaces arising from phase boundaries*, SIAM J. Appl. Math., 48 (1988), 506–518.

[20] C. Cavaterra, C.G. Gal, M. Grasselli, A. Miranville, *Phase-field systems with nonlinear coupling and dynamic boundary conditions*, Nonlinear Anal., 72 (2010), 2375–2399.

[21] X. Chen, *Existence, uniqueness and asymptotic stability of traveling waves in nonlocal evolution equations*, Adv. Differential Equations, 2 (1997), 125–160.

[22] C.K. Chen, P.C. Fife, *Nonlocal models of phase transitions in solids*, Adv. Math. Sci. Appl., 10 (2000), 821–849.

[23] X. Chen, G. Caginalp, E. Esenturk, *A phase field model with non-local and anisotropic potential*, Modelling Simul. Mater. Sci. Eng., 19 (2011) 045006(8).

[24] X. Chen, G. Caginalp, E. Esenturk, *Interface conditions for a phase field model with anisotropic and non-local interactions*, Arch. Rational Mech. Anal., 202 (2011), 349–372.

[25] L. Cherfils, A. Miranville, *Some results on the asymptotic behavior of the Caginalp system with singular potentials*, Adv. Math. Sci. Appl., 16 (2007), 107–129.

[26] R. Chill, E. Fašangová, J. Prüss, *Convergence to steady states of solutions of the Cahn- Hilliard and Caginalp equations with dynamic boundary conditions*, Math. Nachr., 13 (2006), 1448–1462.

[27] A. Damlamian, N. Kenmochi, N. Sato, *Subdifferential operator approach to a class of nonlinear systems for Stefan problems with phase relaxation*, Nonlinear Anal., 23 (1994), 115–142.

[28] N. Dirr, *A Stefan problem with surface tension as the sharp interface limit of a nonlocal system of phase-field type*, J. Statist. Phys., 114 (2004), 1085–1113.

[29] D.B. Duncan, M. Grinfeld, I. Stoleriu, *Coarsening in an integro-differential model of phase transitions*, European J. Appl. Math., 11 (2000), 561–572.

[30] C.M. Elliott, S. Zheng, *Global existence and stability of solutions to the phase-field equations*, in "Free boundary problems," Internat. Ser. Numer. Math. 95, 46–58, Birkhäuser Verlag, Basel, 1990.

[31] E. Feireisl, F. Issard-Roch, H. Petzeltová, *A non-smooth version of the Łojasiewicz-Simon theorem with applications to non-local phase-field systems*, J. Differential Equations, 199 (2004), 1–21.

[32] P.C. Fife, *Well-posedness issues for models of phase transitions with weak interaction*, Nonlinearity, 14 (2001), 221–238.

[33] H. Gajewski, *On a nonlocal model of non-isothermal phase separation*, Adv. Math. Sci. Appl., 12 (2002), 569–586.

[34] H. Gajewski, K. Zacharias, *On a nonlocal phase separation model*, J. Math. Anal. Appl., 286 (2003), 11–31.

[35] C.G. Gal, M. Grasselli, A. Miranville, *Robust exponential attractors for singularly perturbed phase-field equations with dynamic boundary conditions*, NoDEA Nonlinear Differential Equations Appl., 15 (2008), 535–556.

[36] G. Giacomin, J.L. Lebowitz, *Phase segregation dynamics in particle systems with long range interactions. I. Macroscopic limits*, J. Statist. Phys., 87 (1997), 37–61.

[37] G. Giacomin, J.L. Lebowitz, *Phase segregation dynamics in particle systems with long range interactions. II. Interface motion*, SIAM J. Appl. Math., 58 (1998), 1707–1729.

[38] M. Grasselli, A. Miranville, G. Schimperna, *The Caginalp phase-field system with coupled dynamic boundary conditions and singular potentials*, Discrete Contin. Dyn. Syst., 28 (2010), 67–98.

[39] M. Grasselli, H. Petzeltová, G. Schimperna, *Long time behavior of solutions to the Caginalp system with singular potential*, Z. Anal. Anwendungen, 25 (2006), 51–72.

[40] M. Grasselli, H. Petzeltová, G. Schimperna, *A nonlocal phase-field system with inertial term,* Quart. Appl. Math., 65 (2007), 451–469.

[41] M. Grasselli, G. Schimperna, *Nonlocal phase-field systems with general potentials*, Discrete Contin. Dyn. Syst. Ser. A., 33 (2013), 5089–5106.

[42] J.K. Hale, "Asymptotic behaviour of dissipative systems", Amer. Math. Soc., Providence, RI, 1988.

[43] A. Jiménez-Casas, A. Rodríguez-Bernal, *Asymptotic behaviour for a phase field model in higher order Sobolev spaces*, Rev. Mat. Complut., 15 (2002), 213–248.

[44] V.K. Kalantarov, *On the minimal global attractor of a system of phase field equations*, (Russian), Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 188 (1991), Kraev. Zadachi Mat. Fiz. i Smezh. Voprosy Teor. Funktsii. 22, 70–86, 186 [translation in J. Math. Sci., 70 (1994), 1767–1777].

[45] O.V. Kapustyan, *An attractor of a semiflow generated by a system of phase-field equations without uniqueness of the solution* (Ukrainian), Ukraïn. Mat. Zh., 51 (1999), 1006–1009 [Translation in Ukrainian Math. J., 51 (1999), 1135–1139 (2000)].

[46] N.J. Koksch, A.J. Milani, "An introduction to semiflows", Chapman & Hall/CRC, Boca Raton, FL, 2005.

[47] P. Krejčí, J. Sprekels, *Nonlocal phase-field models for non-isothermal phase transitions and hysteresis*, Adv. Math. Sci. Appl., 14 (2004), 593–612.

[48] P. Krejčí, J. Sprekels, *Long time behavior of a singular phase transition model*, Discrete Contin. Dyn. Syst., 15 (2006), 1119–1135.

[49] P. Krejčí, E. Rocca, J. Sprekels, *Nonlocal temperature-dependent phase-field models for non-isothermal phase transitions*, J. Lond. Math. Soc. (2), 76 (2007), 197–210.

[50] P. Krejčí, E. Rocca, J. Sprekels, *A nonlocal phase-field model with nonconstant specific heat*, Interfaces Free Bound., 9 (2007), 285–306.

[51] Ph. Laurençot, *Long-time behaviour for a model of phase-field type*, Proc. Roy. Soc. Edinburgh Sect. A, 126 (1996), 167–185.

[52] Titus Lucretius Carus, "De Rerum Natura", Liber I.

[53] A. Miranville, S. Zelik, *Robust exponential attractors for singularly perturbed phase-field type equations*, Electron. J. Differential Equations, 63 (2002), 1–28.

[54] J.S. Rowlinson, Translation of J.D. van der Waals, *The thermodynamic theory of capillarity under the hypothesis of a continuous variation of density*, J. Statist. Phys., 20 (1979), 197–244.

[55] A. Roy, J.M. Rickman, J.D. Gunton, K.R. Elder, *Simulation study of nucleation in a phase-field model with nonlocal interactions*, Phys. Rev. E, 57 (1998), 2610(8).

[56] N. Sato, T. Aiki, *Phase field equations with constraints under nonlinear dynamic boundary conditions*, Commun. Appl. Anal., 5 (2001), 215–234.

[57] G. Schimperna, *Abstract approach to evolution equations of phase field type and applications*, J. Differential Equations, 164 (2000), 395–430.

[58] J. Sprekels, S. Zheng, *Global existence and asymptotic behaviour for a nonlocal phase-field model for non-isothermal phase transitions*, J. Math. Anal. Appl., 279 (2003), 97–110.

[59] X. Wang, *Metastability and stability of patterns in a convolution model for phase transitions*, J. Differential Equations, 183 (2002), 434–461.

[60] Z. Zhang, *Asymptotic behavior of solutions to the phase-field equations with Neumann boundary conditions*, Commun. Pure Appl. Anal., 4 (2005), 683–693.

ON A THEOREM OF LYAPOUNOV

Nota del s.c. ANTONIO GIORGILLI (*)

(Adunanza del 3 maggio 2012)

SUNTO. Si mostra che un sistema Hamiltoniano nell'intorno di un punto di equilibrio, sotto condizione che gli autovalori soddisfino delle condizioni di non–risonanza del tipo di Melnikov, ammette una forma normale che rende evidente l'esistenza di una varietà invariante (locale) a due dimensioni sulla quale si hanno soluzioni note. Nel caso di un autovalore puramente immaginario tali soluzioni formano una famiglia periodica a due parametri che costituisce la continuazione naturale di un modo normale. Questo secondo risultato è stato dimostrato in precedenza da Lyapounov. In questo lavoro si completa quello di Lyapounov dimostrando la convergenza della trasformazione dell'Hamiltoniana a forma normale e rimuovendo le restrizione che gli autovalori siano puramente immaginari.

***

ABSTRACT. It is shown that a Hamiltonian system in the neighbourhood of an equilibrium may be given a special normal form in case the eigenvalues of the linearized system satisfy non–resonance conditions of Melnikov's type. The normal form possesses a two dimensional (local) invariant manifold on which the solutions are known. If the eigenvalue is pure imaginary then these solutions are the natural continuation of a normal mode of the linear system. The latter result was first proved by Lyapounov. The present paper completes Lyapounov's result in that the convergence of the transformation of the Hamiltonian to a normal form is proven and the condition that the eigenvalues be pure imaginary is removed.

(*) Dipartimento di Matematica, Via Saldini 50, 20133 Milano, Italy.
E-mail: antonio.giorgilli@unimi.it

1. Introduction

Consider a canonical system of differential equations in a neighbourhood of an equilibrium, with Hamiltonian

$$H(x,y) = H_0(x,y) + H_1(x,y) + \dots, \quad (x,y) \in \mathbb{C}^{2n}, \tag{1}$$

where the unperturbed quadratic part of the Hamiltonian is

$$H_0(x,y) = \sum_{j=1}^{n} \lambda_j x_j y_j, \quad (\lambda_1, \dots, \lambda_n) \in \mathbb{C}^n, \tag{2}$$

and $H_s(x,y)$ for $s \geq 1$ is a homogeneous polynomial of degree $s+2$. The form (2) is a typical one for the quadratic part of a Hamiltonian system in the neighbourhood of an equilibrium, as under quite general conditions the system may be given that form via a (complex) linear canonical transformation (see, e.g., [10] or [12], § 15).

The Hamiltonian is assumed to be analytic in some neighborhood of the origin of $\mathbb{C}^{2n}$. Moreover λ_1 will be assumed to satisfy at least the first of the following non–resonance conditions:

(i) *First Melnikov's condition:*

$$\lambda_\nu - k\lambda_1 \neq 0 \quad \text{for } k \in \mathbb{Z} \text{ and } \nu = 2, \dots, n\,. \tag{3}$$

(ii) *Second Melnikov's condition:*

$$\lambda_\nu \pm \lambda_{\nu'} - k\lambda_1 \neq 0 \quad \text{for } k \in \mathbb{Z} \text{ and } \nu, \nu' = 2, \dots, n\,, \tag{4}$$

the case $\nu' = \nu$ being included.

In [10] Lyapounov proved that if $\lambda_1 = i\omega_1$ is pure imaginary and the non resonance condition (i) above is satisfied then there exists a two parameter family of solutions of the form

$$x_j = \varphi_j(\xi_1, \eta_1), \quad y_j = \psi_j(\xi_1, \eta_1) \tag{5}$$

written as convergent power series in the arguments

$$\xi_1 = \overset{\circ}{\xi}_1 e^{ita_1(\overset{\circ}{\zeta})}, \quad \eta_1 = \overset{\circ}{\eta}_1 e^{-ita_1(\overset{\circ}{\zeta})}, \tag{6}$$

where $a_1(\overset{\circ}{\zeta}) = \lambda_1 + \dots$ is a convergent power series in $\overset{\circ}{\zeta}_1 = \overset{\circ}{\xi}_1 \overset{\circ}{\eta}_1$. In the case $n = 1$ this actually describes all solutions of the system. A proof in case all λ's are pure imaginary is reported in [12].

The proof of the theorem is worked out by the authors quoted above by expanding the solution in the form (5) and proceeding by comparison of coefficients. From a formal viewpoint the statement above looks equivalent to the existence of a canonical transformation that gives the system (1) a suitable normal form, making the Hamiltonian to depend at least quadratically on $x_2, \dots, x_n, y_2, \dots, y_n$. A formal construction giving such a normal form can be easily produced. However, proving the convergence of the normalization procedure seems to be more difficult. The aim of this paper is precisely to produce a proof of convergence of the transformation to normal form.

I will actually give two different statements that can be proved with the same method. The first one is

THEOREM 1.1. *With the nonresonance hypothesis (i) above (first Melnikov's condition) on $\lambda_1, \dots, \lambda_n$, there exists a canonical, near the identity transformation in the form of a power series convergent in a neighbourhood of the origin, which gives the Hamiltonian* (1) *the normal form*

$$H(x, y) = H_0(x, y) + \Gamma(x_1 y_1) + F(x, y), \tag{7}$$

where $H_0(x, y)$ as in (1)*, $\Gamma(x_1 y_1)$ depends only on the product $x_1 y_1$, and $F(x, y)$ is at least quadratic in $x_2, \dots, x_n, y_2, \dots, y_n$*

The existence of the Lyapounov orbits for λ_1 pure imaginary is evident from the normal form: just put initially $x_2 = \dots = x_n = y_2 = \dots = y_n = 0$, which defines a local invariant two dimensional manifold on which the dynamics is generated by the Hamiltonian $\lambda_1 x_1 y_1 + \Gamma(x_1 y_1)$. The advantage of the normal form is that it allows also to

investigate the dynamics in the neighbourhood of the orbits so found. To this end the following statement may be even more useful.

THEOREM 1.2. *With the nonresonance hypotheses (i) and (ii) above (first and second Melnikov's conditions) on $\lambda_1, \dots, \lambda_n$, there exists a canonical, near the identity transformation in the form of a power series convergent in a neighbourhood of the origin, which gives the Hamiltonian* (1) *the normal form*

$$H(x,y) = H_0(x,y) + \Gamma(x_1y_1, \dots, x_ny_n) + F(x,y), \tag{8}$$

where $H_0(x,y)$ as in (1)*, $\Gamma(x,y)$ contains only monomials $x_1^j y_1^j x_\nu y_\nu$ with a positive integer j and with $\nu = 2, \dots, n$, and $F(x,y)$ is at least cubic in $x_2, \dots, x_n, y_2, \dots, y_n$*

This requires a stronger non–resonance condition. However this normal form may be more convenient if one is interested in the stability of a Lyapounov orbit. Indeed, let all λ's be pure imaginary, say $\lambda_j = i\omega_j$, and write the Hamiltonian restricted to the invariant manifold $x_2 = \dots = x_n = y_2 = \dots = y_n = 0$ in action–angle variables by transforming $x_1 = \sqrt{p}\, e^{iq}$, $y_1 = -i\sqrt{p}\, e^{-iq}$. Thus one gets the Hamiltonian $\omega_1 p + \Gamma(p)$, which represents a non linear oscillator, with orbits written as $p(t) = p_*$, $q(t) = q(0) + \Omega(p_*)t$, where $\Omega(p_*) = \omega_1 + O(p_*)$ is a fixed frequency. By a translation $p' = p - p_*$ the Hamiltonian may be reexpanded (omitting primes) as

$$H(q,p,x,y) = \Omega p + \sum_{j=2}^{n} \lambda_j x_j y_j + H_1 + H_2 + \dots$$

where H_s is a homogeneous polynomial of degree $s+2$ in $p^{1/2}, x_2, \dots, y_n$ with coefficients periodically depending on q. The dynamics of the latter Hamiltonian may be investigated with known methods from perturbation theory. The advantage with respect to the normal form of theorem 1.1 is that the quadratic part of the Hamiltonian is independent of the angle q.

The proof is based on a previous work by the author [5] concerning the construction of the normal form in a case investigated by Cherry [2]

and Moser [11]. It must be stressed that this problem does not involve small divisors. Rather, the possible source of divergence is due to the use of Cauchy's estimates for the derivatives required by the normalization algorithm. The global effect of accumulation of derivatives is controlled with a technique introduced by the author and U. Locatelli in order to achieve a proof of KAM theorem using classical expansions in a perturbation parameter (see [6, 7, 4, 8]).

2. Formal algorithm

Reducing the Hamiltonian to a normal form is a quite general problem which may be solved in a number of different ways. Moreover, the concept of "normal form" may assume a quite general meaning, depending on what one is looking for. Here I state the algorithm in a general form, using the method of composition of Lie series.

2.1 The algorithm for the normal form

Write the Hamiltonian after r normalization steps as

$$H^{(r)}(x,y) = H_0(x,y) + Z_1(x,y) + \ldots + Z_r(x,y) + \sum_{s>r} H_s^{(r)}(x,y), \quad (9)$$

where $Z_1(x,y), \ldots, Z_r(x,y)$ are in normal form, whatever it means, and are homogeneous polynomials of degree $3, \ldots, r+2$. For $r = 0$ the Hamiltonian (1) is already in the wanted form, with no functions Z.

Assume that the Hamiltonian has been given a normal form (9) up to order $r-1$, so that $H^{(r-1)}$ is known. The generating function χ_r and the normal form Z_r are determined by solving the equation

$$L_{H_0}\chi_r + Z_r = H_r^{(r-1)}. \quad (10)$$

where the common notation $L_\varphi \cdot := \{\cdot, \varphi\}$ has been used. The solution of this equation depends on what is meant by "normal form". At a formal level, any choice thatallows to solve the equation above for Z_r and χ_r is

acceptable. Assume for a moment that a method of solution has been found. Then the transformed Hamiltonian is expanded as

$$H^{(r)}_{sr+m} = \frac{1}{s!}L^s_{\chi_r}Z_m + \sum_{p=0}^{s-1}\frac{1}{p!}L^p_{\chi_r}H^{(r-1)}_{(s-p)r+m}$$

$$\text{for } r \geq 2\,,\ s \geq 1 \text{ and } 1 \leq m < r,$$

$$H^{(r)}_{sr} = \frac{1}{(s-1)!}L^{s-1}_{\chi_r}\left(\frac{1}{s}Z_r + \frac{s-1}{s}H^{(r-1)}_r\right) + \sum_{p=0}^{s-2}\frac{1}{p!}L^p_{\chi_r}H^{(r-1)}_{(s-p)r}$$

$$\text{for } r \geq 1 \text{ and } s \geq 2. \tag{11}$$

The justification of the algorithm requires only some straightforward calculation, and is deferred to appendix A.

Thus, the problem is how to solve the equation (10) for the generating function χ_r and the normal form Z_r. Let me make some general considerations.

Let $\mathcal{P}_s$ denote the linear space of homogeneous polynomials of degree s in the complex variables x, y. Let also $\mathcal{P} = \bigcup_{s\geq 0}\mathcal{P}_s$, so that a formal power series is an element of $\mathcal{P}$. A basis in $\mathcal{P}$ is given by the monomials $x^j y^k := x_1^{j_1}\cdot\ldots\cdot x_n^{j_n}y_1^{k_1}\cdot\ldots\cdot y_n^{k_n}$, where j, k are integer vectors with non–negative components. The linear operator L_{H_0} maps every space $\mathcal{P}_s$ into itself. If, due to the choice of the coordinates, the unperturbed Hamiltonian H_0 has the form (2) then the operator L_{H_0} is diagonal, since

$$L_{H_0}x^j y^k = \langle j-k, \lambda\rangle x^j y^k.$$

The kernel and the range of L_{H_0} are defined as usual, namely $\mathcal{N} = L^{-1}_{H_0}(0)$, the inverse image of the null vector in $\mathcal{P}$, and $\mathcal{R} = L_{H_0}(\mathcal{P})$. Both $\mathcal{N}$ and $\mathcal{R}$ are actually subspaces of the same space $\mathcal{P}$, and it turns out that they are complementary subspaces, i.e., $\mathcal{N}\cap\mathcal{R} = \{0\}$, the null vector, and $\mathcal{N}\oplus\mathcal{R} = \mathcal{P}$. A consequence of the properties above is

that L_{H_0} restricted to the subspace $\mathcal{R}$ is uniquely inverted, i.e., the equation $L_{H_0}\chi = \psi$ with $\psi \in \mathcal{R}$ admits an unique solution χ satisfying the condition $\chi \in \mathcal{R}$. That unique solution will be written as $\chi = L_{H_0}^{-1}\psi$, i.e., $L_{H_0}^{-1}$ is defined as the inverse of L_{H_0} restricted to $\mathcal{R}$. It's easy to identify the subspaces $\mathcal{N}$ and $\mathcal{R}$ using the coordinates. Thanks to the diagonal form of L_{H_0} one has

$$\begin{aligned} \mathcal{N} &= \text{span}\left\{x^j y^k \ : \ \langle j-k, \lambda\rangle = 0\right\}, \\ \mathcal{R} &= \text{span}\left\{x^j y^k \ : \ \langle j-k, \lambda\rangle \neq 0\right\}. \end{aligned} \tag{12}$$

Given $\psi \in \mathcal{R}$ and writing $\psi = \sum_{j,k} \psi_{j,k} x^j y^k$, with $\psi_{j,k} = 0$ for $x^j y^k \in \mathcal{N}$, one has

$$L_{H_0}^{-1}\psi = \sum_{j,k} \frac{\psi_{j,k}}{\langle j-k, \lambda\rangle} x^j y^k. \tag{13}$$

In view of the general considerations above we can conclude that the choice of a normal form is subjected to the constraint that in equation (10) we have $H_r^{(r-1)} - Z_r \in \mathcal{R}$. The simplest choice is to ask also $Z_r \in \mathcal{N}$, i.e., to set Z_r to be the projection of $H_r^{(r-1)}$ on the subspace $\mathcal{N}$. This is known indeed as Birkhoff's normal form.

2.2 Normal form for Lyapounov's orbits

I come now to show that the construction of the normal form of theorem 1.1 is formally consistent. Consider the disjoint subsets of $\mathbb{Z}^n$

$$\begin{aligned} \mathcal{K}^\sharp &= \{k \in \mathbb{Z}^n \ : \ k_2 = \ldots = k_n = 0\}, \\ \mathcal{K}^\natural &= \{k \in \mathbb{Z}^n \ : \ |k_2| + \ldots + |k_n| = 1\}, \\ \mathcal{K}^\flat &= \{k \in \mathbb{Z}^n \ : \ |k_2| + \ldots + |k_n| > 1\}. \end{aligned} \tag{14}$$

One has $\mathbb{Z}^n = \mathcal{K}^\sharp \cup \mathcal{K}^\natural \cup \mathcal{K}^\flat$, of course. Considering only integer vectors $j,\ k$ with non–negative components, introduce the subspaces of $\mathcal{P}$

$$\begin{aligned} \mathcal{P}^\sharp &= \text{span}\left\{ x^j y^k \ :\ j+k \in \mathcal{K}^\sharp \right\} \\ \mathcal{P}^\natural &= \text{span}\left\{ x^j y^k \ :\ j+k \in \mathcal{K}^\natural \right\} \\ \mathcal{P}^\flat &= \text{span}\left\{ x^j y^k \ :\ j+k \in \mathcal{K}^\flat \right\} \end{aligned} \tag{15}$$

These subspaces are clearly disjoint, and moreover one has $\mathcal{P} = \mathcal{P}^\sharp \oplus \mathcal{P}^\natural \oplus \mathcal{P}^\flat$. Finally, let $\mathcal{N}^\sharp = \mathcal{N} \cap \mathcal{P}^\sharp$ and $\mathcal{R}^\sharp = \mathcal{R} \cap \mathcal{P}^\sharp$, and define the subspaces $\mathcal{Z}$ and $\mathcal{W}$ of $\mathcal{P}$ as

$$\mathcal{Z} = \mathcal{N}^\sharp \oplus \mathcal{P}^\flat, \quad \mathcal{W} = \mathcal{R}^\sharp \oplus \mathcal{P}^\natural. \tag{16}$$

It is an easy matter to check that $\mathcal{Z} \cap \mathcal{W} = \{0\}$ and $\mathcal{Z} \oplus \mathcal{W} = \mathcal{P}$. The construction of bases for $\mathcal{Z}$ and $\mathcal{W}$ is quite straightforward: a monomial $x^j y^k$ belongs to $\mathcal{Z}$ in either case ($j+k \in \mathcal{K}^\sharp$ and $\langle j-k, \lambda\rangle = 0$) or ($j+k \in \mathcal{K}^\flat$); else it belongs to $\mathcal{W}$. The hypothesis (i) on λ (first Melnikov's condition) formulated at the beginning of the introduction means that the non-resonance condition

$$\langle k, \lambda\rangle \neq 0 \quad \text{for } 0 \neq k \in \mathcal{K}^\sharp \cup \mathcal{K}^\natural \tag{17}$$

is satisfied. This implies $\mathcal{W} \subset \mathcal{R}$, so that for every $\psi \in \mathcal{W}$ the unique solution $\chi = L_{H_0}^{-1}\psi$, $\chi \in \mathcal{W}$ of the equation $L_{H_0}\chi = \psi$ exists. With this setting, the equation

$$L_{H_0}\chi + Z = \Psi, \tag{18}$$

with Ψ known, admits a straightforward solution. Split $\Psi = \Psi_{\mathcal{Z}} + \Psi_{\mathcal{W}}$ with $\Psi_{\mathcal{Z}} \in \mathcal{Z}$ and $\Psi_{\mathcal{W}} \in \mathcal{W}$; such a decomposition exists and is unique, because $\mathcal{Z}$ and $\mathcal{W}$ are complementary subspaces. Then set $Z = \Psi_{\mathcal{Z}}$, and determine $\chi = L_{H_0}^{-1}\Psi_{\mathcal{W}}$ according to (13).

2.3 Adding the second Melnikov's condition

With minor changes one can also prove that the normal form of theorem 1.2 can be constructed. Let

$$\begin{aligned} \mathcal{K}^\sharp &= \{k \in \mathbb{Z}^n \ : \ k_2 = \ldots = k_n = 0\}\,, \\ \mathcal{K}^\natural &= \{k \in \mathbb{Z}^n \ : \ |k_2| + \ldots + |k_n| = 1, 2\}\,, \\ \mathcal{K}^\flat &= \{k \in \mathbb{Z}^n \ : \ |k_2| + \ldots + |k_n| > 2\}\,. \end{aligned} \tag{19}$$

One has again $\mathbb{Z}^n = \mathcal{K}^\sharp \cup \mathcal{K}^\natural \cup \mathcal{K}^\flat$, of course. The subspaces of $\mathcal{P}$ are defined again as in (15), although they turn out to be different in view of the differences in the sets $\mathcal{K}$. Finally, let $\mathcal{N}^\sharp = \mathcal{N} \cap \mathcal{P}^\sharp$, $\mathcal{R}^\sharp = \mathcal{R} \cap \mathcal{P}^\sharp$, $\mathcal{N}^\natural = \mathcal{N} \cap \mathcal{P}^\natural$ and $\mathcal{R}^\natural = \mathcal{R} \cap \mathcal{P}^\natural$, and define the subspaces $\mathcal{Z}$ and $\mathcal{W}$ of $\mathcal{P}$ as

$$\mathcal{Z} = \mathcal{N}^\sharp \oplus \mathcal{N}^\natural \oplus \mathcal{P}^\flat, \quad \mathcal{W} = \mathcal{R}^\sharp \oplus \mathcal{R}^\natural. \tag{20}$$

The difference with respect to the previous case is just that now $\mathcal{N}^\natural$ is not empty, because it contains all monomials of the form $(x_1 y_1)^j \times x_\nu y_\nu$ with $\nu = 2, \ldots, n$ and positive j. This forces the change in the definition of the subspaces $\mathcal{Z}$ and $\mathcal{W}$. However, the properties $\mathcal{Z} \cap \mathcal{W} = \{0\}$ and $\mathcal{Z} \oplus \mathcal{W} = \mathcal{P}$ remain true. Furthermore, in view of the second Melnikov's condition, also the property $\mathcal{W} \subset \mathcal{R}$ holds true, so that for every $\psi \in \mathcal{W}$ the unique solution $\chi = L_{H_0}^{-1}\psi$, $\chi \in \mathcal{W}$ of the equation $L_{H_0}\chi = \psi$ exists.

3. QUANTITATIVE ESTIMATES

Pick a real vector $R \in \mathbb{R}^n$ with positive components. and consider the domain

$$\Delta_R = \{(x, y) \in \mathbb{C}^n \ : \ |x_j| \le R_j \,, \ |y_j| \le R_j \,, \ 1 \le j \le n\}\,, \tag{21}$$

namely a polydisk which is the product of disks of radii $R_1, \ldots, R_n$ in the planes of the complex coordinates $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, respectively. Let also

$$\Lambda = \min_{1 \le j \le n} R_j. \tag{22}$$

The norm $\|f\|_R$ in the polydisk Δ_R is defined as

$$\|f\|_R = \sum_{|j+k|=r} |f_{j,k}| R^{j+k}. \tag{23}$$

A family of polydisks $\Delta_{\delta R}$ of radii δR, with $0 < \delta \leq 1$ will be considered below. With a minor abuse the simplified notation $\|\cdot\|_\delta$ in place of $\|\cdot\|_{\delta R}$ will be used.

The main result of this section is

LEMMA 3.1. *Let the Hamiltonian $H^{(0)}$ satisfy $\|H^{(0)}_s\|_1 \leq h^{s-1}E$ for $s \geq 1$, with some constants $h \geq 0$ and $E > 0$. Let $0 < d < 1/2$. Then there exist positive constants β and G depending on E, h, Λ, d and on $\lambda_1, \ldots, \lambda_n$ such that*

$$\|\chi_r\|_{1-d} \leq \beta^{r-1} G \quad \text{for all } r \geq 1.$$

The rest of this section is devoted to the proof. Some technical calculations are deferred to appendix B.

3.1 An arithmetic lemma

The following lemma will play a crucial role in the proof of lemma 3.1.

LEMMA 3.2. *Let $\lambda \in \mathbb{C}^n$ be such that λ_1 satisfies the non-resonance condition* (3)*. Then there exists a positive γ such that the inequality*

$$|\langle k, \lambda \rangle| \geq |k| \gamma$$

holds true for all non–zero $k \in \mathcal{K}^\sharp \cup \mathcal{K}^\natural$ defined as in (14)*.*

COROLLARY 3.3. *Let in addition the non-resonance condition* (4) *be satisfied. Then the same statement holds true for all non–zero $k \in \mathcal{K}^\sharp \cup \mathcal{K}^\natural$ defined as in* (19)*.*

The proof of the corollary is just a trivial modification of the PROOF of LEMMA 3.2. For $k \in \mathcal{K}^\sharp$ the claim is obvious, since $|\langle k, \lambda \rangle| = |k_1 \lambda_1|$. So, let $k \in \mathcal{K}^\natural$. Set $\vartheta = \max\big(|\lambda_2|, \ldots, |\lambda_n|\big)$ (for the corollary

maximize also over $|2\lambda_\nu|$ and $|\lambda_\nu \pm \lambda_{\nu'}|$ with $\nu' \neq \nu$). Pick an integer $N \geq 1 + 2\vartheta$ and set

$$\delta = \min_{\substack{k \in \mathcal{K}^\natural \\ |k| \leq N}} \big|\langle k, \lambda\rangle\big|, \quad \gamma = \min\left(\frac{\delta}{N}, \frac{|\lambda_1|}{2}\right);$$

in view of the non-resonance condition (17) one has $\delta > 0$. Then the claim of the lemma holds true with the given value of γ. Indeed, let $k \in \mathcal{K}^\natural$, so that $|k_1| = |k| - 1$. If $|k| \leq N$ then $\big|\langle k, \lambda\rangle\big| \geq \delta \geq N\gamma \geq |k|\gamma$. If $|k| > N$ use $\vartheta \leq (N-1)\delta/2$, which follows from the choice of N, and evaluate

$$\begin{aligned} \big|\langle k, \lambda\rangle\big| &\geq \big|k_1\lambda_1\big| - \vartheta \geq \big(|k| - 1\big)\delta - \tfrac{(N-1)}{2}\,\delta \\ &\geq \frac{|k| - 1}{2}\,\delta + \frac{N}{2}\,\delta - \frac{(N-1)}{2}\,\delta = |k|\frac{\delta}{2} \geq |k|\gamma. \end{aligned}$$ ∎

3.2 Generalized Cauchy estimates

Here I refer to the more restrictive hypotheses of theorem 1.2, and in particular to the spaces $\mathcal{P}$ defined as in sect. 2.3. However, the same arguments with very little simplifications apply also to the setting of sect. 2.2, which applies to theorem 1.1. I will insert short comments in parentheses concerning the latter case, where appropriate.

The estimates in this section strongly depend on a suitable splitting of all functions over the subspaces $\mathcal{P}^\sharp$, $\mathcal{P}^\natural$ and $\mathcal{P}^\flat$. At a formal level, it is useful to keep in mind the following table concerning the Poisson bracket:

$\{\cdot,\cdot\}$	$\mathcal{P}^\sharp$	$\mathcal{P}^\natural$	$\mathcal{P}^\flat$
$\mathcal{P}^\sharp$	$\mathcal{P}^\sharp$	$\mathcal{P}^\natural$	$\mathcal{P}^\flat$
$\mathcal{P}^\natural$	$\mathcal{P}^\natural$	$\mathcal{P}^\sharp \oplus \mathcal{P}^\natural \oplus \mathcal{P}^\flat$	$\mathcal{P}^\natural \oplus \mathcal{P}^\flat$
$\mathcal{P}^\flat$	$\mathcal{P}^\flat$	$\mathcal{P}^\natural \oplus \mathcal{P}^\flat$	$\mathcal{P}^\flat$

(24)

(For the subspaces defined as in sect.2.2 just remove $\mathcal{P}^\natural$ from the central case corresponding to the Poisson bracket between functions in $\mathcal{P}^\natural$.) In

view of the transformation formulæ (11) the situation to be considered is the following. A generating function $\chi \in \mathcal{W} \cap \mathcal{P}_r$ with some $r \geq 1$ is given in the form $\chi = L_{H_0}^{-1}\psi$, with known $\psi \in \mathcal{W} \cap \mathcal{P}_r$. Since $\mathcal{W} = \mathcal{R}^\sharp \oplus \mathcal{R}^\natural$ one has $\chi = \chi^\sharp + \chi^\natural$, with an obvious meaning of the notation. The operator L_χ may be applied either to a generic function $f = f^\sharp + f^\natural + f^\flat \in \mathcal{P}_s$ with $s \geq r$ or to a function in normal form $Z = Z^\sharp + Z^\natural + Z^\flat \in \mathcal{Z} \cap \mathcal{P}_m$ with $0 < m < r$ (in theorem 1.1 one has $Z^\natural = 0$). In particular one has

$$Z^\sharp = \sum_{j>1} z_j x_1^j y_1^j, \quad Z^\natural = \sum_{j>0,\, 2\leq\nu\leq n} z_{j,\nu} x_1^j y_1^j x_\nu y_\nu,$$

due to the non–resonance conditions on λ. For some non–negative δ', δ'', δ satisfying $0 \leq \max(\delta', \delta'') < \delta \leq 1/2$ the norms $\|\psi\|_{1-\delta'}$, $\|f\|_{1-\delta''}$ and $\|Z\|_{1-\delta''}$ are assumed to be known, and one looks for an estimate of the Lie derivative in a domain $\Delta_{(1-\delta)R}$. The following estimates will be used in the rest of the paper.

(i) The generating function χ is estimated by

$$\|\chi\|_{1-\delta'} \leq \frac{1}{\gamma} \|\psi\|_{1-\delta'}, \tag{25}$$

with γ as in lemma 3.2.

(ii) The general estimate for the Lie derivative of a generic function f is

$$\left\|L_\chi f\right\|_{1-\delta} \leq \frac{1}{(\delta-\delta')(\delta-\delta'')\Lambda^2} \|\chi\|_{1-\delta'} \|f\|_{1-\delta''} \tag{26}$$

with Λ as in (22). Denoting by $\left(L_{\chi^\natural} f^\flat\right)^\natural$ the projection of $L_{\chi^\natural} f^\flat$ over $\mathcal{P}^\natural$ one has

$$\left\|\left(L_{\chi^\natural} f^\flat\right)^\natural\right\|_{1-\delta} \leq \frac{4}{(\delta-\delta'')\Lambda^2} \|\chi\|_{1-\delta'} \|f\|_{1-\delta''}. \tag{27}$$

(iii) For a function Z in normal form one has

$$\left\|L_\chi(Z^\sharp + Z^\natural)\right\|_{1-\delta} \leq \frac{1}{(\delta-\delta'')\gamma\Lambda^2} \|\psi\|_{1-\delta'} \|Z\|_{1-\delta''}. \tag{28}$$

I recall the reader's attention on the missing denominator $\delta - \delta'$ in (27) and (28). This is crucial for the convergence proof. For, working out the convergence proof requires a quite accurate control of the accumulation of the divisors $\delta - \delta'$, $\delta - \delta''$ that appear in the generalized Cauchy estimates for derivatives. The scheme in the next section is specially devised in order to allow such a control.

The proof of (25) is a straightforward consequence of the definition of the norm and of (13). For, the denominators are uniformly estimated from below by γ, in view of lemma 3.2.

The proof of the estimates (26), (27) and (28) is a purely technical matter, and is deferred to appendix B.

3.3 Recursive estimates

The aim of this section is to obtain estimates for the norms of the generating functions and of the transformed Hamiltonians, at every step of the normalization procedure.

Consider a sequence of boxed domains $\Delta_{(1-\delta_r)R}$, where $\{\delta_r\}_{r\geq 1}$ is a monotonically increasing sequence of positive numbers converging to some $d < 1/2$. Let also $\delta_0 = 0$, and $d_r = \delta_r - \delta_{r-1}$ for $r \geq 1$, so that $d_r < 1$ for all positive r. The purpose is to look for estimates of the norms of the generating function χ_r and of the normal form Z_r in the polydisk $\Delta_{(1-\delta_{r-1})R}$, and of the functions $H_s^{(r)}$ in the domain $\Delta_{(1-\delta_r)R}$.

Let $\mathcal{J}_{r,s}$ for $1 < r < s$ be the set of integer arrays defined as

$$\mathcal{J}_{r,s} = \Big\{ J = \{j_1, \dots j_k\} \,:\, j_m \in \{1, \dots, r\}\,,\ 1 \leq k \leq 2(s-1)\,, \\ \sum_{m=1}^{k} \log_2 j_m \leq 2(s - 1 - \log_2 s) \Big\}. \tag{29}$$

Let also $\mathcal{J}_{0,s} = \emptyset$ for $s \geq 1$. Recalling that $\{d_r\}_{r\geq 1}$ is a sequence of positive numbers not exceeding 1 define the sequence $\{T_{r,s}\}_{0\leq r<s}$ as

$$T_{0,s} = 1, \quad T_{r,s} = \max_{J\in\mathcal{J}_{r,s}} \prod_{j\in J} d_j^{-1}. \tag{30}$$

The following properties will be used below: for $0 \le r \le r' < s$ one has

$$T_{r,s} \quad \le T_{r',s}, \tag{31}$$

$$\frac{1}{d_r^2} T_{r-1,r} T_{r',s} \quad \le T_{r',r+s}. \tag{32}$$

Checking (31) is easy: for $r = 0$ use $d_l \le 1$ for $l \ge 1$; for $r > 0$ use the inclusion relation $\mathcal{J}_{r,s} \subset \mathcal{J}_{r',s}$ for $r < r'$. In order to prove (32) remark that by definition one has

$$\begin{aligned} \frac{1}{d_r^2} T_{r-1,r} T_{r',s} &= \frac{1}{d_r^2} \max_{J \in \mathcal{J}_{r-1,r}} \prod_{j \in J} d_j^{-1} \max_{J' \in \mathcal{J}_{r',s}} \prod_{j' \in J'} d_{j'}^{-1} \\ &= \max_{J \in \mathcal{J}_{r-1,r}} \max_{J' \in \mathcal{J}_{r',s}} \prod_{j \in \{r,r\} \cup J \cup J'} d_j^{-1}. \end{aligned}$$

It is enough to prove that $\{r, r\} \cup J \cup J' =: \tilde{J} \in \mathcal{J}_{r',r+s}$. First check that

$$\begin{aligned} \#\big(\tilde{J}\big) &= 2 + \#(J) + \#(J') \le 2 + 2(r-1) + 2(s-1) \\ &= 2(r+s-1). \end{aligned}$$

On the other hand, since $1 \le j \le r-1$ for all $j \in J$ and $1 \le j' \le r'$ for all $j' \in J'$, one also has $1 \le \tilde{j} \le r'$ for all $\tilde{j} \in \tilde{J}$. Finally, evaluate

$$\begin{aligned} \sum_{\tilde{j} \in \tilde{J}} \log_2 \tilde{j} &= 2 \log_2 r + \sum_{j \in J} \log_2 j + \sum_{j' \in J'} \log_2 j' \\ &\le 2 \log_2 r + 2(r - 1 - \log_2 r) \\ &\quad + 2(s - 1 - \log_2 s) \\ &\le 2\big[r + s - 1 - (1 + \log_2 s)\big] \\ &\le 2\big[r + s - 1 - \log_2(r+s)\big], \end{aligned}$$

where the elementary inequality $1 + \log_2 s = \log_2 2 + \log_2 s = \log_2(2s) > \log_2(r+s)$ has been used (recall that $r \le r' < s$). Hence, $\tilde{J} \in \mathcal{J}_{r',r+s}$, as claimed.

I shall also use the numerical sequence $\{\mu_{r,s}\}_{r\geq 0, s\geq 0}$ defined as

$$\begin{aligned} &\mu_{0,0} = 0\,, \quad \mu_{0,s} = 1 \quad \text{for } s > 0, \\ &\mu_{r,s} = \sum_{0\leq rp<s} \mu_{r-1,r}^{p} \mu_{r-1,s-rp} \quad \text{for } r > 0 \text{ and } s \geq 0. \end{aligned} \tag{33}$$

The recursive estimates are collected in

LEMMA 3.4. *Let the Hamiltonian $H^{(0)}$ satisfy $\|H_s^{(0)}\|_1 \leq h^{s-1}E$ for some constants $h \geq 0$ and $E > 0$. Let $d_0 = 1$ and $\{d_r\}_{r\geq 1}$ be an arbitrary sequence of positive numbers satisfying $\sum_{r\geq 1} d_r = d$ with $d < 1$. Let also $\delta_0 = 0$ and $\delta_r = d_1 + \ldots + d_r$. Then for $s > r \geq 1$ the following estimates hold true:*

$$\|\chi_r\|_{1-\delta_{r-1}} \leq \mu_{r-1,r} T_{r-1,r} C^{r-1} \frac{E}{\gamma}, \tag{34}$$

$$\|Z_r\|_{1-\delta_{r-1}} \leq \mu_{r-1,r} T_{r-1,r} C^{r-1} \frac{E}{d_{r-1}}, \tag{35}$$

$$\|Z_r^{\sharp} + Z_r^{\natural}\|_{1-\delta_{r-1}} \leq \mu_{r-1,r} T_{r-1,r} C^{r-1} E, \tag{36}$$

$$\|H_s^{(r)}\|_{1-\delta_r} \leq \mu_{r,s} T_{r,s} C^{s-1} \frac{E}{d_r}, \tag{37}$$

$$\|H_s^{(r),\sharp} + H_s^{(r),\natural}\|_{1-\delta_r} \leq \mu_{r,s} T_{r,s} C^{s-1} E, \tag{38}$$

where

$$C = h + \frac{4e^2 E}{\gamma \Lambda^2}, \tag{39}$$

and $\mu_{r,s}$ and $T_{r,s}$ are the sequences defined by (30) *and* (33).

Remark that (34), (36) and (38) differ from (35) and (37), respectively, only because a divisor d_r is missing.

PROOF. By induction. For $r = 0$ only (37) and (38) are meaningful, and hold true in view of $d_0 = \mu_{0,s} = T_{0,s} = 1$ and of $h < C$. The induction consists in first proving that if (37) and (38) hold true up to $r - 1$ then (34), (35) and (36) are true for r; next proving that if (34), (35) and (36) hold true up to r then (37) and (38) are true for r.

Let $r > 0$ and put $r-1$ in place of r and r in place of s in (37) and (38). Recalling that only $H_r^{(r-1),\sharp} + H_r^{(r-1),\natural}$ is used in order to determine χ_r use the definition of the norm, the form of the solution of eq. (10) discussed in sect. 2, and the estimate (25). This immediately shows that (34), (35) and (36) are true for r provided (37) and (38) hold true for $r-1$. Coming to (37) and (38) and recalling the recursive definitions (11) there are only two kinds of terms to be estimated, namely $\frac{1}{s!}L^s_{\chi_r}Z_m$ for $1 \le m < r$ and $\frac{1}{p!}L^p_{\chi_r}H^{(r-1)}_{(s-p)r+m}$ for $0 \le p \le s$ and $0 \le m < r$. For, remarking that Z_r and $H_r^{(r-1)}$ are estimated by exactly the same quantity it is safe to estimate $\|H^{(r)}_{sr}\|$ by replacing Z_r with $H_r^{(r-1)}$ in the second of (11). This is tantamount to extending the sum in the second of (11) to $p = s-1$ and making it identical with the sum in the first of (11), with $m = 0$.

Denote $\varphi_s = L^s_{\chi_r}Z_m$, where $r > 1$, and split $\varphi_s = \varphi_s^\sharp + \varphi_s^\natural + \varphi_s^\flat$. I claim

$$\|\varphi_s\|_{1-\delta_r} \le s! \left(\tfrac{e}{d_r\Lambda}\right)^{2(s-1)} \|\chi_r\|^{s-1}_{1-\delta_{r-1}} \tfrac{D}{d_r}, \tag{40}$$

$$\|\varphi_s^\sharp + \varphi_s^\natural\|_{1-\delta_r} \le s! \left(\tfrac{2e}{d_r\Lambda}\right)^{2(s-1)} \|\chi_r\|^{s-1}_{1-\delta_{r-1}} D, \tag{41}$$

for $s \ge 1$, where

$$D = \mu_{r-1,r}\mu_{m-1,m}T_{r,r+m}C^{r+m-1}E \tag{42}$$

The proof proceeds by induction. Let $s = 1$. By the general estimate (26) one has

$$\|\varphi_1\|_{1-\delta_r} \le \frac{1}{d_r d_m \Lambda^2}\|\chi_r\|_{1-\delta_{r-1}}\|Z_m\|_{1-\delta_{m-1}}.$$

Using (34) and (35) one gets

$$\|\varphi_1\|_{1-\delta_r} \le \frac{1}{d_r}\mu_{m-1,m}\mu_{r-1,r}\,\frac{1}{d_{m-1}d_m}T_{m-1,m}T_{r-1,r}\,C^{r+m-2}\frac{E^2}{\gamma\Lambda^2},$$

so that (40) immediately follows from (31), (32) and (39).

Still keeping $s = 1$, (41) is obtained by remarking that the contributions to $\varphi_1^\sharp + \varphi_1^\natural$ come only from $L_{\chi_r}(Z_m^\sharp + Z_m^\natural)$ and $\big(L_{\chi_r^\natural} Z_m^\flat\big)^\natural$. Proceeding as above, from (28) and (27) one gets

$$\begin{aligned}\big\|\varphi_1^\sharp + \varphi_1^\natural\big\|_{1-\delta_r} \le \frac{1}{\gamma d_m \Lambda^2}\big\|H_r^{(r-1),\sharp} + H_r^{(r-1),\natural}\big\|_{1-\delta_{r-1}}\big\|Z_m\big\|_{1-\delta_{m-1}} \\ + \frac{4}{d_m\Lambda^2}\big\|\chi_r\big\|_{1-\delta_{r-1}}\big\|Z_m\big\|_{1-\delta_{m-1}}.\end{aligned}$$

Then (41) for $s = 1$ follows from (35), (38) for $r - 1$ and (39). Remark that the divisor d_r does not appear here.

Let now $s > 1$, and assume that (40) be true up to $s - 1$. Recalling that the divisor d_r due to the generalized Cauchy estimates is arbitrary, replace d_r with $\frac{s-1}{s} d_r$ in the estimates (40) and (41) for φ_{s-1}, thus getting

$$\begin{aligned}&\big\|\varphi_{s-1}\big\|_{1-\delta_r+d_r/s} \\ &\qquad \le (s-1)!\left(\frac{s}{s-1}\right)^{2s-3}\left(\frac{e}{d_r\Lambda}\right)^{2(s-2)}\|\chi_r\|_{1-\delta_{r-1}}^{s-2}\,\frac{D}{d_r}, \\ &\big\|\varphi_{s-1}^\sharp + \varphi_{s-1}^\natural\big\|_{1-\delta_r+d_r/s} \\ &\qquad \le (s-1)!\left(\frac{s}{s-1}\right)^{2s-4}\left(\frac{2e}{d_r\Lambda}\right)^{2(s-2)}\|\chi_r\|_{1-\delta_{r-1}}^{s-2}\,D.\end{aligned} \tag{43}$$

Consider first the estimate (41). Remarking that the contributions to $\varphi_s^\sharp + \varphi_s^\natural$ come only from $L_{\chi_r}\big(\varphi_{s-1}^\sharp + \varphi_{s-1}^\natural\big)$ and $\big(L_{\chi_r^\natural}\varphi_{s-1}^\flat\big)^\natural$, use (26) and (27) to estimate

$$\begin{aligned}\big\|\varphi_s^\sharp + \varphi_s^\natural\big\|_{1-\delta_r} &\le \big\|L_{\chi_r}\big(\varphi_{s-1}^\sharp + \varphi_{s-1}^\natural\big)\big\|_{1-\delta_r} + \big\|\big(L_{\chi_r^\natural}\varphi_{s-1}^\flat\big)^\natural\big\|_{1-\delta_r} \\ &\le \frac{s}{d_r^2\Lambda^2}\big\|\chi_r\big\|_{1-\delta_{r-1}}\big\|\varphi_{s-1}^\sharp + \varphi_{s-1}^\natural\big\|_{1-\delta_r+d_r/s} \\ &\quad + \frac{2s}{d_r\Lambda^2}\big\|\chi_r\big\|_{1-\delta_{r-1}}\big\|\varphi_{s-1}\big\|_{1-\delta_r+d_r/s}\end{aligned}$$

Replacing (43) in the latter expression one gets

$$\left\|\varphi_s^\sharp + \varphi_s^\flat\right\|_{1-\delta_r} \leq \frac{s!}{d_r^2\Lambda^2}\left(\frac{s}{s-1}\right)^{2s-3}\left(\frac{2e}{d_r\Lambda}\right)^{2(s-2)}\left\|\chi_r\right\|_{1-\delta_{r-1}}^{s-1} D,$$

so that (41) follows from the trivial inequality $\left(\frac{s}{s-1}\right)^{s-1} < e$. The estimate (40) is checked with a similar calculation, just taking into account that (26) must be used in order to estimate $L_{\chi_r}\varphi_{s-1}$. This produces an extra divisor d_r with respect to the calculation above.

Finally, replace (34) and (42) in (40) and (41). Using also (39), one gets

$$\begin{aligned}\left\|\varphi_s\right\|_{1-\delta_r} &\leq s!\,\mu_{r-1,r}^{s-1}\left(\frac{1}{d_r^2}\,T_{r-1,r}\right)^{s-1} T_{r,r+m}C^{sr+m-1}\frac{E}{d_r}.\\ \left\|\varphi_s^\sharp + \varphi_s^\flat\right\|_{1-\delta_r} &\leq s!\,\mu_{r-1,r}^{s-1}\left(\frac{1}{d_r^2}\,T_{r-1,r}\right)^{s-1} T_{r,r+m}C^{sr+m-1}E.\end{aligned}$$

Using $s-1$ times the inequalities (31) and (32) one easily gets

$$\left(\frac{1}{d_r^2}\,T_{r-1,r}\right)^{s-1} T_{r,r+m} \leq \left(\frac{1}{d_r^2}\,T_{r-1,r}\right)^{s-2} T_{r,2r+m} \leq \ldots \leq T_{r,sr+m}.$$

Thus one concludes

$$\begin{aligned}&\frac{1}{s!}\left\|L_{\chi_r}^s Z_m\right\|_{1-\delta_r}\\ &\qquad\leq \mu_{r-1,r}^s\mu_{m-1,m}T_{r,sr+m}C^{sr+m-1}\frac{E}{d_r},\end{aligned}\tag{44}$$

$$\begin{aligned}&\frac{1}{s!}\left\|\left(L_{\chi_r}^s Z_m\right)^\sharp + \left(L_{\chi_r}^s Z_m\right)^\natural\right\|_{1-\delta_r}\\ &\qquad\leq \mu_{r-1,r}^s\mu_{m-1,m}T_{r,sr+m}C^{sr+m-1}E.\end{aligned}\tag{45}$$

The estimate for $\frac{1}{p!}L^p_{\chi_r}H^{(r-1)}_{(s-p)r+m}$ is a minor *variazione* of the scheme above. Only the first step must be omitted. E.g., set $\varphi_p = L^p_{\chi_r}H^{(r-1)}_{(s-p)r+m}$ and proceed as follows. Using (37) for $r-1$ get

$$\left\|\varphi_0\right\|_{1-\delta_{r-1}} \leq \mu_{r-1,sr+m}T_{r,sr+m}C^{sr+m-1}E;$$

this starts the induction on p. Then proceed for $p > 0$ as above. The conclusion is

$$\begin{aligned}
&\frac{1}{p!}\ \left\|L^p_{\chi_r}H^{(r-1)}_{(s-p)r+m}\right\|_{1-\delta_r} \leq \mu^p_{r-1,r}\mu_{r-1,(s-p)r+m}T_{r,sr+m}C^{sr+m-1}\frac{E}{d_r},\\
&\frac{1}{p!}\ \left\|\left(L^p_{\chi_r}H^{(r-1)}_{(s-p)r+m}\right)^\sharp + \left(L^p_{\chi_r}H^{(r-1)}_{(s-p)r+m}\right)^\natural\right\|_{1-\delta_r}\\
&\qquad\qquad \leq \mu^p_{r-1,r}\mu_{r-1,(s-p)r+m}T_{r,sr+m}C^{sr+m-1}E.
\end{aligned}$$

Collecting the latter equations together with (44) and (45), and referring to the transformation formulæ (11) it is now an easy matter to verify that (37) and (38) hold true provided the sequence $\mu_{r,s}$ for $0 < r < s$ is defined as

$$\begin{aligned}
\mu_{0,s} &= 1 && \text{for } s > 0,\\
\mu_{r,sr+m} &= \mu^s_{r-1,r}\mu_{m-1,m} + \sum_{p=0}^{s-1}\mu^p_{r-1,r}\mu_{r-1,(s-p)r+m}\\
&&& \text{for } r \geq 2,\ s \geq 1,\ 1 \leq m < r.\\
\mu_{r,sr} &= \sum_{p=0}^{s-1}\mu^p_{r-1,r}\mu_{r-1,(s-p)r} && \text{for } r \geq 1,\ s \geq 2.
\end{aligned} \tag{46}$$

This looks quite different from (33). However, I claim that (33) is just a harmless extension of (46). Indeed, just redefine the indexes by replacing $sr + m$ with s, also accepting $s \geq 0$, which removes the implicit restriction $s > r$. This is harmless, because for $s \leq r$ one gets for (33) $\mu_{r,s} = \mu_{r-1,s}$. Therefore, in the second line one can replace $\mu_{m-1,m} = \mu_{r-1,m}$ and include it into the sum. This completes the proof of lemma 3.4. ∎

3.4 Completion of the proof of lemma 3.1

The statement of the lemma concerns only the sequence of generating functions, that are estimated by (34). The completion of the proof rests on a suitable choice of the sequence $\{d_r\}_{r\geq 1}$, that was left arbitrary, and on a suitable estimate of the sequence $\{\mu_{r,s}\}_{s\geq r\geq 0}$. As a matter of fact, only the diagonal elements of the latter sequence need to be estimated, because the estimate for the generating functions in lemma 3.4 involves only $\mu_{r-1,r} = \mu_{r,r}$.

First, pick a value for d, with $0 < d < 1/2$, and set

$$d_r = \frac{b}{r^2}, \quad b = \frac{6d}{\pi^2},$$

so that $\sum_{r\geq 1} d_r = d$ in view of $\sum_{r\geq 1} 1/r^2 = \pi^2/6$. The immediate consequence is that

$$T_{r,s} \leq \left(\frac{16}{b^2}\right)^{s-1}. \tag{47}$$

For, use the definition (30) and recall the definition (29) of $\mathcal{J}_{r,s}$; then let $J \in \mathcal{J}_{r,s}$ and evaluate

$$\prod_{j\in J} \frac{1}{d_j} \leq \frac{1}{b^{2(s-1)}} \prod_{j\in J} j^2,$$

because $\#(J) \leq 2(s-1)$. On the other hand one has

$$\log_2 \prod_{j\in J} j^2 = 2\sum_{j\in J} \log_2 j \leq 4(s-1).$$

This proves (47)

Coming to the sequence (33), the problem is to show that $\mu_{r-1,r} \leq \eta^{r-1}$ for some positive η. For, only $\mu_{r-1,r}$ enters the estimate (34). By separating the term $p = 0$ in the sum one gets

$$\begin{aligned}\mu_{r,s} &= \mu_{r-1,s} + \mu_{r-1,r} \sum_{0\leq q<s-r} \mu_{r-1,r}^q \mu_{r-1,s-r-rq} \\ &= \mu_{r-1,s} + \mu_{r-1,r}\mu_{r,s-r}.\end{aligned}$$

Putting $r-1$ in place of r and r in place of s in the latter formula one gets

$$\begin{aligned}\mu_{r-1,r} &= \mu_{r-2,r} + \mu_{r-2,r-1}\mu_{r-1,1} \\ &= \mu_{r-3,r} + \mu_{r-3,r-2}\mu_{r-2,2} + \mu_{r-2,r-1}\mu_{r-1,1} \\ &\dots \\ &\leq \mu_{0,r} + \mu_{0,1}\mu_{1,r-1} + \dots \mu_{r-2,r-1}\mu_{r-1,1} \\ &\leq \mu_{0,1}\mu_{r-2,r-1} + \dots + \mu_{r-2,r-1}\mu_{0,1}.\end{aligned}$$

The last inequality requires a justification. Just use

$$\mu_{0,r} < \mu_{1,r} < \dots < \mu_{r-1,r} = \mu_{r,r} = \dots,$$

which is an immediate consequence of the definition. Then, for $r-j \geq j$ substitute $\mu_{r-j,j} = \mu_{j-1,j}$, and for $r-j < j$ use $\mu_{r-j,j} < \mu_{j-1,j}$, and so also $\mu_{r-j,j} + 1 \leq \mu_{j-1,j}$ which gets rid of the extra term $\mu_{0,r} = 1$.

Thus, the sequence $\{\nu_r\}_{r\geq 1}$ defined as

$$\nu_1 = 1, \quad \nu_r = \sum_{j=1}^{r-1} \nu_j \nu_{r-j} \quad \text{for } r > 1 \tag{48}$$

is a majorant of $\{\mu_{r-1,r}\}_{r\geq 1}$. This is known as the Catalan's sequence, and one has

$$\nu_r = \frac{2^{r-1}(2r-3)!!}{r!} \leq 4^{r-1}, \tag{49}$$

where the common notation $(2n+1)!! = 1 \cdot 3 \cdot \ldots \cdot (2n+1)$ has been used.

Thus, we conclude that $\mu_{r-1,r} \leq 4^{r-1}$. Inserting the latter inequality and (47) in (34) the statement of lemma 3.1 follows.

4. Proof of theorem 1.1

Having established the estimate of lemma 3.1 on the sequence of generating functions it is now a standard matter to complete the proof of theorem 1.1. Hence this section will be less detailed with respect to the previous ones.

The situation to be dealt with is the following. An infinite sequence $\{\chi_r\}_{r\geq 1}$ of generating functions is given, with $\chi_r \in \mathcal{P}_{r+2}$ (a homogeneous polynomial of degree $r+2$) satisfying $\|\chi_r\|_R \leq \beta^{r-1}G$ for some real vector R with positive components and some positive β

and G. Define a corresponding sequence of canonical transformations $(x^{(r-1)}, y^{(r-1)}) = \exp(L_{\chi_r})(x^{(r)}, y^{(r)})$. By composition one also constructs a sequence $\{\mathcal{C}^{(r)}\}_{r\geq 0}$ of canonical transformations $(x^{(0)}, y^{(0)}) = \mathcal{C}^{(r)}(x^{(r)}, y^{(r)})$ recursively defined as

$$\mathcal{C}^{(0)} = \mathrm{Id}, \quad \mathcal{C}^{(r)} = \exp(L_{\chi_r}) \circ \mathcal{C}^{(r-1)},$$

Id denoting the identity operator. The problem is to prove the following statements.

(i) Every near the identity canonical transformation defined via the exponential operator $\exp(L_{\chi_r})$ is expressed as a power series which is convergent in a polydisk $\Delta_{\varrho R}$ for some positive ϱ.

(ii) For any function $f(x^{(r-1)}, y^{(r-1)})$ analytic in $\Delta_{\varrho R}$ the transformed function is analytic in the same polydisk, and moreover

$$f(x^{(r-1)}, y^{(r-1)})\Big|_{(x^{(r-1)},y^{(r-1)})=\exp(L_{\chi_r})(x^{(r)},y^{(r)})} = \big[\exp(L_{\chi_r})f\big](x^{(r)}, y^{(r)}).$$

(iii) The sequence $\{\mathcal{C}^{(r)}\}_{r\geq 0}$ of canonical transformations converges for $r \to \infty$ to a canonical transformation $\mathcal{C}^{(\infty)}$ which is analytic in a polydisk $\Delta_{(1-d)\varrho R}$ for some positive $d < 1/2$.

(iv) For any function f analytic in $\Delta_{\varrho R}$ the sequence recursively defined as $f^{(0)} = f$, $f^{(r)} = \exp(L_{\chi_r})f^{(r-1)}$ converges for $r \to \infty$ to a function $f^{(\infty)}$ that is analytic in $\Delta_{(1-d)\varrho R}$, and moreover one has

$$f^{(\infty)} = f \circ \mathcal{C}^{(\infty)}.$$

The statement (i) actually reduces to Cauchy's proof of the existence and uniqueness of the local solution of an analytic system of differential equations. For, the transformation defined via the exponential operator is the time–one canonical flow induced by the Hamiltonian vector field generated by χ_r. The statement (ii) actually claims that the substitution of variables in a function f may be effectively replaced by the application of the exponential operator to f; this is indeed the basis of the algorithm for constructing the normal form used in sect. 2. A detailed proof of both

these statements may be found, e.g., in [9]; however, the reader may be able to reconstruct the proof by following the hints in [3].

The proof of (iii) rests on the following remarks. In the polydisk $\Delta_{\varrho R}$ one has $|\chi_r(x,y)| \le \varrho^{r+2}\,\|\chi_r\|_{\varrho R}$; this, in turn, implies that $\left|x^{(r)} - x^{(r-1)}\right| \sim \beta^{r-1}\varrho^{r+2}$ and $\left|y^{(r)} - y^{(r-1)}\right| \sim \beta^{r-1}\varrho^{r+2}$. The geometric bound on the latter quantities implies that $\sum_{r>1}\left|x^{(r)} - x^{(r-1)}\right|$ and $\sum_{r>1}\left|y^{(r)} - y^{(r-1)}\right|$ behave like geometric series, i.e., converge for ϱ small enough. Thus, the claim follows from Weierstrass theorem. Finally, the statement (iv) follows from (ii) being true for all $r > 0$, which implies that both sequences $f^{(r)} = \mathcal{C}^{(r)} f$ and $f \circ \mathcal{C}^{(r)}$ converge to the same limit. This concludes the proof of theorem 1.1.

APPENDIX A: JUSTIFICATION OF THE NORMALIZATION ALGORITHM

Justifying the normalization algorithm of sect. 2 is just matter of rearranging terms in the expansion of $\exp(L_{\chi_r})H^{(r-1)}$. Considering first H_0 and $H_r^{(r-1)}$ together, one has

$$\begin{aligned}\exp(L_{\chi_r})\left(H_0 + H_r^{(r-1)}\right) = H_0 + &L_{\chi_r}H_0 + \sum_{s\ge 2}\frac{1}{s!}L_{\chi_r}^s H_0 \\ &+ H_r^{(r-1)} + \sum_{s\ge 1}\frac{1}{s!}L_{\chi_r}^s H_r^{(r-1)}.\end{aligned}$$

Here, H_0 is the first term in the transformed Hamiltonian $H^{(r)}$ in (9). In view of (10) one has $L_{\chi_r}H_0 + H_r^{(r-1)} = Z_r$, which kills the unwanted term $H_r^{(r-1)}$ and replaces it with the normalized term Z_r. The two sums may be collected and simplified by calculating

$$\begin{aligned}\sum_{s\ge 2}\frac{1}{s!}L_{\chi_r}^s\, H_0 + \sum_{s\ge 1}\frac{1}{s!}L_{\chi_r}^s H_r^{(r-1)} &\\ = \sum_{s\ge 2}\frac{1}{(s-1)!}L_{\chi_r}^{s-1}\left[\frac{1}{s}\left(L_{\chi_r}H_0 + H_r^{(r-1)}\right) + \frac{s-1}{s}H_r^{(r-1)}\right] &\\ = \sum_{s\ge 2}\frac{1}{(s-1)!}L_{\chi_r}^{s-1}\left(\frac{1}{s}Z_r + \frac{s-1}{s}H_r^{(r-1)}\right).&\end{aligned}$$

Here, both $L^{s-1}_{\chi_r} Z_r$ and $L^{s-1}_{\chi_r} H^{(r-1)}_r$ are homogeneous polynomials of degree $sr+2$, that are added to $H^{(r)}_{sr}$ in the second of (11).

Proceed now by transforming the functions $Z_1, \ldots, Z_{r-1}$ that are already in normal form. Recall that no such term exists for $r = 1$. For $r > 1$ calculate

$$\exp(L_{\chi_r}) Z_m = Z_m + \sum_{s \geq 1} \frac{1}{s!} L^s_{\chi_r} Z_m, \quad \text{for } 1 \leq m < r.$$

The term Z_m is copied into $H^{(r)}$ in (9). The term $L^s_{\chi_r} Z_m$ is a homogeneous polynomial of degree $sr + m + 2$ that is added to $H^{(r)}_{sr+m}$ in the first of (11).

Finally, consider all terms $H^{(r-1)}_s$ with $s > r$, that may be written as $H^{(r-1)}_{lr+m}$ with $l \geq 1$ and $0 \leq m < r$, the case $l = 1\,,\ m = 0$ being excluded. One gets

$$\exp(L_{\chi_r}) H^{(r-1)}_{lr+m} = \sum_{p \geq 0} \frac{1}{p!} L^p_{\chi_r} H^{(r-1)}_{lr+m}$$

where $L^p_{\chi_r} H^{(r-1)}_{lr+m}$ is a homogeneous polynomial of degree $(p+l)r+m$. Collecting all homogeneous terms with $m = 0$, $l \geq 2$ and $p + l = s \geq 2$ one gets $\sum_{p=0}^{s-2} \frac{1}{p!} L^p_{\chi_r} H^{(r-1)}_{(s-p)r}$, that is added to $H^{(r)}_{sr}$ in the second of (11). Similarly, collecting all homogeneous terms with $0 < m < r,\ l \geq 1$ and $p+l = s \geq 1$ one gets $\sum_{p=0}^{s-1} \frac{1}{p!} L^p_{\chi_r} H^{(r-1)}_{(s-p)r+m}$, that is added to $H^{(r)}_{sr+m}$ in the first of (11). The latter case does not occur for $r = 1$. This completes the justification of the formal algorithm.

APPENDIX B: TECHNICAL CALCULATIONS

The aim is to check the estimates (26), (27) and (28). Write, generically, $\chi = \sum_{j,k} c_{j,k} x^j y^k$ and $f = \sum_{j,k} f_{j,k} x^j y^k$. Then compute

$$L_\chi f = \sum_{j,k,j',k'} \sum_{l=1}^{n} \frac{j'_l k_l - j_l k'_l}{x_l y_l} c_{j,k} f_{j',k'} x^{j+j'} y^{k+k'}. \tag{50}$$

Using the definition of norm evaluate

$$\begin{aligned}
&\|L_\chi f\|_{1-\delta} \\
&\le \sum_{j,k,j',k'} \sum_{l=1}^{n} \frac{|j'_l k_l - j_l k'_l|}{R_l^2} |c_{j,k}|\, |f_{j',k'}| (1-\delta)^{|j+k|+|j'+k'|-2} R^{j+k+j'+k'} \\
&\le \frac{1}{\Lambda^2} \sum_{j,k} \sum_{j',k'} \sum_{l=1}^{n} \left|j'_l k_l - j_l k'_l\right|\, |c_{j,k}| \left((1-\delta') - (\delta - \delta')\right)^{|j+k|-1} R^{j+k} \\
&\quad \times |f_{j',k'}| \left((1-\delta'') - (\delta - \delta'')\right)^{|j'+k'|-1} R^{j'+k'}.
\end{aligned}$$

If f is a generic function, then in view of $|j_l| \le |j+k|$ and $k_l \le |j+k|$ one has

$$\sum_{l=1}^{n} |j'_l k_l - j_l k'_l| < |j+k| \sum_{l=1}^{n} |j'_l + k'_l| = |j+k| \cdot |j'+k'|. \tag{51}$$

Replacing in the estimate above and using the elementary inequality

$$m(\lambda - x)^{m-1} < \frac{\lambda^m}{x} \quad \text{for } 0 < x < \lambda \text{ and } m \ge 1 \tag{52}$$

one gets

$$\begin{aligned}
\|L_\chi f\|_{1-\delta} &\le \frac{1}{\Lambda^2} \sum_{j,k} |c_{j,k}|\, |j+k| \left((1-\delta') - (\delta - \delta')\right)^{|j+k|-1} R^{j+k} \\
&\quad \times \sum_{j',k'} |f_{j',k'}|\, |j'+k'| \left((1-\delta'') - (\delta - \delta'')\right)^{|j'+k'|-1} R^{j'+k'} \\
&\le \frac{1}{(\delta - \delta')(\delta - \delta'')\Lambda^2} \sum_{j,k} |c_{j,k}|\, (1-\delta')^{|j+k|} R^{j+k} \\
&\quad \times \sum_{j',k'} |f_{j',k'}| (1-\delta'')^{|j'+k'|} R^{j'+k'},
\end{aligned} \tag{53}$$

from which (26) immediately follows in view of the definition of the norm.

In order to prove (27) recall that $\chi^\natural$ contains only monomials $c_{j,k}x^j y^k$ with $j+k \in \mathcal{K}^\natural$. The projection $\left(L_{\chi^\natural} f^\flat\right)^\natural$ is just part of the general expression (50). In particular the value $l=1$ in the sum must be discarded because the resulting monomials belong to $\mathcal{P}^\flat$. Moreover, for $k \in \mathcal{K}^\natural$ one has $\sum_{l=2}^n |j_l + k_l| \le 2$. Thus the estimate (51) may be replaced by

$$\sum_{l=2}^n |j'_l k_l - j_l k'_l| < 2 \sum_{l=2}^n |j'_l + k'_l| = 2|j' + k'|. \tag{54}$$

Hence the inequality (52) must be used only for the term involving $|j'+k'|$, and there is no need to introduce the divisor $\delta - \delta'$. Use instead $1/(1-\delta) < 2$ in view of $\delta < 1/2$.

Coming finally to (28), replace f in the general expression (50) by $Z^\sharp + Z^\natural = \sum_{\nu \in \mathcal{K}^\sharp \cup \mathcal{K}^\sharp} z_{\nu,\nu} x^\nu y^\nu$. Recall also that the coefficients $c_{j,k}$ of χ have the form $c_{j,k} = \frac{\psi_{j,k}}{\langle k-j,\lambda\rangle}$, in view of (13). Then (51) may be replaced by

$$\sum_l |\nu_l (j_l - k_l)| \le |\nu| \sum_l |j_l - k_l| \le |\nu|\,|j-k| \tag{55}$$

On the other hand, by lemma 3.2 one has

$$|c_{j,k}| \le \frac{|\psi_{j,k}|}{|j-k|\gamma},$$

so that the factor $|j-k|$ in (55) is compensated by the divisor here. This removes the need to introduce the divisor $\delta - \delta'$ in the rest of the estimates. Use instead $(1-\delta)^{|j+k|-1} \le 2(1-\delta)^{|j+k|}$, which holds true

in view of $\delta < 1/2$. Then (53) is replaced by

$$\begin{aligned}
&\|L_\chi(Z^\sharp + Z^\natural)\|_{1-\delta} \\
&\quad \le \frac{1}{\Lambda^2}\sum_{j,k}\frac{2}{\gamma}|\psi_{j,k}|\,(1-\delta)^{|j+k|}R^{j+k} \\
&\qquad \times \sum_{\nu\in\mathcal{K}^\sharp}|z_{\nu,\nu}|\,|\nu|\big((1-\delta'')-(\delta-\delta'')\big)^{2|\nu|-1}R^{2\nu} \\
&\quad \le \frac{1}{(\delta-\delta'')\gamma\Lambda^2}\sum_{j,k}|\psi_{j,k}|\,(1-\delta')^{|j+k|}R^{j+k} \\
&\qquad \times \sum_{\nu\in\mathcal{K}^\sharp}|z_{\nu,\nu}|(1-\delta'')^{2|\nu|}R^{2\nu}.
\end{aligned}$$

Thus (28) follows in view of the definition of the norm.

REFERENCES

[1] G.D. Birkhoff: *Dynamical systems*, New York (1927).

[2] T.M. Cherry: *On the solutions of Hamiltonian systems in the neighborhood of a singular point*, Proc. London Math. Soc., Ser. 2, **27**, 151–170 (1926).

[3] A. Giorgilli: *Quantitative methods in classical perturbation theory*, proceedings of the Nato ASI school "From Newton to chaos: modern techniques for understanding and coping with chaos in N–body dynamical systems", A.E. Roy e B.D. Steves eds., Plenum Press, New York (1995).

[4] A. Giorgilli: *Classical constructive methods in KAM theory*, PSS, **46**, 1441–1451 (1998).

[5] A. Giorgilli: *Unstable equilibria of Hamiltonian systems*, Disc. and Cont. Dynamical Systems, Vol. 7, N. 4, 855–871 (2001).

[6] A. Giorgilli and U. Locatelli: *Kolmogorov theorem and classical perturbation theory*, ZAMP **48**, 220–261 (1997).

[7] A. Giorgilli and U. Locatelli: *On classical series expansions for quasi–periodic motions*, MPEJ **3** N. 5 (1997).

[8] A. Giorgilli and U. Locatelli: *A classical self–contained proof of Kolmogorov's theorem on invariant tori*, in *Hamiltonian systems with three or more degrees of freedom*, Carles Simó ed., NATO ASI series C, Vol. 533, Kluwer Academic Publishers, Dordrecht–Boston–London (1999).

[9] W. Gröbner: *Die Lie–Reihen und Ihre Anwendungen*, VEB Deutscher Verlag der Wissenschaften (1967).

[10] A.M. Lyapunov: *The General Problem of the Stability of Motion* (In Russian), Doctoral dissertation, Univ. Kharkov (1892). French translation in: *Problème général de la stabilité du mouvement*, Annales de la Faculté des Sciences de Toulouse, deuxième série, Tome IX, 203–474 (1907). Reprinted in: Ann. Math. Study, Princeton University Press, n. 17, (1949).

[11] J.K. Moser: *On the generalization of a theorem of A. Liapounoff*, Comm. Pure Appl. Math. **11**, 257–271 (1958).

[12] C.L. Siegel and J.K. Moser: *Lectures in Celestial Mechanics*, Springer–Verlag, Berlin Heidelberg New York (1971).

 Istituto Lombardo (Rend. Scienze) 146, 161-172 (2012)

GLI ACCELERATORI NUCLEARI NELLA TERAPIA DEI TUMORI

DOMENICO SCANNICCHIO (*)

Nota presentata dal m.e. Ferdinando Borsa
(Adunanza del 24 Maggio 2012)

SUNTO. – Vengono riportati in forma schematica i principali vantaggi dell'adroterapia rispetto alla radioterapia convenzionale. Dopo aver passato in rassegna le principali caratteristiche degli acceleratori di particelle vengono descritte le attrezzature del Centro Nazionale di Adroterapia Oncologica, in particolare i sistemi di posizionamento del paziente e di rilascio della dose di radiazione.

***

ABSTRACT. – The main advantages of hadrontherapy against the traditional radiotherapy are reported. After a brief review of particles accelerators, the facilities of the 'Centro Nazionale di Adroterapia Oncologica' (National Center of Oncological Hadrontherapy) in Pavia are described with details on the patient positioning and beam irradiation systems.

1. INTRODUZIONE

Da oltre 60 anni vengono sviluppati sistemi di accelerazione di particelle elettricamente cariche per impiegarle come proiettili al fine di indagare la struttura della materia (atomi, nuclei e particelle elementari). Solo da circa 40 anni questi fasci di particelle vengono impiegati a scopo terapeutico, in particolare per distruggere cellule neoplastiche (tumori).

(*) Dipartimento di Fisica, Università degli Studi di Pavia, Italy.
E-mail: domenico.scannicchio@pv.infn.it

Prima di descrivere tale applicazione degli acceleratori è opportuno prendere in considerazione le particelle utilizzate come proiettili. Come noto la materia è formata da atomi e/o da strutture costituite da molti atomi diversi (molecole e macromolecole), con ciascun atomo costituito da una nuvola di elettroni che circonda un nucleo centrale, centomila volte più piccolo dell'atomo stesso, dotato di una carica elettrica contraria a quella degli elettroni orbitanti (l'atomo normalmente è elettricamente neutro) e contenente quasi tutta la massa dell'atomo.

I proiettili prodotti dagli acceleratori sono le particelle citate, gli *elettroni* (di massa molto piccola) e quelle costituenti i nuclei: *protoni* e neutroni (2000 volte più massicci degli elettroni) oppure i nuclei stessi, chiamati *ioni* essendo dotati di carica elettrica (positiva). Oltre a queste particelle gli acceleratori permettono la formazione di fasci di radiazione elettromagnetica, cioè raggi X e raggi gamma di energia molto elevata.

Per comprendere l'efficacia di particelle e radiazioni X nel distruggere le cellule tumorali è indispensabile conoscere in che modo esse interagiscono con la materia in cui si propagano, diminuendo gradualmente la propria energia iniziale che viene rilasciata al materiale stesso con molteplici processi in successione (*processi secondari*). Si definisce *dose assorbita* il rapporto tra l'energia assorbita dal materiale e la massa del materiale, la cui unità di misura nel Sistema Internazionale è il gray (Gy), corrispondente ad un joule per chilogrammo. Radiazioni X e particelle (elettroni, protoni e ioni) rilasciando energia nel materiale gradualmente perdono energia fino ad arrestarsi: la distanza media percorsa si chiama *range* ed è tanto più elevato quanto maggiore è l'energia iniziale (*Fig. 1*, dove il mezzo attraversato è l'acqua, principale costituente del materiale biologico).

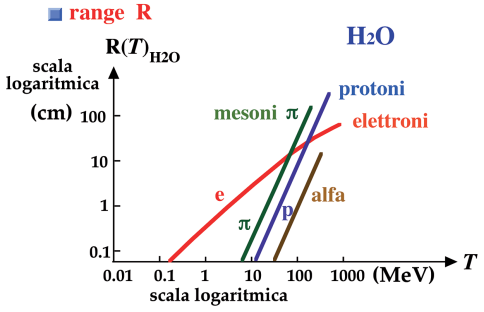


Fig. 1 – Viene riportato il percorso medio (range) in acqua di vari tipi di particelle.

La cura dei tumori mediante fasci di particelle o radiazioni X è chiamata radioterapia. Quella *convenzionale* impiega raggi X ed elettro-

ni. La radioterapia che impiega particelle nucleari (protoni o nuclei stessi) è detta *adroterapia* (le particelle nucleari sono chiamate adroni, cioè particelle che interagiscono fra loro con una forza attrattiva enorme, chiamata interazione nucleare forte).

L'adroterapia presenta in generale 3 vantaggi relativamente alla radioterapia convenzionale: elevato rilascio di energia a fine percorso, collimazione del fascio nel penetrare la materia, modalità di rilascio dell'energia alle macromolecole presenti nelle cellule, in particolare al DNA. Vediamo ora i dettagli di queste 3 caratteristiche vantaggiose.

2. Rilascio di energia nel materiale attraversato

In *Fig. 2* sono riportati i rilasci di energia da parte di varie radiazioni (corpuscolari e raggi X) man mano esse penetrano nel materiale (acqua): come si può facilmente rilevare i protoni rilasciano gran parte della loro energia a fine percorso (picco di Bragg), mentre raggi X e gamma ed elettroni rilasciano l'energia prevalentemente all'ingresso del materiale. Calibrando l'energia dei protoni l'energia può essere concentrata in una determinata regione danneggiandola, contrariamente alle altre radiazioni che danneggiano il materiale prevalentemente all'ingresso con limitata distruzione del bersaglio in profondità. Questa caratteristica dei protoni (picco di Bragg) viene ulteriormente accentuata nel caso di radiazioni costituite da nuclei (ad esempio di ioni di carbonio o di ossigeno). In pratica gli adroni agiscono in modo simile a bombe di profondità nel materiale.

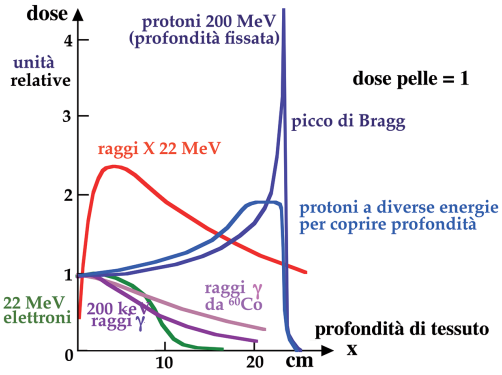


Fig. 2 – Rilascio di energia lungo il percorso in acqua da parte di varie particelle. Per i protoni monoenergetici è evidente il picco di Bragg.

3. Collimazione del fascio

In *Fig. 3a* e *3b* sono mostrate le sezioni di fasci di elettroni, raggi X, protoni e ioni carbonio. A partire da un fascio di date dimensioni (rettangolo iniziale) gli elettroni già a 10 cm di profondità si sono allargati in quello che è chiamato uno sciame elettromagnetico (3a) e analogamente i raggi X a 15 cm di profondità (3b) mentre i protoni e gli ioni carbonio (la cui energia comporta una profondità di fine percorso di 15,5 cm) alle diverse profondità restano molto collimati. Nella pratica ciò significa che elettroni e radiazione X e gamma causano distruzioni lateralmente al fascio, dove il materiale biologico se sano deve subire il minor danno possibile e se vitale per l'organismo non deve assolutamente essere danneggiato.

La particolare elevata collimazione del fascio di adroni permette di "centrare" il bersaglio con elevata precisione (dell'ordine di un millimetro), precisione che tuttavia richiede un posizionamento del paziente dal punto di vista tecnico assai complesso, come vedremo nel seguito, trattandosi di un organismo vitale e non di un corpo inerte.

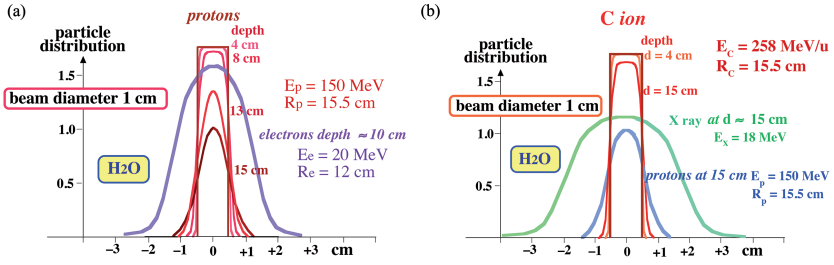


Fig. 3 – a) Sezioni di un fascio di protoni in acqua a varie profondità paragonate alla sezione di un identico fascio di elettroni alla profondità di 12 cm. Si osservi la maggiore collimazione del fascio di protoni rispetto agli elettroni che diffondono lateralmente. b) Sezioni di un fascio di ioni carbonio a varie profondità paragonate con la sezione di un fascio identico di raggi X: la collimazione degli ioni carbonio è molto elevata.

4. Modalità di rilascio dell'energia

Quando una radiazione (corpuscolare o raggi X) attraversa la materia, in particolare la materia biologica, rilascia energia agli atomi e molecole del mezzo: questa interazione con le strutture molecolari comportano soprattutto la rottura di legami chimici. Quando la macromolecola coinvolta è il DNA si hanno rotture di legami che spesso il DNA riesce ad autoriparare.

La grandezza fisica che rappresenta il rilascio di energia per unità di percorso è chiamato LET (Linear Energy Transfer). Le varie radiazioni presentano un LET diverso: mentre raggi X, raggi gamma, elettroni sono radiazioni a basso LET, cioè creano pochi danni per unità di percorso, gli adroni (protoni ma soprattutto ioni) hanno un LET elevato creando un tale numero di danni (rotture di legami chimici) da rendere in particolare la macromolecola DNA irreparabile. La distruzione del DNA per una cellula significa l'impossibilità di replicarsi e quindi la sua morte.

In realtà le macromolecole nelle cellule si trovano immerse in un ambiente acquoso e le radiazioni a basso LET provocano la rottura delle molecole d'acqua in due frammenti dotati di elevata energia (acquisita dall'interazione con la radiazione) chiamati *radicali liberi*. Questi radicali liberi sono molto reattivi e interagiscono chimicamente con le macromolecole presenti nei dintorni: questa interazione chimica danneggia le macromolecole causando la rottura di legami chimici, legami che il DNA per sua natura è in parte in grado di riparare. Dunque le radiazioni a basso LET provocano un danno di *tipo chimico* al materiale biologico.

Nel caso della radiazione ad alto LET, oltre al danno chimico, l'elevato numero di rottura di legami nelle macrostrutture (DNA) provoca un danno di *tipo fisico*, praticamente "tagliando" la macromolecola in modo non riparabile. In effetti quando con un coltello affilato si taglia la carne, l'azione di "tagliare" comporta la rottura di innumerevoli legami chimici (dell'ordine del numero di Avogadro).

In conclusione l'efficacia distruttiva di cellule è assai maggiore per la radiazione costituita da adroni rispetto alle radiazioni della radioterapia convenzionale: per gli adroni il danno fisico irreparabile si aggiunge a quello chimico da radicali liberi. In *Fig. 4* sono mostrate delle simulazioni di danni determinati da diverse radiazioni, nella scala del DNA mostrato a sinistra. L'efficacia dell'adroterapia si evidenzia in modo particolare nel caso di *tumori radio-resistenti* alla terapia convenzionale.

5. Piani di trattamento

Per eseguire un irraggiamento è indispensabile predisporre un piano di trattamento che viene sviluppato a partire dalle immagini radiografiche e tomografiche che forniscono la posizione e la forma tridimensionale del tumore e dei tessuti circostanti, dalla natura delle cellule neoplastiche e dalla prossimità di organi vitali. Il piano di trattamento definisce la dose da impartire in ciascun irraggiamento, i para-

metri del fascio (energia, intensità, dimensione), le direzioni di irraggiamento (definite *campi*) e il numero di irraggiamenti. Il piano di trattamento viene messo a punto dalla collaborazione del medico radioterapista e del fisico sanitario. Tutti i relativi dettagli vengono inseriti in un sistema di calcolo che provvede alla esecuzione del piano, come descritto nei paragrafi successivi.

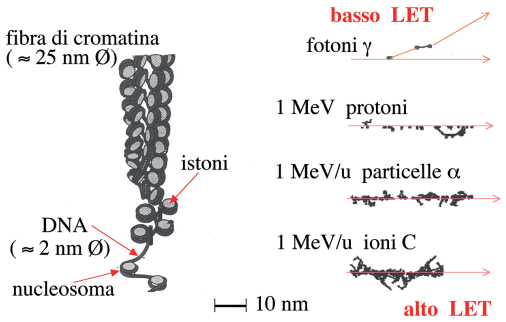


Fig. 4 – Rilascio di energia di varie radiazioni (corpuscolari e non): a destra ogni punto rappresenta una interazione, a sinistra, sulla stessa scala dimensionale, è mostrata la doppia elica del DNA avvolta sulle strutture chiamate istoni. Il numero di interazioni per gli ioni carbonio (in basso a destra) sovrapposte al DNA sulla sinistra è tale da creare danni non riparabili al DNA.

6. Acceleratori di particelle

Dalla fine degli anni ’30 del secolo scorso sono stati sviluppati vari tipi di acceleratori di particelle circolari e lineari. L’accelerazione di particelle dotate di carica elettrica avviene impiegando campi elettrici in successione (acceleratori lineari) o attraversati numerose volte dalle particelle (acceleratori circolari).

In linea generale per ottenere fasci di particelle di elevata energia sono impiegati acceleratori circolari in cui le particelle nell’orbita circolare attraversano una regione dove un intenso campo elettrico imprime una accelerazione che viene ripetuta ad ogni passaggio fino a raggiungere l’energia richiesta ed essere poi espulsi con un campo magnetico deviatore verso la zona del bersaglio (sala di trattamento dove si trova il paziente). Le particelle sono trattenute nell’orbita circolare da campi magnetici: magneti curvanti e una sistema di accelerazione (*cavità a radiofrequenza*) costituiscono la base di tali acceleratori.

Nel caso degli adroni (protoni e ioni) l'acceleratore di solito impiegato è chiamato *sincrotrone*. Un sistema di produzione delle particelle da accelerare (*sorgente*) invia "pacchetti" di particelle con una opportuna frequenza in un pre-acceleratore (di solito lineare) che con campi elettrici accelera ad una energia minima di ingresso nel sincrotrone in cui i pacchetti di particelle in orbita circolare sono accelerati da campi elettrici alternati in opportuni dispositivi (*cavità a radiofrequenza*). Ad ogni passaggio nella cavità il pacchetto incrementa la propria energia per cui, per permanere sempre nell'orbita circolare, il campo magnetico dei magneti curvanti viene aumentato, in contemporanea all'accelerazione. Il processo di accelerazione procede con numerosissime orbite, in tempi dell'ordine di frazione di secondo, fino all'energia richiesta.

I pacchetti di particelle sono composti da particelle della stessa carica elettrica tra le quali agisce quindi la repulsione coulombiana. Per evitare la disgregazione del pacchetto in brevissimo tempo, tra i magneti curvanti sono inseriti magneti focheggianti il pacchetto di particelle, mantenendolo compattato. Come detto, un campo magnetico devia il fascio di pacchetti una volta raggiunta l'energia desiderata inviandolo verso la sala del bersaglio tramite una successione di campi magnetici curvanti e focheggianti (*linea di trasporto magnetica*).

7. Il Centro Nazionale di Adroterapia Oncologica

Dopo circa 10 anni di progettazioni, proposte e richieste ai vari governi che si sono succeduti, a partire dal 2002 viene finanziata la costruzione del Centro Nazionale di Adroterapia Oncologica (CNAO) basato su un sincrotrone da 250 MeV di energia massima per i protoni e con 3 sale di trattamento (*Fig. 5*). Il sistema è in grado di accelerare nella stessa ciambella (sotto vuoto spinto) sia i protoni che gli ioni carbonio. Quest'ultima importante caratteristica permette di irraggiare con protoni e poi con ioni senza che il paziente venga spostato e riposizionato. Il Centro è situato a Pavia in un terreno compreso tra gli IRCCS S. Matteo, la Fondazione Mondino e la Fondazione Maugeri. Al fine di evitare sprofondamenti e deformazioni dell'edificio a causa dell'elevato peso del cemento necessario, essendo il terreno alluvionale, l'acceleratore del Centro è stato costruito su una piattaforma di cemento armato spessa 2 metri che poggia su oltre 1500 palafitte costituite da cilindri di cemento armato alti 12 metri del diametro di 1,5 metri.

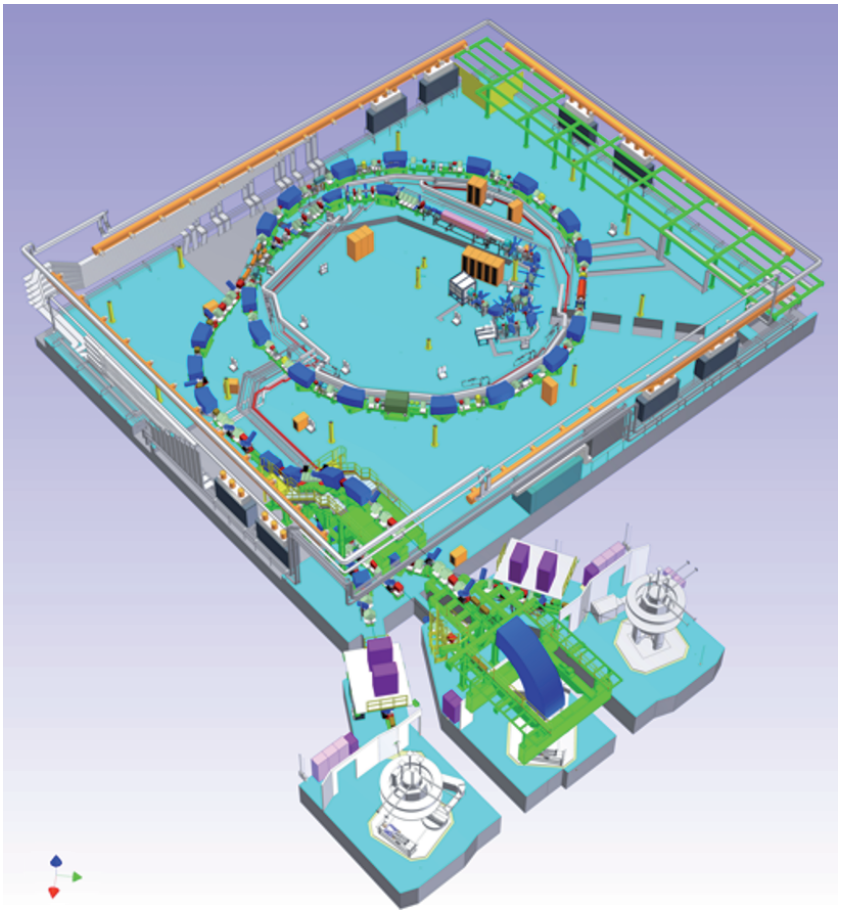


Fig. 5 – Il sincrotrone del CNAO con le 3 sale di trattamento. In blu sono i magneti curvanti, il cilindro rosso a destra nell'anello circolare è la cavità risonante di accelerazione, al centro il pre-acceleratore lineare in rosa e i due cubi grigio-bianchi che contengono le sorgenti (di protoni e di ioni carbonio). La sala di trattamento centrale riceve dall'alto, tramite il magnete curvante blu, un fascio anche verticale. I lunghi cilindri arancioni ai lati sono il sistema di condizionamento della sala mentre le righe grigie al suolo contengono le tubature dell'acqua di raffreddamento dei numerosi magneti.

Il sincrotrone ha un diametro di circa 21 metri. Le rilevanti dimensioni sono determinate dalla massa dei protoni, 2000 volte più massicci degli elettroni, e degli ioni carbonio (di massa 24mila volte quella dell'elettrone) e quindi altrettanto più difficili da curvare a parità

di campi magnetici: infatti gli acceleratori di elettroni (che producono anche fasci di raggi X) nella radioterapia convenzionale sono lineari e hanno dimensioni di circa solo 1 o 2 metri.

Il sincrotrone è stato basato su magneti tradizionali soprattutto per motivi di sicurezza e stabilità nel funzionamento a scopi terapeutici, escludendo l'impiego di magneti superconduttori (di complessa manutenzione e non sicura affidabilità) che avrebbero permesso una drastica riduzione delle dimensioni dell'acceleratore. Delle tre sale di trattamento quella centrale, oltre che disporre di un fascio orizzontale, permette anche l'irraggiamento verticale tramite un sistema di trasporto magnetico del fascio che devia il fascio verso l'alto fino a farlo giungere verticalmente sul paziente. In *Fig. 6* è mostrata una visione d'insieme del sincrotrone.

Fig. 6 – Visione d'insieme del sincrotrone.
La cavità risonante in cui avviene l'accelerazione è in rosso in basso a sinistra.

8. Tecniche di posizionamento

Come riferito in precedenza, la posizione del bersaglio (e quindi del paziente) deve avere una precisione millimetrica, non facile da ottenere e tenere sotto controllo per un organismo vitale. A questo scopo sono state applicate tecnologie di posizione estremamente accurate, derivate da atti-

vità spaziali. Infatti il paziente viene cosparso nella zona cutanea da irraggiare (sotto la quale vi è il tumore da distruggere) con *markers* che riflettono la radiazione infrarossa e contenenti pallini di piombo, visibili nelle radiografie (nelle regioni capo-collo o torace o addome). Con radiografie e/o tomografie in raggi X si studia il volume da irraggiare avendo visibili anche i pallini di piombo contenuti nei markers: in questo modo si correlano le posizioni nelle 3 dimensioni dei pallini col volume del tumore. In un locale in prossimità della sala di trattamento il paziente viene disteso su un tavolo di carbonio, anch'esso dotato di markers, e posizionato, note la posizione dei markers sul proprio corpo e la posizione e forma del tumore, in riferimento ad alcuni markers sul tavolo. Il paziente viene posizionato seguendo le indicazioni del piano di trattamento e poi trasportato nell'adiacente sala di trattamento sul fascio dove telecamere digitali riprendono le immagini dei markers illuminati da radiazione infrarossa a 100 Hertz (100 volte al secondo). Le posizioni dei markers sul paziente e sul tavolo devono corrispondere a quelle richieste dal piano di trattamento entro l'accuratezza di circa 1 millimetro (le posizioni dei markers ottenute dalle immagini digitali infrarosse sono confrontate in tempo reale con quanto richiesto dal sistema di calcolo). Durante l'irraggiamento qualsiasi spostamento dei markers maggiore della precisione richiesta (ad esempio uno starnuto del paziente) causa l'interruzione istantanea del trattamento. Tutto ciò per la sicurezza del paziente, data la notevole efficacia distruttiva degli adroni. Il posizionamento per la sua accuratezza e importanza fondamentale dura alcune decine di minuti. Il tavolo su cui è posto il paziente è incastrato su un braccio robotico che su comando del calcolatore, lo sposta e lo orienta, secondo il piano di trattamento. In *Fig. 7* è mostrata una sala di trattamento mentre in *Fig. 8* è mostrata la disposizione e l'impiego dei markers nel caso di un tumore cerebrale (nella parte destra sono visibili le telecamere digitali e gli apparati radiografici per il controllo della posizione).

9. Tecniche di irraggiamento

Anche l'irraggiamento, eseguito seguendo il piano di trattamento con metodica attiva, è tecnologicamente complesso. Il fascio sottile (*pencil beam*) a una determinata energia, proveniente dall'acceleratore viene deviato da opportuni magneti lungo le coordinate in x e in y, esegue una scansione di una sezione del tumore distruggendola. Variando l'energia del fascio vengono distrutte via via le altre sezioni conformando la fettina

da distruggere al contorno noto del tumore seguendo sempre le istruzioni del piano di trattamento (vedi *Fig. 9*). L'intera procedura di irraggiamento dura solo alcuni minuti. Quando il piano di trattamento prevede l'impiego non solo di protoni ma anche di ioni (ad esempio per distruggere la parte radio-resistente del tumore), il paziente non viene spostato mentre in tempi relativamente brevi si procede all'accensione della sorgente di ioni e alla accelerazione del nuovo fascio. Lungo la ciambella entro cui passano le particelle opportuni rivelatori permettono di conoscere e controllare l'intensità e la conformazione del fascio in tempo reale.

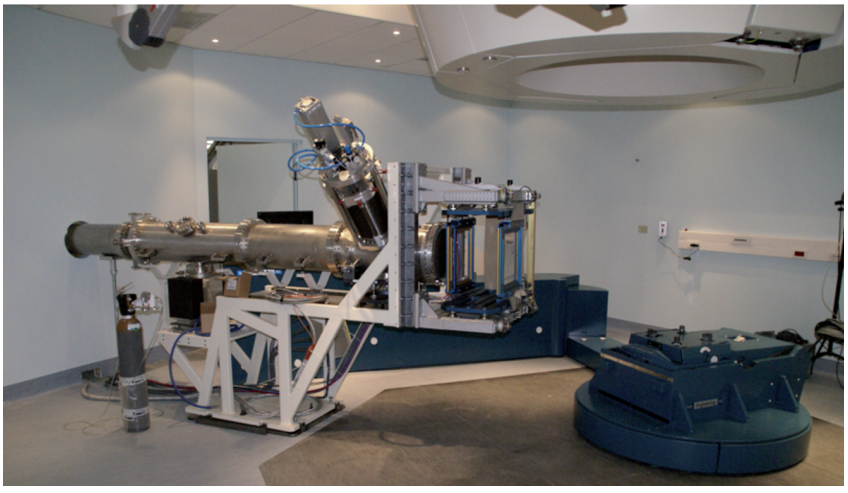


Fig. 7 – Sala di trattamento laterale in cui è visibile il lettino in carbonio. Dal rettangolo sulla destra esce il fascio.

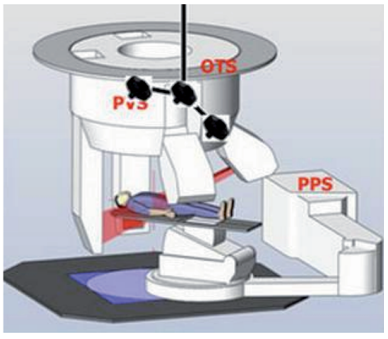


Fig. 8 – Posizionamento di un paziente (a sinistra) con i markers riflettenti. A destra sono mostrate le telecamere e le lampade ad infrarosso (PVS),

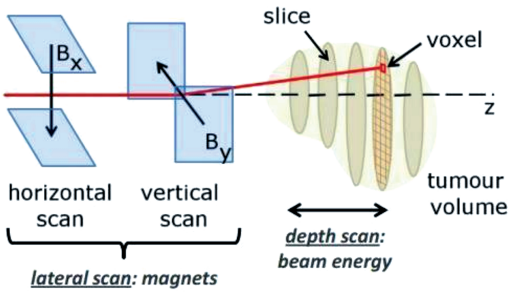


Fig. 9 – Schema di irraggiamento attivo tramite la scansione di una sezione dopo l'altra. Lo spostamento lungo la profondità z si ottiene variando l'energia del fascio.

10. Conclusioni

La radioterapia con adroni è iniziata al CNAO nel 2012 con irraggiamenti sperimentali di pazienti selezionati ed è divenuta parte del Servizio Sanitario Nazionale dal 1° gennaio 2014. Allo stato attuale sono in funzione tutte le tre sale di trattamento, con il Centro in piena operatività terapeutica durante il giorno mentre nella notte proseguono le attività di ricerca scientifica sia tecnologica sia sullo studio degli effetti delle radiazioni sui vari tipi di cellule tumorali con lo scopo di migliorare l'efficacia dell'adroterapia. L'operatività del Centro è pertanto di 24 ore su 24 come avviene in tutti i centri di ricerca dotati di acceleratori, con interruzioni periodiche per manutenzione. In futuro si prevede di sviluppare un fascio destinato ad una sala di ricerca che opererà a livello internazionale. L'attività del Centro viene controllata da un Comitato Scientifico internazionale di cui fanno parte i maggiori esperti nel settore dell'adroterapia provenienti dal Giappone, dagli USA e dalla Germania, tutti Paesi ove sono sorti i principali centri di adroterapia nel mondo.

IL PROFUMO DI PULITO

CLAUDIO FUGANTI (*), MARIA ELISABETTA BRENNA (*)

Nota presentata dal m.e. Claudio Fuganti
(Adunanza del 24 maggio 2012)

SUNTO. – La nascita della profumeria moderna risale a circa un secolo fa, quando la chimica organica sintetica ha cominciato a mettere a disposizione, accanto ai materiali naturali estrattivi, una grande varietà di prodotti: composti isolati puri o in forma arricchita da fonti naturali, prodotti semi-sintetici da materiali naturali, prodotti aromatici di sintesi totale. Per le varie classi di sostanze odorose vengono presentati i risultati più significativi ottenuto negli ultimi anni.

***

ABSTRACT. – Modern perfumery was born nearly a century ago, when synthetic organic chemistry made available a great variety of products, besides those traditionally obtained by extraction from nature: pure or enriched compounds isolated from natural sources, semi-synthetic products prepared from natural starting materials, aromatic compounds obtained by total synthesis. The most relevant results achieved in the last decade are herein presented for several types of odorous substances.

L'uomo ha da tempo immemore considerato importanti le sostanze odorose profumate, come suggerito anche dall'uso di preziosi contenitori (*Fig. 1*) per gli estratti vegetali e animali che con il tempo era riuscito d ottenere. Tuttavia, la moderna profumeria è iniziata circa un secolo fa,

(*) Dipartimento di Chimica del Politecnico di Milano, Italy.
E-mail: mariaelisabetta.brenna@polimi.it

quando la chimica organica sintetica ha cominciato ad aggiungere alla palette del profumiere creativo ai materiali naturali estrattivi (costituiti da estratti con solvente, oli essenziali e materiali resinosi da fiori, frutti, foglie, radici e legno da molte piante: a) prodotti isolati puri o in forma arricchita da fonti naturali; b) prodotti semi-sintetici da materiali naturali e c) prodotti aromatici di sintesi totale) (*Fig. 2*).

Questo è il caso della etilvanillina, mai trovata in Natura. Questo materiale è 10 volte più potente della vanillina ed è utilizzato come **aroma artificiale** soprattutto nei gelati. La prima utilizzazione della etilvanillina in profumeria fu in Jicky (1898) e in Salimar (1925) di Guerlain (*Fig. 3*).

Fra i prodotti semi-sintetici è importante l'idrossi citronellale (*Fig. 4*). Philippe Cruit nel 1907 trattando con acido cloridrico diluito l'olio essenziale di citronella ottenne la conversione del componente principale, citronellale, nell'idrossicitronellale (Cyclosia), attraverso una formale addizione di acqua al doppio legame. Questo prodotto rappresenta la base del profumo della crema Nivea ed è componente importante della nota floreale di Quelques Fleurs (Houbigants) (*Fig. 5*). Nella *Fig. 6* sono indicati prodotti di sintesi analoghi del Cyclosia.

Il profumo più famoso al mondo, Chanel N° 5, è stato creato combinando con un estratto naturale (olio di Ylang-Ylang) con aldeidi sintetiche contenenti da 8 a 12 atomi di carbonio (*Fig. 7*).

Il profumo di violetta, rappresentato dal chetone α-ionone, è stato ottenuto per caso trattando con acido solforico concentrato il chetone ottenuto condensando l'acetone con il citrale naturale. Più di 70 anni dopo la sua sintesi l'α-ionone è stato trovato nell'olio essenziale di viola (*Fig. 8*). Altro prodotto completamente di sintesi, essenziale per i profumi di rose, è l'estere metilico dell'acido 2-ottinoico, usato nella creazione di Fahranheit di Dior (1988) (*Fig. 9*).

Altra classe di prodotti di grande uso in profumeria è quella dei salicilati, strutturalmente correlati all'Aspirina (*Fig. 10*). Questi sono caratterizzanti di prodotti commerciali di grande successo quali Fidji di Guy Laroche (1966), Charlie di Norell e Super Estee di Estee Lauder (*Fig. 11*).

Un punto di rottura nell'evoluzione della profumeria è costituito dall'enorme espansione dell'impiego di polveri detergenti e di saponi la cui produzione mondiale raggiunge circa 26 milioni di ton/anno (*Fig. 12*).

Nessuno dei materiali odorosi sopra nominati è stabile alle condizioni alcaline della maggior parte dei detergenti impiegati comunemen-

te. Per di più, è noto che i tessuti lavati in lavatrice e asciugati in ambiente non fortemente ventilato posseggono un odore definito tecnicamente di 'cane bagnato'. Ci si pose il problema di quale tonalità odorosa potesse essere usata per conferire il cosiddetto '**profumo di pulito**'.

Si desidera che le sostanze utili allo scopo, una volta addizionate alle polveri detergenti, resistano al lavaggio, aderendo alle fibre in modo sufficientemente forte da rimanere anche dopo gli abbondanti risciacqui, non si allontanino quando il tessuto rimane all'aria e, infine, siano rilevabili nell'armadio in cui si ripone il capo lavato. Questi requisiti si ritrovano abbastanza soddisfacentemente in materiali odorosi di origine animale (*Fig. 13*). Analoghi di sintesi del muscone (*Fig. 14*) sono riportati in *Fig. 15*. Paradossalmente, fra i più potenti odoranti muschiati ci sono derivati aromatici simili all'esplosivo tritolo, quali il musk toluene (*Fig. 16*). Questi materiali di sintesi raggiunsero una produzione annua ragguardevole (5.000,00 ton/anno) (*Fig. 17*), ma la loro fototossicità portò al declino e poi all'esclusione dall'impiego. Sostituti quanto mai efficienti dei nitro muschi sono i muschi policiclici, il più importante dei quali è il galaxolide (*Fig. 18*). Ovviamente, la maggior parte degli odorizzanti addizionati alle polveri per lavatrice, viene asportata assieme allo sporco e, alla fine, va a finire nei fiumi. E' stato quindi utile studiate il metabolismo da parte dei pesci di molti odorizzanti. Nel caso del galaxolide, miscela di diastereoisomeri usata come racemo, si è osservato che la carpa degrada in modo più efficiente i due isomeri più odorosi del materiale impiegato (*Fig. 19*). La capacità di rimanere aderenti alla fibra di tessuto dei vari muschi policiclici si ritiene sia legata al numero di gruppi metilici presenti nella struttura, così come la loro resistenza alla biodegradazione. Ciò in analogia con quanto avviene per i componenti della benzina (*Fig. 20*).

L'enantioselettività della percezione odorosa è quanto mai spiccata nell'uomo. Un esempio è quello del Doremox (*Fig. 21*) e del Magnolan (*Fig. 22*). Referenze sulla enantioselettività della percezione odorosa sono in *Fig. 23*.

Fig. 1 – Porta-unguenti di alabastro (Museo Egizio del Cairo).

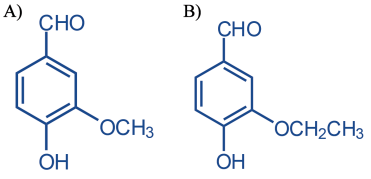


Fig. 2 – A) Vanillina. B) Etilvanillina.

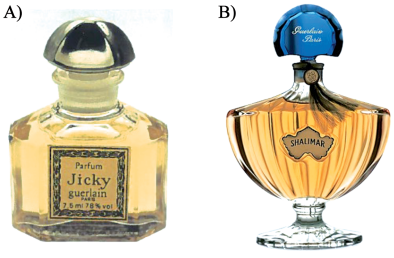


Fig. 3 – A) Jicky (Guerlain, 1898). B) Shalimar (Guerlain, 1925).

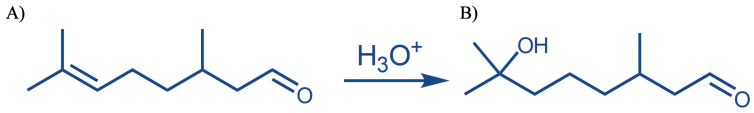
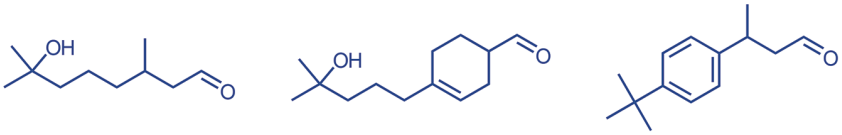


Fig. 4 – A) Citronellale. B) Idrossicitronellale (Cyclosia®).

Fig. 5 – A) Citronellale. B) Idrossicitronellale (Cyclosia®).



Cyclosia (1907)

Lyral (tonalità simile a Cyclosia) (1960)

Lylial (fiori di tiglio)

Fig. 6.

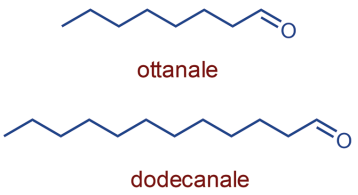
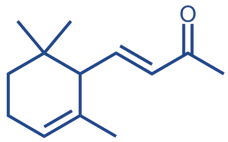


Fig. 7.

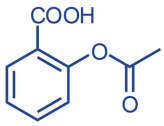


α-ionone (profumo di violetta)

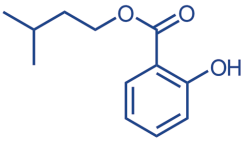
Fig. 8.

metil 2-octinoato

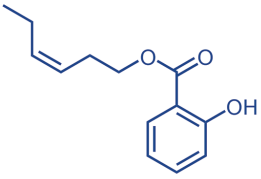
Fig. 9.



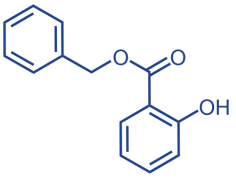
acido acetil salicilico (Aspirina)



salicilato di isoamile



salicilato di cis-3-esenile
(nota verde, di grande potenza)



salicilato di benzile
(gelsomino)

Fig. 10 – Salicilati.

Fidji di *Guy Laroche (1966)*

Charlie di *Norell*

Super Estee di *Estee Lauder*

Fig. 11.

Produzione mondiale recente di alcuni composti organici (ton/a)

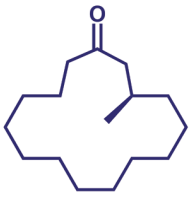
Prodotti petrolchimici	180.000.000,00
Oli e grassi da piante	78.000.000,00
Polveri detergenti, saponi	26.000.000,00
Essenza di trementina	260.000,00
Fragranze (conc.)	200.000,00
Oli essenziali	40.000,00
Oli del genere *Citrus*	80.000,00

Fig. 12.

Moschus moschiferus

Viverra civetta

Fig. 13.

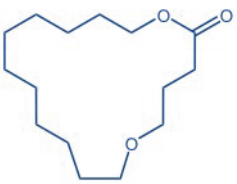


(-)-muscone

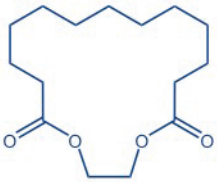
Estratto da ghiandole dei maschi
di *Moschus moschiferus* e *Viverra civetta:*
liquido, temp. eboll.: 328 °C
MOLTO POCO SOLUBILE IN ACQUA

Isolamento:
H. Waldbaum, *Das natuerliche Moschusaroma J.Prakt. Chem.*, **1906**, *73*, 488.
Caratterizzazione strutturale della forma naturale:
L. Ruzicka *Helv. Chim. Acta,* **1926**, *9*, 715.
Sintesi:
Ziegler et al., *Liebigs Ann.* **1934**, *514*, 164.
Ruzicka e Stoll, *Helv. Chim. Acta* **1934**, *17*, 1308.

Fig. 14.

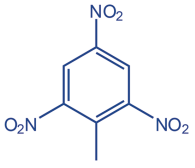


11-oxa-esadecanolide (Musk R-1 [18m])

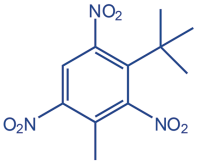


etilene brassilato (Astrotone)
(prodotto di partenza: acido brassilico dalla colza)

Fig. 15.

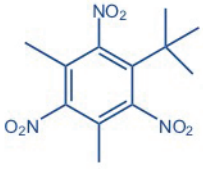


tritolo

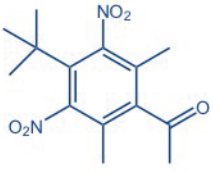


musk toluene

Fig. 16.



musk xilene



musk ketone

Fig. 17.

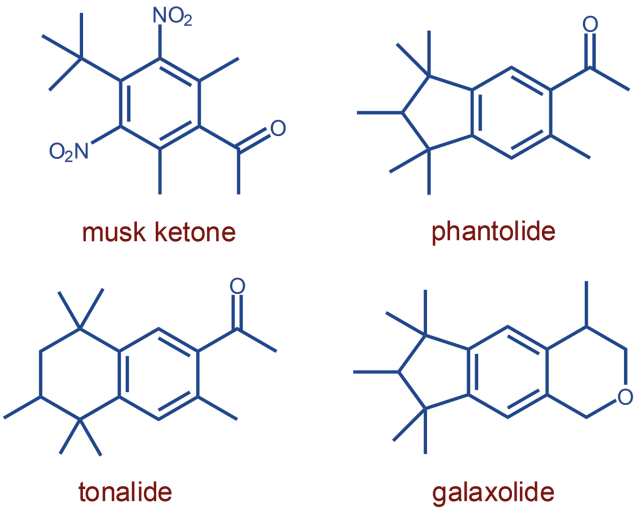

NO2
O2N
O
musk ketone
O
phantolide
O
tonalide
O
galaxolide

Fig. 18.

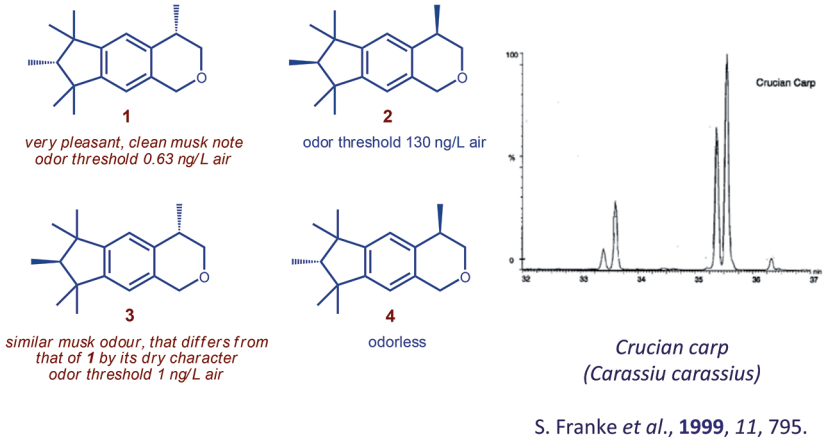
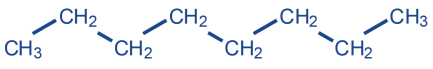
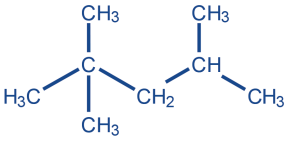

1
very pleasant, clean musk note
odor threshold 0.63 ng/L air
2
odor threshold 130 ng/L air
3
similar musk odour, that differs from
that of 1 by its dry character
odor threshold 1 ng/L air
4
odorless
Crucian Carp
Crucian carp
(Carassiu carassius)
S. Franke et al., 1999, 11, 795.

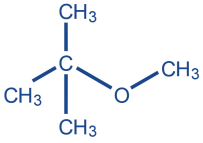
Fig. 19.



ottano



isottano

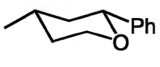


MTBE

Fig. 20.

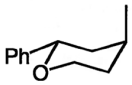
(2*S*,4*R*)-*cis*-**Doremox**

rose oxide, powerful, nice



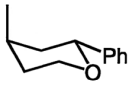
(2*R*,4*S*)-*cis*-**Doremox**

rose oxide, diphenyl oxide, metallic, slightly plastic



(2*S*,4*S*)-*trans*-**Doremox**

weak, rosy, plastic, citronellol, a rose oxide note is also present

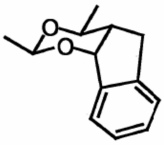


(2*R*,4*R*)-*trans*-**Doremox**

rosy, rose oxide, metallic, off note

E. Brenna, C. Fuganti, S. Ronzani, S. Serra, *Can.J. Chem*, **2002**, *80*, 714.

Fig. 21.



(-)-(1*S*,3*S*,4a*R*,9a*S*)-magnolan

highly appreciated, and found rosy, floral (geranium, magnolia), citronellyl acetate, citric-fruity, with a slight green nuance

Fig. 22.

Pergamon

SCIENCE DIRECT®

Tetrahedron: *Asymmetry* 14 (2003) 1–42

TETRAHEDRON: *ASYMMETRY*

TETRAHEDRON: ASYMMETRY REPORT NUMBER 54

Enantioselective perception of chiral odorants

Elisabetta Brenna,* Claudio Fuganti and Stefano Serra

Dipartimento di Chimica, Materiali ed Ingegneria Chimica del Politecnico, Istituto CNR per la Chimica del Riconoscimento Molecolare, Via Mancinelli 7, I-20131 Milano, Italy

Received 18 October 2002; accepted 28 October 2002

CHEMICAL REVIEWS

REVIEW

pubs.acs.org/CR

Biocatalytic Methods for the Synthesis of Enantioenriched Odor Active Compounds

Elisabetta Brenna,*[,†] Claudio Fuganti,[†] Francesco G. Gatti,[†] and Stefano Serra[‡]

[†]Politecnico di Milano, Dipartimento di Chimica, Materiali, Ingegneria Chimica, Via Mancinelli 7, I-20131 Milano, Italy.
[‡]Istituto di Chimica del Riconoscimento Molecolare—CNR, Via Mancinelli 7, I-20131 Milano, Italy

dx.doi.org/10.1021/cr100289r | *Chem. Rev.* 2011, 111, 4036–4072

Fig. 23.

CARATTERIZZAZIONE MOLECOLARE E FUNZIONALE DI NEURONI STRIATALI DERIVATI DA CELLULE STAMINALI EMBRIONALI UMANE

ALESSIA DELLI CARRI (*)

Nota presentata dal s.c. Elena Cattaneo
(Adunanza del 28 giugno 2012)

SUNTO. – Lo sviluppo di protocolli per il differenziamento neurale finalizzato all'ottenimento in vitro di neuroni funzionali rappresenta il futuro in campo medico e farmacologico. Il contributo illustra il protocollo sviluppato da laboratorio Cattaneo per ottenere in vitro neuroni dello striato da cellule staminali umane embrionali e pluripotenti e pubblicati nel 2013 sulla rivista Development (Delli Carri A., Onorati M., *et al.*; 2013). Questi neuroni muoiono nella Corea di Huntington, una malattia degenerativa del sistema nervoso centrale, che determina la degenerazione dei neuroni di tipo striatale localizzati al livello dei gangli della base. La patologia dal punto di vista clinico è caratterizzata da movimenti involontari, da un deterioramento cognitivo progressivo ed alterazione del comportamento.

***

ABSTRACT. – The development of neural differentiation protocol aimed at obtaining functional neurons *in vitro* represents the future in medicine. The text will explain the protocol developed by Cattaneolab in order to obtain *in vitro* striatal neurons from human embryonic and pluripotent stem cells. The work has been published in 2013 on Development (Delli Carri A., Onorati M., *et al.*; 2013). Striatal neurons at ganglia level die in Huntington's Disease, a neurodegenerative disorder of the central nervous system, characterized by a gradual development of involuntary muscle movements, progressive disorientation and confusion, personality disintegration, impairment of memory control.

(*) Center for Stem Cell Research, Università degli Studi di Milano, 20133 Milan, Italy.

E-mail: alessia.dellicarri@unimi.it; elena.cattaneo@unimi.it

C'è molto interesse nel mondo scientifico per lo sviluppo di protocolli per il differenziamento neurale finalizzato all'ottenimento *in vitro* di neuroni funzionali.

Molti laboratori lavorano e collaborano per il raggiungimento di questo obiettivo comune, che rappresenta il futuro in campo medico e farmacologico.

Le cellule staminali embrionali e, più recentemente, quelle pluripotenti indotte (iPS) sono state utilizzate per la messa a punto di protocolli innovativi di differenziamento neuronale. In particolar modo, il trattamento di alcune patologie neurodegenerative attraverso neuroni ottenuti *in vitro* da cellule dello stesso paziente o da embrionali staminali è un'ipotesi attualmente in considerazione per il Parkinson.

Il lavoro condotto dal nostro laboratorio ha coinvolto diversi gruppi, tanti ricercatori e tanti anni di esperimenti ed è stato possibile grazie alla collaborazione con i colleghi facenti parte del consorzio NeuroStemcell (2008-2013), un network di ricerca su scala europea e con il coinvolgimento anche di partner americani. Queste collaborazioni proseguono ancora oggi nel progetto Neurostemcellrepair (2013-2017).

I risultati ottenuti hanno permesso di pubblicare nel 2013 sulla rivista Development (Delli Carri A., Onorati M., *et al.*; 2013) il protocollo per ottenere *in vitro* neuroni dello striato da cellule staminali umane pluripotenti. Questi neuroni muoiono nella Corea di Huntington, una malattia degenerativa del sistema nervoso centrale, che determina la distruzione dei neuroni di tipo striatale localizzati al livello dei gangli della base. Questi, che rappresentano la popolazione cellulare più numerosa presente nello striato, utilizzano come neurotrasmettitore il GABA, responsabile di segnali di tipo inibitorio.

Clinicamente è caratterizzata da movimenti involontari, da un deterioramento cognitivo progressivo ed alterazione del comportamento, conseguenza della mancanza dell'azione inibitoria del GABA.

Ad oggi non vi sono farmaci in grado di bloccare o rallentare la progressione della malattia, o curarla; le sostanze attualmente disponibili, benché utili, hanno solo un effetto a livello sintomatico.

Grazie a studi eseguiti sullo sviluppo dello striato nel topo si è riusciti a ricapitolare ciò che succede al livello embrionale durante lo sviluppo del sistema nervoso, *in vitro*, partendo da cellule staminali embrionali umane.

Nelle fasi precoci dello sviluppo del cervello, le cellule staminali sono localizzate in una zona che circonda i ventricoli cerebrali. Quelle

che genereranno i neuroni striatali umani presentano un preciso codice molecolare identificativo, che si modifica, raggiungendo una nuova "impronta molecolare", quando si allontanano dalla zona, per popolare lo striato. Un terzo codice identificativo viene acquisito, infine, nel momento in cui le cellule raggiungono la zona dello striato dove risiederanno definitivamente.

Il protocollo messo a punto dal nostro laboratorio e pubblicato nel 2013, permette di mimare tutte queste fasi *in vitro* e, dopo 80 giorni, di ottenere neuroni striatali elettrofisiologicamente funzionali. Prevede tre fasi principali: la prima detta d'induzione, in cui le cellule staminali abbandonano lo stato di pluripotenza e sono indirizzate verso un destino neurale. Questa prima fase dura circa 15 giorni e prevede l'uso di particolari terreni di coltura e sostanze sintetiche neuralizzanti.

La seconda fase è di specificazione, in cui l'uso di morfogeni, particolari molecole con specifici segnali regionalizzanti, permette alle cellule di diventare progenitori neurali. Questi saranno capaci di dare origine ad una ridottissima popolazione neuronale, tra cui quelli striatali. La terza fase consiste nell'impartire a questi progenitori una specifica identità di neuroni striatali, inibendo altre identità, e prevede l'utilizzo di una particolare sostanza neurotrofica, che ne permette il loro differenziamento terminale. Questo protocollo permette di ottenere neuroni maturi in 80 giorni.

I colleghi che lavorano sul Parkinson hanno invece dimostrato che è possibile ottenere neuroni dopaminergici autentici da embrionali umane.

Un grande successo è stato la recente pubblicazione, sulla prestigiosa rivista Stem Cell, del lavoro condotto da Malin Parmar dell'Università di Lund in Svezia, sull'applicazione clinica delle cellule staminali nei pazienti malati di Parkinson. Anche questo laboratorio ha fatto parte del consorzio NeuroStemcell (2008-2013).

Diverse sono le tappe del sensazionale studio svedese. Gli scienziati, inizialmente, hanno ottenuto neuroni che producono dopamina, gli stessi che vanno incontro a degenerazione nei malati di Parkinson, a partire da cellule staminali embrionali umane. Li hanno trapiantati in topolini da laboratorio e, con sorpresa, hanno constatato che erano in grado di connettersi alle altre cellule nervose del tessuto ospite, attraverso un'estesa rete di ramificazioni. Si tratta di un risultato che ha richiesto tanti anni di ricerca e la speranza dei ricercatori è di poterlo affinare ulteriormente, fino a riuscire a produrre le cellule nel rispetto dei parametri necessari per l'utilizzo clinico. Questo rappresenta uno dei potenziali scenari per il trattamento di alcune patologie neurodegenerative.

BIOENGINEERING THE ARTIFICIAL NERVE

Nota del m.s. GIORGIO TERENGHI (*)

(Adunanza del 4 ottobre 2012)

SUNTO. – Attualmente, la riparazione di un nervo periferico a seguito di un danno traumatico, richiede una operazione chirurgicamente complessa e spesso risulta in un recupero non ottimale con conseguente deficit funzionale. La neurobiologia della rigenerazione del nervo non può essere risolta unicamente con le tecniche chirurgiche esistenti. La bioingegneria rappresenta una possibile soluzione a questo problema, con l'applicazione di condotti neuronali che potenzialmente possono migliorare la rigenerazione dei nervi. L'utilizzo di nuovi bio-materiali e di metodi di fabbricazione sempre più raffinati, usando micro- e nano-tecnologie, hanno portato ad un ulteriore perfezionamento dei condotti. Un successivo progresso nella rigenerazione del nervo è l'utilizzo di cellule trapiantate all'interno dei condotti, creando così un nervo bio-artificiale da innestare nel nervo danneggiato. Le cellule staminali adulte sono l'alternativa ideale alle cellule di Schwann, ed in particolare il tessuto adiposo è la fonte più promettente di cellule staminali per la medicina rigenerativa. Le cellule staminali, isolate facilmente da questo tessuto in donatori adulti, si moltiplicano rapidamente e possono essere pertanto differenziate in cellule di Schwann. Per questa ragione, queste cellule rappresentano una nuova strategia nell'applicazione clinica e per il miglioramento nella rigenerazione dei nervi periferici.

***

ABSTRACT. – At present, peripheral nerve injury represents a complex challenge for the surgeons, and the repair of these injuries often results in sub-optimal recovery and functional deficit. The neurobiology of nerve regeneration cannot be adequately addressed by the existing surgical techniques, and it has become apparent that tissue engineering and the creation of nerve conduits have significant potential to improve the results of nerve reconstruction. The use of novel biomaterials and more refined fabrication methods, by using micro- and nano-technology, are a promising development for

(*) Professore di Ingegneria Tissutale Centre for Tissue Injury & Repair University of Manchester.
E-mail: giorgio.terenghi@manchester.ac.uk

these experimental constructs. Also, the use of cells transplanted into the conduit to create a bioartificial nerve graft represents a new development. Adult stem cells constitute the ideal alternative to Schwann cells, and among the various sources of stem cells with potential application for regenerative medicine the adipose tissue has been proven to be the most promising. Adult adipose-derived stem cells are easily obtained, they expand rapidly and can be differentiated to Schwann cells, hence they represent a new strategy for clinical application in order to improve the results of nerve re generation.

1. Introduction

Injuries to peripheral nerves constitute a substantial clinical problem, although the peripheral nervous system possesses an intrinsic capacity to regenerate, which allows the nerve to regenerate spontaneously when there is nerve continuity after damage.[1] However, the functional recovery is often not satisfactory, leaving patients with diability and poor quality of life. This is exacerbated by the delay that may occur between injury and repair, which causes peripheral atrophy in the muscles lacking innervation, and neuronal cells death of those neurons that have lost contact with the periphery.[2]

Current surgical techniques for nerve repair consist in joining the proximal and distal stumps of the damaged nerve. When the nerve gap is nor repairable without creating tension, a nerve autograft is carried out, using a nerve from another area of the body, typically the sural nerve in the leg. The nerve autograft contains all the elements that are required to promote nerve regeneration, but often this approach is not ideal in term of size match for the recipient nerve. Also, autografting involves donor site morbidity such as scarring, sensory loss and formation of neuroma. An alternative solution would be the use of a bioenegineered nerve graft. This approach consists in the creation of a construct composed of a nerve guide, or conduit,[3] enriched with the elements normally presents in a nerve autografts which can improve the outcome of regeneration, such as transplanted cells, extracellular matrix proteins and growth factors.

2. Nerve conduits

A simple biologically inert tube in which the nerve stumps are inserted to bridge the gap would help in directing the regenerating

axons, concentrate the regenerative factors and block scar formation to the nerve. A range of bio-compatible materials have been tested experimentally as empty tubes with various degrees of success.[4,5,6] A direct comparison between different studies is difficult due to the variability in animal models, nerve gap size, surgical technique and outcome measures. However, all have indicated that the entubulation of the nerve is a feasable approach for the repair of a gap injury. Some materials have also been used clinically, although their use is limited to specific applications and not always they produce good results.[7]

More recent studies have developed modification of the internal structure of the conduit, for example adding internal fibres or porous structures to enhance directional axonal growth,[8,9,10,11] but further work will be needed before full clinical application will be possible.

An alternative way to making the conduits bioactive is the incroporation of elements that partecipate actively to the regeneration process, such as the targeted delivery of exogenous neurotrophic factors,[12] as they have a great impact upon the initiation of the regeneration process. Despite experimental evidence that isolated factors improve regeneration following their incorporation into nerve constructs, they are not effective for a clinical application since each neuronal sub-population exhibits specificity for individual neurotrophic factor.[13] Hence a cocktail of growth factors would be necessary for optimal effect, but this mix would need to be changed for each type of injured nerve.

Cell transplanted within a conduit rapresent a further alternative. Schwann cells are the obvious choice given their active role in peripheral nerve injury and regeneration,[14] and several experiments have demonostrated the improvement obtained by adding these cells to a conduit.

3. Transplanted cells

Autologous Schwann cells are the favourable choice for seeding in a nerve conduit, as they are normally present in the autologous nerve graft, and they drive the regeneration process by secreting growth factors and producing extracellular matrix proteins. They also form structures called bands of Büngner, which are essential to guide the regenerating axons towards the target organs.[15] Autologous Schwann cells,

both human and animal [16,17] have been successfully cultured and have been shown experimentally to enhance regeneration across nerve gaps.[18] However, the practicalities for their use are not ideal, as it takes several weeks to culture sufficient autologous cells to fill even a short conduits bridging a nerve gap, and this delay would lead to increase neuronal death and nerve atrophy.[2]

Allogeneic Schwann cells could be cultured and stored for rapid use of nerve injury repair, but they are rejected too quickly to give a significant improvement to the regeneration process,[19] unless immunosuppression is carried out. Immunosuppression is not detrimental to nerve regeneration [20] and it has been tested clinically.[21] However, ethical issue relating to transplantation and, more importantly, the potential of infection to the patient have precluded a widespread use of this approach.

For all these reasons, the research has moved towards the use of alternative cells, in particular stem cells, which presents all the advantages of an ideal candidate cell for regenerative medicine.

4. Stem cells

Stem cells are undifferentiated precursors that can divide into cells with identical potential, in a process called self-renewal. Also they can differentiate to a variety of cell lineages.[22] Embryonic stem cells are defined as pluripotent, as they can generate all the three germ layers of the embryo, *i.e.* ectoderm, mesoderm and endoderm.[22] Although they have shown great potential for regenerative medicine, embryonic stem cells present moral and ethical issues, the risk of solid tumour formation and possible differentiation into undesirable cells.[22,23] Hence their clinical use has been somehow limited and controversial.

Most adult organs contain small populations of adult stem cells, originally considered unipotent as generating only one type of differentiated cells. Their role is primarily to repair and replace the functional cells of the tissue to which they belong in case of damage or injury.[24] New findings on cell plasticity have shown that adult stem cells can trans-differentiate into cells of a different lineage under specific stimulation.[24,25] Adult stem cells are distributed throughout the body, including bone marrow, fat, skeletal muscle, liver and skin.[25] Bone marrow derived stem cells (MSC) were the first cell population to be

characterised and used experimentally and clinically, as they can differentiate *in vitro* into mesenchymal cells leading to the formation of bone, cartilage and fat.[25,26] Under specific stimulation, MSC can also differentiate into Schwann cell-like cells, with similar molecular markers, morphology and secretion of growth factors.[27,28] Experimentally, MSC have also been shown to promote nerve regeneration after injury.[24,29]

The harvest of MSC from bone marrow is a highly invasive procedure, with pain at the harvest site and with limited volume of harveste fluid and cells.[30] However, humans have abundant subcutaneous fat deposits containing stem cells, defined as adipose-derive stem cells, which can be easily isolated by conventional liposuction procedures avoiding the problems related to the harvest of MSC.[31]

5. Adipose-derived stem cells

Adipose-derived stem cells (ASC) present similar characteristic to MSC, such as self-renewal capacity and ability to differentiate into multiple lineage.[32] ASC have the advantage to be more numerous in the processed lipoaspirate compared to the MSC in the bone marrow,[33] and given the amount of fat that can be harvested subcutaneously ASC represent a more attractive source of stem cell for regenerative medicine.

Different groups have also demonostrated that ASC can be differentiated *in vitro* into Schwann cell-like cells, similarly to MSC.[31,34] Interestingly, depending on the anatomical site of harvest, ASC have shown different properties in culture models. Human ASC obtained from superficial fat deposits show higher yields, faster proliferation and increased promotion of neurite outgrowth when compared to ASC obtained from deeper layer.[35] Also, the harvest site and donor age influence the growth rate, differentiation and neurotrophic factor secretion of mouse and rat ASC.[36,37,38] These studies indicate that more has to be done in order to characterise and standardise ASC for a possible clinical use.

Experimental studies have also confirmed the ability of differentiated ASC to promote nerve regeneration, when the cell are seeded into fibrin conduit bridging a short gap in the rat sciatic nerve.[39, 40] Good regeneration results were also obtained in experiment using poly-

mer conduits [41] seeded with differentiated ASC, showing increased function index and prevention of neuronal cell death respectively.

6. Conclusions

There are increase evidence that ASC may represent a potential tool for the treatment of nerve injury. In a clinical scenario, fat could be harvested from the patients suffering from nerve injury to extract and culture ASC. These can then be differentiated to Schwann cell-like before seeding them in a nerve conduit, to be grafted at the site of injury creating an "artificial nerve" substituting the presently used nerve autograft.

Considerable technical progress has been made and good experimental results have defined the elements required for nerve repair and regeneration. However, many factors are still to be clarified before this can be achieved, such as shortening the presently time of culture and differentiation of the ASC, or alternatively finding a way to differentiate these cells within the conduit after being seeded as undifferentiated stem cells. Emerging technologies may also have a further impact on the design of the nerve conduit, including internal nanatotechnology design to modulate cell behaviour or even differentiation. Although clinical translation of nerve conduit has already began using inert constructs, the results are still below optimal and it will be only when a bioactive conduit will be available to obtain the best nerve regeneration results.

REFERENCES

[1] Chen ZL, Yu WM, Strickland S. Peripheral nerve regeneration, *Annual Review of Neuroscience*, 2007: 30: 209-233.

[2] Hart AM, Terenghi G, Wiberg M. Neuronal death after peripheral nerve injury and experimental strategies for neuroprotection, *Neurological Research*, 2008: 30: 999-1011.

[3] De Ruiter GC, Spinner RJ, Yaszemski MJ, Windebank AJ, Malessy MJ. Nerve tubes for peripheral nerve repair. *Neurosurgery Clinical North America,* 2009: 20: 91-105.

[4] Chiu DT, Janecka I, Krizek TJ, Wolff M, Lovelace RE. Autogenous vein graft as a conduit for nerve regeneration, *Surgery*, 1982: 91: 226-233.

[5] Ijpma FF, Van De Graaf RC, Meek MF. The early history of tubulation in nerve repair, *Journal of Hand Surgery Europe*, 2008: 33: 581-586.

[6] Lundborg G. Regeneration and experimental nerve repair, in: *Nerve injury and repair*, Lundborg G editor, p. 114-167, Elsevier Churchill Livingstone, Philadelphia, 2004.

[7] Meek MF, Coert JH. US Food and Drug Administration / Conformity Europe – approved absorbable nerve conduits for clinical repair of peripheral and cranial nerve, *Annual of Plastic Surgery*, 2008: 60: 466-472.

[8] Allmeling C, Jokuszies A, Reimers K, Kall S, Vogt PM. Use of spider silk fibres as an innovative material in a biocompatible artificial nerve conduit, *Journal of Cell Molecular Medicine*, 2006: 10: 770-777.

[9] Chew SY, Mi R, Hoke A, Leong KW. The effect of allignment of electrospun fibrous scaffold on Schwann cell maturation, *Biomaterials*, 2008: 29: 653-661.

[10] Ngo TT, Waggoner PJ, Romero AA, Nelson KD, Eberhart RC, Smith GM. Poly(L-lactide) microfilaments enhance peripheral nerve regeneration across extended nerve lesions, *Journal of Neuroscience Research*, 2003: 72: 227-238.

[11] Toba T, Nakamura T, Shimizu Y, Matsumoto K, Ohnishi K, Fukuda S. Regeneration of canine peroneal nerve with the use of polyglycolic acid-collagen tube filled with laminin-soaked collagen sponge: a comparative study of collagen sponge and collagen fibres as filling material for nerve conduits, *Journal of Biomedical Material Research*, 2001: 58: 622-630.

[12] Yin Q, Kemp GJ, Frostick SP. Neurotrophins, neurones and peripheral nerve regeneration, *Journal of Hand Surgery*, 1998: 23B: 433-436.

[13] Terenghi G. Peripheral nerve regeneration and neurotrophic factors, *Journal of Anatomy*, 1999: 194: 1-14.

[14] Gordon T. The role of neurotrophic factors in nerve regeneration, *Neurosurgery Focus*, 2009: 26: E3.

[15] Ide C. Peripheral nerve regeneration, *Neuroscience Research,* 1996: 25: 101-121.

[16] Levi AD, Bunge RP. Studies of myelin formation after transplantation of human

Schwann cells into a severe combined immunodeficient mousse, *Experimental Neurology*, 1994: 130: 41-52.
[17] Levi AD, Guenard V, Aebischer P, Bunge RP. The functional characteristics of Schwann cells cultured from human peripheral nerve after transplantation into a gap within the rat sciatic nerve, *Journal of Neuroscience*, 1994: 14: 1309-1319.
[18] Mosahebi A, Woodward B, Wiberg M, Martin R, Terenghi G. Retroviral labelling of Schwann cells: *in vitro* characterization and *in vivo* transplantation to improve peripheral nerve regeneration, *Glia*, 2001: 34: 8-17.
[19] Mosahebi A, Fuller P, Wiberg M, Terenghi G. Effect of allogeneic Schwann cell transplantation on peripheral nerve regeneration, *Experimental Neurology*, 2002: 173: 213-223.
[20] Snyder AK, Fox IK, Nichols CM, Rickman SR, Hunter DA, Tung TH. Neuroregenerative effects of pre-injury FK506 administration, *Plastic and Reconstructive Surgery*, 2006: 118: 360-367.
[21] Mackinnon SE, Doolabh VB, Novak CB, Trulock EP. Clinical outcome following nerve allograft transplantation, *Plastic and Reconstructive Surgery*, 2001: 107: 1419-1429.
[22] Bajada S, Mazakova I, Richardson JB, Ashammakhi N. Updates on stem cells and their applications in regenerative medicine, *Journal of Tissue Engineering and Regenerative Medicine*, 2008: 2: 169-183.
[23] Choumerianou DM, Dimitriou H, Kalmanti M. Stem cells: promises *versus* limitations, *Tissue Engineering, Part B, Reviews,* 2008: 14: 53-60.
[24] Tohill M, Terenghi G. Stem cell plasticity and therapy for injuries of the peripheral nervous system, *Biotechnology and Applied Biochemistry*, 2004: 40: 17-24.
[25] Raff M, Adult stem cell plasticity: fact or artifact?, *Annual Review of Cell and Developmental Biology*, 2003: 19: 1-22.
[26] Tropel P, Noel D, Platet N, Legrand P, Benbid AL, Berger F. Isolation and characterisation of mesenchymal stem cells from adult mouse bone marrow, *Experimental Cell Research*, 2004: 295: 395-406.
[27] Caddick J, Kingham PJ, Gardiner NJ, Wiberg M, Terenghi G. Phenotypic and functional characteristics of mesenchymal stem cells differentiated along a Schwann cell lineage, *Glia*, 2006: 54: 840-849.
[28] Mahay D, Terenghi G, Shawcross SG. Growth factors in mesenchymal stem cells following glial cell differentiation, *Biotechnology and Applied Biochemistry*, 2008: 51: 167-176.
[29] Dezawa M, Takahashi I, Esaki M, Takano M, Sawada H. Sciatic nerve regeneration in rats induced by transplantation of *in vitro* differentiated bone-marrow stromal cells, *The European Journal of Neuroscience*, 2001: 14: 1771-1776.
[30] Locke M, Windsor J, Dunbar PR. Human adipose-derived stem cells: isolation, characterisation and application in surgery, *ANZ Journal of Surgery*, 2009: 79: 235-244.
[31] Kingham PJ, Kalbermatten DF, Mahay D, Armstrong SJ, Wiberg M, Terenghi G. Adipose-derived stem cells differentiate into a Schwann cell phenotype and promote neurite outgrowth *in vitro*, *Experimental Neurology,* 2007: 207: 267-274.

[32] Zuk PA, Zhu M, Ashjian P, De Ugarte DA, Huang JI, Mizuno H. Human adipose tissue is a source of multipotent stem cells, *Molecular Biology of the Cell*, 2002: 13: 4279-4295.

[33] Strem BM, Hicok KC, Zhu M, Wulur I, Alfonso Z, Schreiber RE. Multipotent differentiation of adipose tissue-derived stem cells, *The Keio Journal of Medicine*, 2005: 54: 132-141.

[34] Jiang L, Zhu JK, Liu XL, Xiang P, Hu J, Yu WH. Differentiation of rat adipose tissue-derived stem cells into Schwann cell-like cells *in vitro, Neuroreport*, 2008: 19: 1015-1019.

[35] Kalbermatten DF, Schaakxs D, Kingham PJ, Wiberg M. Neurotrophic activity of human adipose stem cells isolated from deep and superficial layers of abdominal fat, *Cell and Tissue Research*, 2011: 344: 251-260.

[36] Kaewkhaw R, Scutt AM, Haycock JW. Anatomical site influences the differentiation of adipose-derived stem cells for Schwann cell phenotype and function, *Glia,* 2011: 59: 734-749.

[37] Mantovani C, Raimondo S, Haneef MS, Geuna S, Terenghi G, Shawcross SG. Morphological, molecular and functional differences of adult bone marrow and adipose derived stem cells isolated from rats of different ages, *Experimental Cell Research*, 2012: 318: 2034-2048.

[38] Engels PE, Tremp M, Kingham PJ, di Summa PG, Largo RD, Schaefer DJ. Harvest site influences the growth properties of adipose derived stem cells, *Cytotechnology*, 2013: 65: 437-445.

[39] Di Summa PG, Kingham PJ, Raffoul W, Wiberg M, Terenghi G, Kalbermatten DF. Adipose-derived stem cells enhance peripheral nerve regeneration, *Journal of Plastic, Reconstructive and Aesthetic Surgery*, 2010: 63: 1544-1552.

[40] Di Summa PG, Kalbermatten DF, Pralong E, Raffoul W, Kingham PJ, Terenghi G. Long-term *in vivo* regeneration of peripheral nerve through bioengineered nerve grafts, *Neuroscience*, 2011: 181: 278-291.

[41] Reid AJ, Sun M, Wiberg M, Downes S, Terenghi G, Kingham PJ. Nerve repair with adipose-derived stem cells protects dorsal root ganglia neurons from apoptosis, *Neuroscience*, 2011: 199: 515-522.

 Istituto Lombardo (Rend. Scienze) 146, 199-213 (2012)

ESTENSIONE DELLA TEORIA QUANTISTICA DI CAMPO A TEORIA DI AUTOMA CELLULARE QUANTISTICO

Nota del m.e. GIACOMO MAURO D'ARIANO (*)

(Adunanza del 4 ottobre 2012)

SUNTO. In questa nota riferisco brevemente sulla nuova teoria di campo basata sul paradigma informazionale. Il programma prosegue quello della derivazione assiomatica della teoria quantistica da principi di teoria dell'informazione, programma del quale presentai una memoria il 31 gennaio 2008. Il passaggio dalla teoria quantistica di sistemi generici alla teoria di campo si ottiene aggiungendo agli assiomi della teoria quantistica nuovi principi generali per la rete di interazioni di un insieme numerabile di sistemi quantistici. Tali principi si possono sintetizzare nella richiesta di minima complessità dell'algoritmo quantistico che descrive la legge fisica. Da principi generali quali omogeneità, isotropia, località, linearità, unitarietà della rete di interazioni, nonché dimensione minima del vettore di campo, si ricava la teoria di campo libera, ovvero i campi di Weyl, Dirac, e Maxwell. Ne deriva una descrizione del campo in termini di automa cellulare quantistico sul grafo di Cayley di un gruppo.

I vantaggi della nuova teoria sono molteplici. Ad esempio, vengono tautologicamente a mancare tutti i problemi che derivano dal continuo, quali le divergenze ultraviolette, il problema della localizzazione, e le violazioni della causalità. La meccanica stessa non viene assunta, bensì consegue dai principi, e la teoria è *quantistica ab initio*, ovvero non necessita della quantizzazione di una teoria di campo classica. La covarianza di Lorentz non è assunta, bensí consegue essa stessa nel limite relativistico di piccoli vettori d'onda, mentre la nozione di sistema di riferimento è riformulata in termini di rappresentazione irriducibile dell'automa quantistico. Infine, la nuova teoria, derivata da principi puramente matematici–e quindi con variabili adimensionali–contiene essa stessa al suo interno le unità di misura per distanza, massa, e tempo, e le unità sono date in termini di valori estremali delle variabili, in principio misurabili sperimentalmente. Dalla teoria ad automa emergono anche effetti con sapore relativistico-generale, quali un limite superiore alla massa a riposo della particella di Dirac (discendente dall'unitarietà

(*) Dipartimento di Fisica, Via Bassi 6, 27100 Pavia, Italy.
E-mail: dariano@unipv.it

dell'automa), in corrispondenza del quale la relazione di dispersione diviene completamente piatta, in stretta analogia con la nozione di mini buco nero.

***

ABSTRACT. In the present note I will briefly report recent results on the novel field theory based on the information-theoretical paradigm. This program continues the previous one on the axiomatic derivation of quantum theory from information-theoretical principles, of which I presented a note on January 31st 2008. The passage from the quantum theory of abstract systems to the theory of fields is achieved by adding to the axioms of quantum theory new general principles for the network of interation among a denumerable set of quantum systems. Such principles can be synthesized with the requirement of minimal complexity of the quantum algorithm describing the physical law. From general principles such as homogeneity, isotropy, locality, linearity, and unitarity of the interaction network, along with minimal dimension of the field vector, one derives the free quantum field theory, namely Weyl, Dirac, and Maxwell fields. The principles lead to a description in terms of a quantum cellular automaton on the Cayley graph of a group. The advantages of the new theory are numerous. For example the theory tautologically solves all the problems originating from the continuum, such as the ultraviolet divergencies, the localization problem, and causality violations. The mechanics itself is not assumed, but it is a consequence of the principles, and the theory is quantum *ab initio*, namely it does not need quantization of a classical field theory. Lorentz covariance is not assumed, but follows itself from the principles, in the relativistic limit of small wavevectors. The notion of reference system is restated in terms of irreducible representation of the quantum automaton. Finally, the new theory, derived from purely mathematical principles–whence with adimensional variables–contains itself the units of measure for distance, time, and mass, which are given in terms of extremal values of the variables, which are in principle experimentally detectable. Also effects with GR flavor emerge from the automaton theory, such as an upper bound for the Dirac particle rest-mass (from the unitariety condition), where the dispersion relation becomes completely flat, in strict analogy with the notion of mini black hole.

1. INTRODUZIONE

Nell'adunanza del 31 gennaio 2008 presentai una memoria [1] su un mio lavoro in corso, iniziato nell'agosto 2003, sulla fondazione assiomatica della Teoria Quantistica–la struttura teorica portante di tutta la fisica contemporanea. Il programma di assiomatizzazione ha raggiunto il risultato desiderato con due lunghi lavori pubblicati su Physical Review A [2, 3] in collaborazione con i miei ex allievi Paolo Perinotti e Giulio Chiribella[1]. Nel lavoro [3], la Teoria Quantistica è derivata da assiomi di

[1] Il lavoro [3] è stato selezionato dall'American Physical Society (APS) per il *viewpoint* di Časlav Brukner [4] (l'APS seleziona per un *viewpoint* sulla Physical Review circa un

natura puramente informazionale. In sintesi, la Teoria Quantistica è una vera e propria teoria dell'informazione, e gli assiomi (dotati di interpretazione fisica a differenza degli assiomi Hilbertiani) sono principi generali di epistemologia, e riguardano la falsificabilità di una qualsiasi proposizione della teoria nella situazione di completo controllo sperimentale. La nuova formulazione assiomatica permette anche di dimostrare la maggior parte dei risultati chiave della teoria quantistica senza utilizzare spazi di Hilbert e operatori, conducendo ad un nuovo modo di comprendere la teoria.

Nella presente memoria, desidero invece parlare del passo successivo alla deduzione assiomatica della teoria quantistica, ovvero la riformulazione della parte meccanica della teoria. Ciò si ottiene derivando la teoria quantistica dei campi a partire da principi puramente informazionali. Questo programma realizza il paradigma informazionale profetizzato da John Archibald Wheeler e sintetizzato dalla frase del suo slogan: *It from Bit*–un mondo fisico fatto di bits. Con il senno di poi, oggi diciamo: un mondo fatto di *qubits*, ovvero "bit quantistici". Un mondo fatto da puro software, *software senza hardware: Fisica come Informazione*. Cosa vuol dire?

Si deve immaginare che non esistono atomi, ne' particelle, ne' fotoni, ma che questi sono *patterns* risultanti da una computazione su un calcolatore infinito quantistico[2], come se la realtà fisica fosse un'enorme simulazione, ovvero tutto ciò che è fisico–inclusi spazio, tempo, massa, energia, gravitazione–emerge dalla computazione.

Le domanda più frequente che mi viene posta è: da che tipo di hardware è supportato il software se è la materia stessa che viene simulata? Di cosa sono fatti i qubits? Questa domanda non ha senso, e sorge perchè siamo abituati a pensare tutto in termini di materia o di energia. Ma il noto teorema di Malament [6] della teoria algebrica di campo dimostra che il concetto di particella è logicamente insostenibile: l'alternativa è la cosiddetta *ontologia del campo* [7]. Non dobbiamo dimenticare che il modo in cui immaginiamo la realtà deriva dalla teoria

lavoro su 2000). È in corso di pubblicazione un testo della Cambridge University Press, che sarà presentato prossimamente a questa Accademia.

[2]Del computer quantistico parlai in Accademia la prima volta il 18 novembre del 1998 nell'ambito di un ciclo di conferenze su *Scienza e Tecnologie Avanzate*, e suggerisco al lettore il relativo estratto per un breve presentazione divulgativa [5].

che ne abbiamo, e fra le varie teorie non falsificate dall'esperimento la scelta va alla teoria che deriva le osservazioni sperimentali dal minor numero di assunzioni. Il paradigma informazionale di Wheeler promette di derivare l'intera fisica dalla sola teoria quantistica, con l'aggiunta di poche semplici regole generali di semplicità dell'architettura del software. E queste nuove regole, insieme a quelle della teoria quantistica, risultano essere tutti principi metodologici alla base dello stesso approccio scientifico. Il vantaggio notevole è la possibilità di derivare la relatività stessa dalla teoria quantistica, aprendo la strada maestra alla riconciliazione fra le due teorie–la soluzione del problema più rilevante della fisica teorica contemporanea, nonchè chiave di volta della teoria della gravità.

Sostituire un mondo fatto di particelle che viaggiano in uno spazio-tempo Minkowskiano con un cosmo evanescente fatto di pura informazione è indiscutibilmente un cambiamento di ontologia notevole, e di primo acchito il fisico teorico potrebbe essere riluttante ad adottare questo nuovo punto di vista. Ma le nozioni di spazio-tempo e di particella sono ontologicamente inconsistenti se guardate con gli occhi della fisica teorica contemporanea: la particella, oltrechè origine della schizofrenia ontologica del dualismo onda-corpuscolo, è solo uno stato "quantistico del campo", ovvero una mera regola di probabilità–una entità soggettiva nell'interpretazione della probabilità secondo Bayes e de Finetti.

Nel breve spazio di questa memoria riassumerò i risultati recenti della nuova teoria, illustrando come la teoria libera quantistica relativistica di campo emerga dalla computazione in termini di equazioni descriventi il flusso libero dell'informazione sulla rete computazionale e, insieme ad essa, emerga anche lo spazio-tempo e l'invarianza relativistica. Alcuni risultati preliminari su una derivazione euristica dell'equazione di Dirac nel paradigma informazionale sono stati presentati nella memoria del 16 Dicembre 2010 [8]. Per esposizioni divulgate si vedano anche i lavori [9, 10, 11, 12].[3]

[3] I lavori [11, 12] sono derivati da due essays premiati dal Foundational Question Institute (FQXi).

2. I PRINCIPI GENERALI DELLA NUOVA TEORIA DI CAMPO

La teoria quantistica è una teoria generale applicabile ad ogni tipo di sistema. Nella assiomatizzazione informazionale [3] la teoria è una *teoria operazionale*, ovvero una teoria probabilistica dotata di regole di connettività fra gli eventi. Matematicamente ciò corrisponde ad avere le probabilità congiunte dipendenti parametricamente da un circuito che descrive relazioni di input-output tra gli eventi. La nozione di evento viene "vestita" con i "sistemi" che rappresentano le connessioni fra gli eventi, e questi si connettono fra loro in circuiti aciclici chiusi. Matematicamente le regole di connessione sono formalizzate dalla *categoria monoidale* in teoria delle categorie.[4] La teoria quantistica, in quanto teoria operazionale, è quindi un'estensione della teoria delle probabilità, la quale è a sua volta un estensione della logica [13, 14]: in tal senso il programma di assiomatizzazione informazionale è la naturale continuazione del programma della *quantum logic* di von Neumann [15].

Nel framework operazionale si definiscono i seguenti assiomi per la teoria quantistica [3]:

1. Causalità
2. Discriminabilità locale
3. Discriminabilità perfetta
4. Atomicità della composizione
5. Compressione ideale
6. Purificazione

Per una esposizione divulgata rimando alla memoria [1], mentre per la formulazione matematica dei principi e la derivazione tecnica della teoria quantistica si veda il lavoro [3].

La teoria di campo libera in spazio piatto viene derivata aggiungendo gli assiomi seguenti, riguardanti la rete di interazione di un insieme numerabile di sistemi [16]:

[4]Per un esposizione dettagliata si veda [3].

7. Minimalità della dimensione s del sistema
8. Linearità
9. Unitarietà
10. Località
11. Omogeneità
12. Embeddabilità quasi-isometrica in spazio Euclideo
13. Isotropia

Come vedremo, in accordo all'assioma 7 di minimalità dimensionale i sistemi hanno dimensionalità è $s = 2$.[5] L'assioma 8 di linearità corrisponde a scrivere l'evoluzione del vettore di stato del qubit in termini di combinazione lineare dei sistemi che con esso interagiscono mediante matrici di transizione in $\mathbb{M}_s(\mathbb{C})$. L'assioma 9 di unitarietà si traduce in un insieme di equazioni che le matrici di transizione devono soddisfare. L'assioma 10 di località impone che l'insieme dei sistemi che interagiscono con un dato sistema (ovvero connessi da matrici di transizione non nulle) è uniformemente limitato. L'assioma 11 di omogeneità stabilisce che tale numero di sistemi interagenti è uguale per tutti i sistemi, che l'insieme delle matrici di transizione è lo stesso per tutti i sistemi, ed infine che tutte le loop di catene di sistemi interagenti sono le stesse a partire da ogni sistema. Degli assiomi 12 e 13 parlerò nel seguito.

Dagli assiomi 10 e 11 discende che le connessioni fra i sistemi interagenti sono descritti dal grafo di Cayley di un gruppo G, ovvero un grafo direzionato e con links colorati, dove i siti rappresentano elementi del gruppo G, i links sono in corrispondenza con coppie di elementi, il colore del link è in corrispondenza con un generatore del gruppo, e la freccia sul link rappresenta l'operazione del generatore del gruppo sull'elemento alla coda della freccia con risultato l'elemento alla punta della freccia. Quindi il grafo delle interazioni è in corrispondenza con una particolare presentazione $G = \langle h_1, h_2, \ldots . | r_1, r_2, \ldots \rangle$ del gruppo G

[5] Nella teoria interagente o se si considera l'equivalente di più particelle, il qubit diviene un campo Fermionico.

mediante generatori $h_1, h_2, \ldots$, e relatori $r_1, r_2, \ldots$. Ad ogni generatore h è associata una matrice di transizione in $\mathbb{M}_2(\mathbb{C})$.

Si dimostra che gli assiomi 8 e 9 danno luogo a un operatore unitario sullo spazio di Hilbert $\ell^2(G) \otimes \mathbb{C}^s$ di singolo sistema quantistico tensorizzato lo spazio della rappresentazione *right-regular* di quadrato sommabile del gruppo G. La condizione di unitarietà si traduce in un insieme non-banale di condizioni bilineari sulle matrici di transizione.

La Geometric Group Theory. La suddetta costruzione matematica, che segue direttamente dagli assiomi 7-11 collega la nuova teoria di campo ad una branca della matematica pura contemporanea: la Geometric Group Theory. L'idea alla base, dovuta Mikhail Gromov, è di studiare le proprietà del gruppo assegnato dalla presentazione $G = \langle l_1, l_2, \ldots . | r_1, r_2, \ldots \rangle$ in termini delle proprietà dello spazio metrico $M(G)$ nella quale il grafo di Cayley si può embeddare quasi-isometricamente,[6] conducendo ad una geometrizzazione del problema gruppale. Ad esempio, restringendoci a gruppi infiniti che rappresentano il caso di interesse, un teorema dovuto a Gromov dimostra che lo spazio metrico $M(G)$ in cui si embedda il gruppo è Euclideo se e solo se il gruppo G è virtualmente Abeliano,[7] altrimenti G è non virtualmente Abeliano e $M(G)$ è iperbolico. La logica è semplicemente il confronto delle crescite dei volumi della palla[8] di raggio r in funzione di r, rispettivamente per $M(G)$ e G, e tale crescita è polinomiale se $M(G)$ è Euclideo e G è virtualmente Abeliano, mentre è esponenziale se $M(G)$ è iperbolico e G è infinito e non virtualmente Abeliano.

Il lettore si chiederà cosa centra tutto ciò con la teoria di campo: la risposta è che lo spazio metrico $M(G)$ nella nuova teoria risulta essere lo spazio fisico emergente dalle interazioni fra i sistemi. Il postulato

[6] La metrica sul grafo è la *counting-metric* intera, mentre quella sullo spazio metrico è reale. Il termine *embedding quasi isometrico* vuol dire che la discrepanza fra le due metriche è uniformemente limitata, ad esempio è data dal fattore $\sqrt{d}$, dove d è la dimensione dello spazio metrico. Inoltre l'embedding deve anche essere *quasi-onto*, ovvero ogni punto dello spazio metrico dista da un sito del grafo non più di una costante.

[7] Un gruppo si dice virtualmente Abeliano se contiene un sottogruppo Abeliano di indice finito.

[8] Il volume della palla di raggio r nel grafo è data dal numero di punti che distano non più di r steps da un punto qualunque del grafo.

11 restringe la teoria a quella su spazi Euclidei. Tale restrizione è fisicamente motivata dal fatto che nel caso iperbolico l'informazione trasmessa decresce esponenzialmente con la distanza (il caso di gruppo iperbolico sarà comunque analizzato in futuro).

Il postulato dell'isotropia. Resta infine il postulato di isotropia. Innanzitutto si richiede che nell'embedding quasi-isometrico di G in $M(G)$ la sfera unitaria sul grafo sia embeddata nella sfera unitaria in $M(G)$. Viene richiesta quindi l'esistenza di un gruppo P di permutazioni dei generatori di G transitivo che agisce come automorfismo del grafo di Cayley, ed avente una rappresentazione (generalmente proiettiva) unitaria in dimensione s. Si richiede quindi che l'evoluzione unitaria sia invariante sotto l'azione congiunta del gruppo P sul grafo di Cayley e sullo spazio s-dimensionale dei vettori di stato del singolo sistema.

3. La derivazione della teoria quantistica di campo libera nel limite relativistico

Nel lavoro [16] con Paolo Perinotti ci restringiamo al caso di gruppo G Abeliano, anziché al più generale caso virtualmente Abeliano.[9] In questo lavoro dimostriamo che l'unico grafo di Cayley embeddabile quasi-isometricamente in $\mathbb{R}^3$ che permette di soddisfare i principi 7-13 è il reticolo BCC (body centered cubic). Nessuno degli altri reticoli permette di soddisfare le condizioni di unitarietà per dimensione $s = 2$ dello spazio di Hilbert di singolo sistema, mentre per dimensione $s = 1$ (equivalente a campo scalare) sussiste solo il caso di evoluzione banale. Inoltre, esistono solo due possibili trasformazioni unitarie (a meno di cambi di rappresentazione locali e di particolari traslazioni nello spazio dei vettori d'onda), ovvero esistono due automi cellulari quantistici che soddisfano i principi, e i due automi sono connessi fra loro dalla simmetria di parità. Si noti che il concetto di vettor d'onda è ben definito, essendo il gruppo con Cayley graph BCC Abeliano. Nel cosiddetto limite relativistico di vettori d'onda piccoli (ovvero vicino al centro della

[9] Da studi tuttora in corso sembra che la restrizione al caso Abeliano non sia rilevante per quanto segue nel testo. L'utilizzazione dei gruppi virtualmente Abeliani riduce l'analisi al caso scalare per un sottogruppo Abeliano di indice 4, e permette una derivazione unificata degli automi di Weyl, Maxwell, e Dirac (si veda il seguito).

zona di Brillouin) si trova che l'evoluzione unitaria è data esattamente dall'equazione di Weyl (i due automi corrispondono all'evoluzione per spinore rispettivamente sinistrorso e destrorso). Con questi due automi elementari di Weyl si possono ottenere automi composti in due modi: mediante somma diretta e mediante prodotto tensore dell'operatore unitario con il suo aggiunto.[10] Nel caso della somma diretta per preservare il principio di località è necessario usare un accoppiamento indipendente dal vettori d'onda nei blocchi fuori diagonale. Dalla composizione degli automi, per ognono dei due possibili automi di Weyl si ottiene un automa con dimensione $s = 4$.

In sintesi, si ha:

Accoppiamento in somma diretta [16]. Si hanno due nuovi automi che nel limite relativistico danno entrambi l'equazione di Dirac. Questi due automi sono connessi da CPT, e la violazione della simmetria denota la rottura della simmetria di Lorentz. Il parametro di accoppiamento nella somma diretta gioca il ruolo della massa a riposo m, e l'unitarietà comporta che i blocchi sulla diagonale siano riscalati di una quantità n, con $n^2 + m^2 = 1$, in modo da mantenere la normalizzazione delle righe dell'operatore, e il fattore di riscalamento gioca il ruolo della velocità della luce. Ne deriva che per soddisfare l'unitarietà la massa a riposo $m \leq 1$ è limitata superiormente, e che all'approssimarsi al valore $m = 1$ la relazione di dispersione dell'automa diventa piatta ($n = 0$). Un tale comportamento ricorda quello di un mini buco nero, laddove la massa della particella è limitata superiormente dall'uguaglianza della lunghezza d'onda Compton con il raggio di Schwarzschild. Si noti infine che la massa inerziale nel contesto della nuova teoria viene introdotta senza dover utilizzare la definizione meccanica circolare basata sulla la legge di Newton [17]: è questo un vantaggio fondazionale notevole, che deriva dal fatto che la teoria è *quantum ab initio*. Per altri aspetti riguardanti l'automa di Dirac, si vedano anche i lavori [18, 19, 20].

[10] Nel caso del prodotto tensore con lo stesso operatore si ottiene l'evoluzione di più particelle identiche, mentre il tensore con l'aggiunto dà luogo ad un nuovo automa. Invece, l'unico modo di costruire un automa per somma diretta che preserva il principio di località è quello di sommare a blocchi l'operatore unitario con il suo aggiunto, e usare un accoppiamento indipendente dal vettor d'onda nei blocchi fuori diagonale.

Accoppiamento a prodotto tensore [21]. Si hanno due nuovi automi con lo stesso limite relativistico, corrispondente ad un campo vettoriale che evolve secondo l'equazione di Maxwell, e un campo scalare costante. Il campo vettoriale non è fatto di Bosoni, bensì di coppie di Fermioni entangled. La statistica che ne risulta è comunque indistinguibile da quella Fermionica, e non sono rilevabili effetti di saturazione Fermionica anche con potenze di radiazione dell'ordine del Petawatt, come le potenze ottenibili da un laser Vulcan laser. L'equazione di Maxwell risulta debolmente distorta. Si hanno principalmente due effetti: 1) esiste una componente longitudinale della polarizzazione, e il vettore di Pointing, la velocità di gruppo e il vettor d'onda hanno direzioni leggermente diverse; 2) la velocità di gruppo differisce di poco da c, dipendendo dalla lunghezza d'onda e dalla direzione. L'effetto 1 è praticamente non rilevabile sperimentalmente, mentre l'effetto 2 si accumula durante la propagazione, e può essere rivelato come shift di pacchetti di radiazione a frequenze diverse provenienti da un evento cosmico ai limiti dell'universo visibile. La teoria a Fermioni della radiazione era stata già considerata da De Broglie, ma fu in seguito scartata perchè non obbediva alla statistica di Bose-Einstein.

3.1 Unità di misura all'interno della teoria

Il bound sul valore della massa permette anche di stabilire le unità di misura all'interno della teoria. Infatti, una teoria fondata su assiomi necessariamente matematici (sebbene dotati di interpretazione fisica), avrà necessariamente variabili solo adimensionali, che potranno essere scritte come rapporto fra il valore fisico misurato e il valore campione. Nella teoria ad automa quantistico la variabile tempo adimensionale è intera, e si scriverà quindi come il rapporto fra il tempo fisico e il valore campione t_P. Analogamente per la distanza spaziale in termini della distanza campione x_P. Invece, la massa divisa per la massa campione m_P è un numero compreso fra 0 e 1. La massa campione m_P è quindi il valore massimo della massa di una particella, e può essere misurato, in principio, mediante determininazione sperimentale dell'indice di rifrazione di massa (come già accennato, a causa dell'unitarietà n^{-1} giuoca il ruolo di indice di rifrazione). Dal confronto della teoria dell'automa con l'usuale teoria quantistica di campo nel limite relativistico si ricavano le relazioni

$c = x_P/t_P$ e $\hbar = m_P x_P c$, dove c rappresenta la velocità della luce, e $\hbar$ la costante di Planck. La teoria di automa cellulare quantistico prevede per la radiazione una velocità che nel limite relativistico dipende linearmente dal vettor d'onda. Da tale dipendenza si può determinare ad esempio x_P, mentre t_P sarà dato da x_P/c. Quindi, conoscendo il valore di $\hbar$ e di c, dalla determinazione sperimentalmente di x_P oppure di m_P otteniamo i valori delle tre misure campione. Un ordine di grandezza di tali valori può essere ottenuto assumendo ad esempio che l'accelerazione gravitazionale dovuta a una massa m_P a distanza minima x_P dia la costante di gravitazione G secondo la legge di Newton, nel qual caso $m_P = 2.176 * 10^{-8}$kg è la massa di Planck, $x_P = 1,616 * 10^{-35}$m è la lunghezza di Planck e $t_P = 5.391 * 10^{-44}$s è il tempo di Planck.

3.2 Bontà del limite relativistico

La teoria di campo ad automa cellulare quantistico è un'estensione della teoria quantistica di campo a scale molto piccole, dell'ordine della scala di Planck, e contiene la teoria di campo quantistica usuale come limite relativistico di piccoli vettor d'onda (come abbiamo visto, per "piccoli" si intende rispetto al vettor d'onda di Planck).[11] La teoria rappresenta in modo molto fedele la teoria di campo usuale nel limite relativistico: ad esempio si calcola [16] che l'evoluzione dell'automa per un protone relativistico è indistinguibile da quella di Dirac, con qualunque tecnologia, per un evoluzione temporale di milioni di anni, mentre effetti alla portata della tecnologia attuale sono visibili su evoluzioni di miliardi di anni.

4. COVARIANZA DI LORENTZ NELLA TEORIA DI CAMPO AD AUTOMA CELLULARE, E NOZIONE DI PARTICELLA

Uno dei punti concettualmente più rilevanti della nuova teoria di campo ad automa è che essa permette di derivare la teoria di campo senza assumere la meccanica, e quindi nemmeno l'invarianza di Lorentz, ma come conseguenza solo di principi molto generali, come quelli elencati nella Sezione 2. Inoltre, oltre che essere *quantistica ab initio*–ovvero

[11] Ad esempio una particella di un *ultra high energy cosmic ray* ha vettor d'onda 10^{-8} volte il valore del vettor d'onda di Planck.

non necessita della quantizzazione della teoria classica–la nuova teoria, essendo discreta, elimina per definizione tutti i problemi dalla teoria di campo usuale derivanti dal continuo: in particolare, oltre alle divergenze ultraviolette, il problema della localizzazione della particella. Il calcolo del propagatore come *path integral* è ben definito, ed è stato calcolato analiticamente per il caso di Dirac in una dimensione spaziale [22] e per quello di Weyl in due dimensioni [23]. L'invarianza di Lorentz è riottenuta esattamente nel limite di piccoli vettori d'onda.

Cosa succede del principio di relatività alla scala discreta? Accade che il gruppo di Lorentz diventa non lineare [24], ed avvengono nuovi fenomeni, quali ad esempio, il fenomeno della *relative locality* [25]–fenomeno previsto una quindicina di anni fa da Amelino-Camelia e Smolin–in base al quale non solo la simultaneità temporale di eventi, ma anche la loro vicinanza spaziale dipendono dall'osservatore. Nel contesto della teoria ad automa il cambiamento di sistema di riferimento è definito come un cambiamento di rappresentazione dell'automa, corrispondente a rimappare il vettor d'onda $k \mapsto k'(k)$ e corrispondentemente la frequenza $\omega(k) \mapsto \omega'(k')$. È interessante notare che per massa $m = 0$ sopravvive il concetto di particella come invariante del gruppo di Poincaré (nel nostro caso il gruppo di Poincaré nonlineare), e che ad esempio, per l'automa di Dirac si hanno quattro particelle (e rispettive antiparticelle) diverse in corrispondenza di quattro zone invarianti all'interno della zona di Brillouin del BCC. Infine, se la massa a riposo m è nonnulla, nel caso ultrarelativistico le trasformazioni nonlineari hanno anche una dipendenza dal boost della massa a riposo, e ciò forza ad estendere il gruppo di invarianza al gruppo di de Sitter [26], mentre la simmetria di Lorentz è ritrovata all'ordine m^2.

Ringraziamenti

Desidero vivamente ringraziare la collaborazione intensa, costante, e cruciale del Prof. Paolo Perinotti, e il notevole lavoro dei Dott. Alessandro Bisio e Alessandro Tosini. Desidero anche ringraziare l'incoraggiamento e il sostegno dei membri dell'Accademia prof. Attilio Rigamonti e Prof. Giorgio Goggi. Questa ricerca è finanziata dalla Templeton Fundation (project ID# 43796 *A Quantum-Digital Universe*), e dalla *Fondazione della Banca del Monte di Lombardia*.

REFERENCES

[1] G. M. D'Ariano, *Da quali principi discende la Meccanica Quantistica?*, Rendiconti Istituto Lombardo - Accademia di Scienze e Lettere **142** 23 (2008)

[2] G. Chiribella, G. M. D'Ariano, P. Perinotti, *Probabilistic Theories with Purification*, Phys. Rev. A **81** 062348 (2010)

[3] G. Chiribella, G. M. D'Ariano, P. Perinotti, *Informational derivation of Quantum Theory*, Phys. Rev A **84** 012311 (2011)

[4] Č. Brukner, *Questioning the rules of the game*, Physics **4** 55 (2011).

[5] G. M. D'Ariano, and C. Macchiavello, *Introduzione alla teoria dell'Informazione Quantistica*, Rendiconti Istituto Lombardo - Accademia di Scienze e Lettere **137** 289 (2003)

[6] H. Halvorson and R. Clifton, *No Place for Particles in Relativistic Quantum Theory?*, Phil. Sc. **69** 1 (2002)

[7] M. Kuhlmann, *What is Real?*, Sc. Am. **32** August, (2013) 1

[8] G. M. D'Ariano, *La Fisica come Informazione: l'universo è un grande automa quantistico*, Rendiconti Istituto Lombardo - Accademia di Scienze e Lettere **144** 253 (2010)

[9] G. M. D'Ariano, *A Quantum Digital Universe*, Il Nuovo Saggiatore **28** 13 (2012)

[10] G. M. D'Ariano, *L'Univers est-il un ordinateur quantique?*, La Recherche, Juillet-Aout 2014, n.489, pag.48

[11] G. M. D'Ariano, *It from Qubit*, in A. Aguirre et al. (eds.), *It From Bit or Bit From It?*, The Frontiers Collection (Springer 2015)

[12] G. M. D'Ariano, *Quantum-Informational Principles for Physics*, in A. Aguirre et al. (eds.), *Questioning the Foundations of Physics*, The Frontiers Collection (Springer 2015)

[13] R. T. Cox, The Algebra of Probable Inference (Johns Hopkins University Press, Baltimore 1961)

[14] E. T. Jaynes, Probability Theory: The Logic of Science (Cambridge University Press, Cambridge 2003)

[15] G. Birkhoff and J. von Neumann, *The logic of quantum mechanics*, Ann. Math. **37** 823 (1936)

[16] G. M. D'Ariano, P. Perinotti, *Derivation of the Dirac Equation from Principles of Information processing*, Phys. Rev. A **90** 062106 (2014)

[17] M. Jammer, Concepts of Mass in Contemporary Physics and Philosophy, (Princeton University Press, Princeton 1999)

[18] A. Bisio, G. M. D'Ariano, A. Tosini, *Quantum Field as a Quantum Cellular Automaton: the Dirac free evolution in one dimension*, Ann. Phys. **354** 244 (2015)

[19] G. M. D'Ariano, F. Manessi, P. Perinotti, A. Tosini, The Feynman problem and Fermionic entanglement: Fermionic theory versus qubit theory Int. J. Mod. Phys. A **29** 1430025 (2014)

[20] G. M. D'Ariano, F. Manessi, P. Perinotti, A. Tosini, Fermionic computation is non-local tomographic and violates monogamy of entanglement, EPL **107** 20009 (2014)

[21] A. Bisio, G. M. D'Ariano, P. Perinotti, Quantum Cellular Automaton Theory of Light, arXiv: 1407.6928

[22] G. M. D'Ariano, N. Mosco, P. Perinotti, A. Tosini, *Path-integral solution of the one-dimensional Dirac quantum cellular automaton*, Phys. Lett. A **378** 3165 (2014)

[23] G. M. D'Ariano, N. Mosco, P. Perinotti, A. Tosini, *Discrete Feynman propagator for the Weyl quantum walk in 2+1 dimensions*, EPL **109** 40012 (2015)

[24] A. Bisio, G. M. D'Ariano, P. Perinotti, *Lorentz symmetry for 3d Quantum Cellular Automata*, arXiv:1503.01017

[25] A. Bibeau-Delisle, A. Bisio, G. M. D'Ariano, P. Perinotti, A. Tosini, Doubly-Special Relativity from Quantum Cellular Automata, EPL 109 50003 (2015)

[26] A. Bisio, G. M. D'Ariano, P. Perinotti, in preparazione.

LA SABBIA E I FISICI

Nota del m.s. JACQUES VILLAIN (*)

(Adunanza del 25 ottobre 2012)

SUNTO. – Partendo da considerazioni sulla utilizzazione della sabbia come strumento di visualizzazione di alcuni fenomeni, la nota esamina il comportamento dinamico di sistemi granulari sottoposti a sollecitazioni. Si illustrano in particolare le modificazioni da apportare alla descrizione termodinamico-cinetica classica dei gas di molecole in ragione della non-conservazione della energia nelle collisioni tra i grani e le inaspettate conseguenze concettuali e pratiche di tali modificazioni.

***

ABSTRACT. – The dynamical response of granular systems under certain excitations is considered. In particular the breakdown of the energy conservation in the collisions among grains, with respect to the classical termodynamical- kinetical description of the molecule in gases, is taken into account and it is addressed how unexpected conceptual and pratical consequences arise.

1. LA SABBIA CHE SI ARRAMPICA

Da molto tempo la sabbia è stata usata dai fisici come uno strumento. Un impiego di particolare successo fu quello del tedesco Chladni che, due secoli fa, usò la sabbia per evidenziare i modi di vibrazione di una lastra. Le *figure di Chladni* [1] sono anche utili per mostrare le onde stazionarie in un violino o una chitarra.[2] Invece della sab-

(*) Theory group, ESRF, 6 Rue Jules Horowitz, 3800 Grenoble.
E-mail: jvillain@esrf.fr

bia, altri materiali granulari possono essere usati, e nella presente comunicazione, la parola "sabbia" deve essere considerata come un'espressione generale per indicare "materia granulare".

Alcuni anni dopo, Michael Faraday cominciò a usare la sabbia come specifico oggetto di studio. Scuotendo uno strato di sabbia inizialmente piano, osservò che si formano mucchi.[3]

Da un quarto di secolo, la sabbia scossa è diventata oggetto di studio abbastanza comune, che ha attratto l'attenzione di alcuni fisici illustri. In questa nota sono ricordate alcune proprietà legate a questo problema e fatte alcune osservazioni ad integrazione.

Si può dapprima ricordare un lavoro di Bizon *et al.* [4] che, scuotendo uno strato di sabbia, hanno osservato la formazione di una grande varietà di figure, molto più complicate dei mucchi di Faraday. Infatti la sabbia viene proiettata nell'aria, dove i grani subiscono collisioni, poi ricadono sulla lastra (Douady et al. *Europhys. Lett.,* **8,** 621, 1989). Durante ogni periodo, la sabbia è ora nell'aria, ora sulla lastra, e la sua posizione sulla lastra cambia ad ogni periodo.

In un altro tipo di esperimenti, la sabbia è messa in una scatola scossa con violenza, e la maggioranza dei grani rimane quasi sempre nell'aria. Parlerò più in dettaglio di questa situazione perché il gas formato dalla sabbia ha proprietà sorprendenti.

2. La sabbia in volo

Consideriamo una scatola in forma di parallelepipedo, scossa nella direzione verticale z, e molto grande nelle direzioni orizzontali. La quantità caratteristiche sono soltanto funzioni di z. Il caso più semplice è quello senza gravità, studiato sperimentalmente per esempio da Evesque.[5] (*Fig. 1*)

I grani di sabbia (che in realtà sono generalmente sfere identiche di massa m) formano un gas, le cui particelle collidono l'un l'altra. Le collisioni conservano l'impulso, ma non conservano l'energia cinetica! In ogni collisione una parte dell'energia cinetica è trasformata in calore, e questo calore è in qualche modo dissipato e può essere dimenticato. La perdita di energia in ogni urto può essere dell'ordine del 10%. La velocità delle particelle quindi diminuisce quando aumenta la distanza alle pareti vibranti.

Il caso più interessante è il regime stazionario. In questo caso

abbiamo due ugualianze utili. In primo luogo, il flusso di particelle deve essere nullo, cioè, se $\rho(z,v_z)$ è la densità locale di particelle di velocità v_z, avremo

$$\int_{-\infty}^{\infty} v_z \rho(z,v_z) dv_z = 0 \qquad (1)$$

Un'altra particolarità del caso stazionario è che lo sforzo σ deve essere uniforme. La componente più interessante è σ_{zz}, che chiamerò P. È facile dimostrare [6] che, se ρ è la massa per unità di volume, si ha

$$P = \sigma_{zz} = \rho \langle v_z^2 \rangle \qquad (2)$$

dove <...> è il valore medio e v è la velocità delle particelle. Questa relazione è del tutto analoga a quella di un gas perfetto all'equilibrio, $PV=Nk_BT$, ma la temperatura locale T definita da $m\langle v_z^2\rangle = k_BT$ dipende da z e, come abbiamo visto, diminuisce quando aumenta la distanza dalle pareti vibranti. Siccome la pressione (2) è uniforme, la densità ρ aumenta con questa distanza.

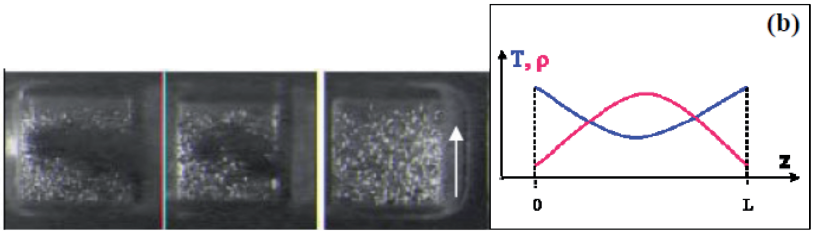


Fig. 1 – Contenitore vibrante con (dalla sinistra alla destra) 3, 2 et 1 strati di palline (Evesque 2009). La zona centrale scura è una zona da alta densità. A destra, variazione della densità ρ *(in rosso) e dell'energia cinetica media* T *(in blu) in una scatola di altezza* L *in gravità nulla.*[5]

Se lo sforzo può facilmente essere definito nel caso stazionario, sarebbe difficile dare una tale definizione fuori dall'equilibrio....

Evesque ha introdotto uno sforzo asimmetrico:

$$P^+(z) = \sigma_{zz}^{+} = \rho \langle v_z^2 \rangle^+ = \int_0^{\infty} v_z^{\,2} \rho(z,v_z) dv_z \quad : \quad P^-(z) = \sigma_{zz}^{-} = \rho \langle v_z^2 \rangle^- = \int_{-\infty}^{0} v_z^{\,2} \rho(z,v_z) dv_z \quad (3)$$

Queste quantità non sono uguali, particolarmente nella vicinanza di una parete vibrante. Consideriamo la parete inferiore. Le particelle da velocità positiva sono molto più veloci delle particelle a componente della velocità negativa, che hanno perso energia in seguito a collisioni. Ma la (1) impone che il prodotto $\rho|v_z|$ sia uguale in valore medio per i due tipi di particelle. Il prodotto $\rho|v_z|^2$ deve quindi essere maggiore per le particelle a velocità positiva, e $P^+ > P^-$. Vicino alla parete superiore $P^+ < P^-$, e al centro, $P^+ = P^-$.

L'interesse di P^+ e P^- è che la loro differenza ha un significato fisico anche fuori dalla stazionarità. Infatti, se mettiamo nel sistema una particella pesante, la forza che agisce su questa particella è

$$F = k[P+(z) - P-(z)] \qquad (4)$$

che non è nulla, tranne al centro della scatola. Il coefficiente k è dell'ordine di grandezza della sezione della particella, cioè πR^2 per una sfera di raggio R. La particella pesante è spinta verso il centro della scatola.

3. Segregazione e diavoletto di Maxwell

Delle osservazioni precedenti si deduce che, nell'assenza di gravità, particelle pesanti vengono attratte dalla parte media della scatola. Questo è un caso di segregazione frequentemente osservato nella sabbia scossa. Un altro esempio classico è quello di una miscela densa di grani grossi e piccoli. Se la miscela è scossa durante un tempo abbastanza lungo, le particelle più grandi salgono alla superficie (*Fig. 2*). Il fenomeno è differente di quello che abbiamo descritto, e a nostra conoscenza non è ancora del tutto compreso.

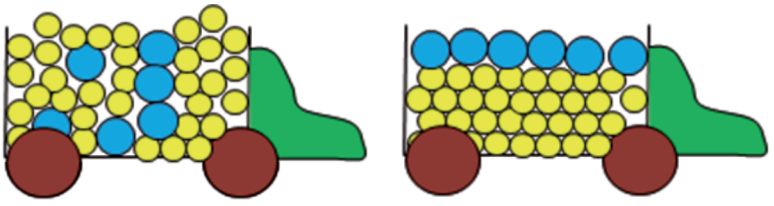


Fig. 2 – Un camion, partendo con un carico di noci mescolate, arriva con le noci più grandi sullo strato superiore.

Un caso più strano di segregazione può essere osservato scuotendo una scatola a due compartimenti (*Fig. 3*). In certi casi un compartimento si svuota in parte, e nello stato finale i due compartimenti contengono un numero disuguale di particelle. Questo fatto, degno di un "diavoletto di Maxwell", dimostra la caduta di validità delle proprietà generali della meccanica statistica usuale, che sono il principio del bilancio dettagliato, il principio di Carnot e il teorema H. L'effetto è stato osservato sperimentalmente in presenza di gravità [7,8] se la scatola è scossa con un'energia sufficiente, e senza gravità [9] è stato ottenuto in simulazioni quando il numero di particelle è abbastanza grande.

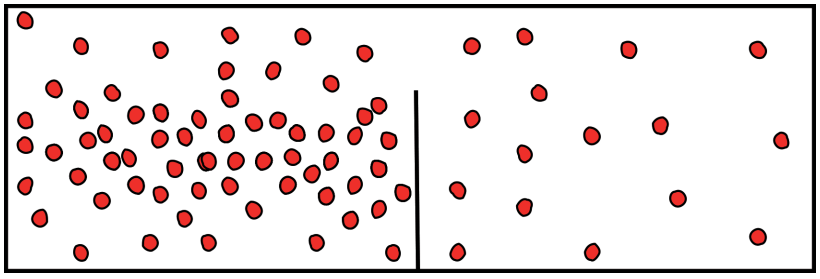


Fig. 3 – Il diavoletto di Maxwell. In certi casi la situazione stabile corrisponde a densità diverse in due compartimenti comunicanti.

Meno strano, ma interessante, è il fatto (*Fig. 4*) che una clessidra si svuota più lentamente quando viene scossa.[10] Un nuovo diavoletto impedisce alla gravità di operare. Un modo di spiegarlo è di dire che la densità è più bassa nella vicinanza della parete, e quindi della buca. Non lasciamo la clessidra senza osservare che, se non la scuotiamo, la portata è costante, indipendente dell'altezza di sabbia. Questa proprietà è caratteristica della materia granulare.

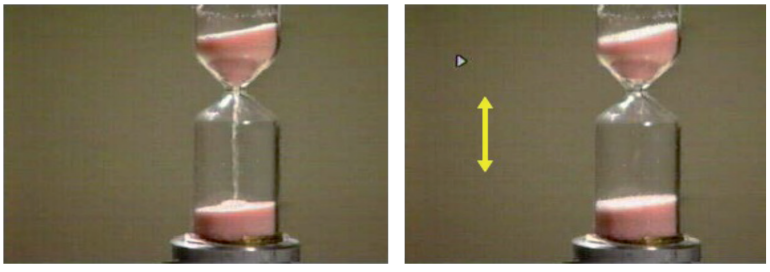


Fig. 4 – Una clessidra a riposo (a sinistra) si svuota normalmente, quando viene scossa quasi smette di svuotarsi (a destra).

4. Tracce sulla sabbia

In un loro libro, Rigamonti e Varlamov [11] riportano le seguenti parole di Reynolds, il grande esperto di idrodinamica, scritte nel 1885: "*Quando il piede preme sulla sabbia, sul tratto compatto dopo il ritiro della marea, la regione che si trova attorno al piede diventa immediatamente asciutta. La pressione del piede rende la sabbia meno compatta e quanto più tale pressione è forte, tanta più acqua fuoriesce... Essa rende la sabbia asciutta finché dal basso non arriva altra acqua*".

Se capisco bene, questo significa che dopo il ritiro della marea la sabbia ha trovato una struttura ben compatta che lo sforzo verticale del piede perturba, così formando lacune dove l'acqua può penetrare in maggiore quantità (*Fig. 5*). È usuale, particolarmente in elasticità, che una compressione in una direzione produca un'espansione nelle alte, ma generalmente il volume diminuisce. In elasticità, non credo che sia conosciuto un caso dove la compressione in una direzione produca un aumento di volume. Per la sabbia (che ovviamente non è mezzo elastico) è così.

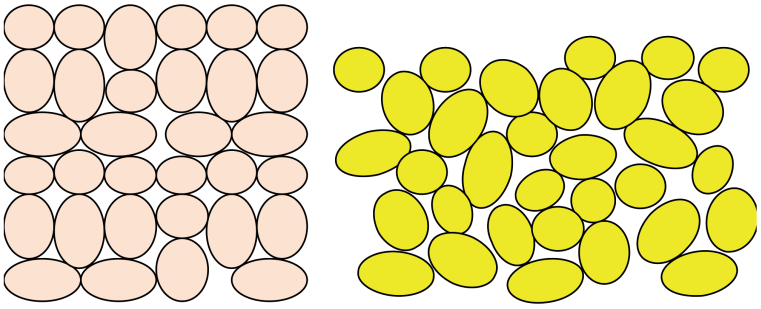


Fig. 5 – Una sabbia che ha formato una struttura ben compatta (a sinistra) prende una forma meno compatta sotto l'effetto di uno sforzo.

Il lettore troverà altri aspetti sulla sabbia nel libro citato, che presenta versioni in inglese e in russo. Una versione innovata ed arricchita è disponibile in francese, con aggiornamenti e integrazioni.[12] Ma ci sono libri interi sulla fisica della sabbia, dei quali la presente nota è soltanto una parziale illustrazione.[13,14] Per quanto riguarda in particolare l'effetto dello sforzo sulla densità, si può utilmente ricorrere all'articolo di rassegna di Evesque.[15]

Per finire voglio ricordare, citando de Gennes,[16] che la mecca-

nica della sabbia non è soltanto importante dal punto di vista dei fondamenti della fisica, è anche importante per la sopravvivenza degli uomini che vivono intorno al deserto.

BIBLIOGRAFIA

[1] Ernst F.F. Chladni, *Entdeckungen über die Theorie des Klange* (1787).
[2] G. Benedetto, R. Pisani, Articolo Acustica musicale (Tavola V) Enciclopedia delle scienze fisiche, Treccani, Roma (1992).
[3] M. Faraday, Philos. Trans. R. Soc. London 52, 299 (1831).
[4] C. Bizon, M.D. Shattuck, J.B. Swift, W.D. McCormick, and Harry L. Swinney., Physical Review Letters 80, 57 (1998).
[5] P. Evesque, Poudres et grains 17, 577 (2009).
[6] J. Villain, Poudres & grains 20, 29 (2012).
[7] J. Eggers, Phys. Rev. Lett. 83, 5322 (1999).
[8] Una bella realizzazione sperimentale del diavoletto di Maxwell, fra altri esperimenti sulla materia granulare, è visibile sul sito http://www.palais-decouverte.fr/index.php?id=1662
[9] Y. Li, M. Hou, P. Evesque, Journal of Physics: Conference Series 327 (2011) 01203410.
[10] P. Evesque & W. Meftah, C.R. Acad. Sciences 314, série II, 1125-32, (1992).
[11] A. Rigamonti, A. Varlamov, *Magico caleidoscopio della fisica*. La Goliardica pavese (2007).
[12] A. Rigamonti, A. Varlamov, J. Villain, *Le kaléidoscope de la Physique,* Ed. Belin (Paris 2014).
[13] Jacques Duran *Poudres, sables et grains* (Eyrolles, 1997) *Sand, powders and grains* (Springer 2000) http://books.google.fr/books?id=9kK1ZxsAS_EC&printsec=frontcover&hl=fr#v=onepage&q&f=false
[14] Jacques Duran *Sables émouvants* (Belin, 2003).
[15] P. Evesque*, Eléments de mécanique quasi-statique des milieux granulaires mouillés ou secs,* Poudres & grains NS 1, 1-60 (2000). http://www.poudres-et-grains.ecp.fr/IMG/pdf/poudresNS1.pdf
[16] http://archives.arte.tv/hebdo/archimed/19980915/ftext/sujet5.html

VELOCITÀ LIMITE, CAUSALITÀ E TEST DELLA RELATIVITÀ RISTRETTA

Nota del m.e. GIOVANNI MARIA PROSPERI (*)

(Adunanza del 25 ottobre 2012)

SUNTO. – Prendendo spunto dall'annuncio, poi rientrato, della collaborazione OPERA dei Laboratori del Gran Sasso dell'osservazione di neutrini superluminali, si richiamano brevemente alcuni principi base della Teoria della Relatività e si discutono alcune conferme molto stringenti della teoria che vengono dalla mancate osservazioni di fenomeni tipo effetto Cherenkov e indirettamente dagli stessi risultati di OPERA.

***

ABSTRACT. – With reference to the claim, subsequently withdrawn, by the "OPERA" collaboration of "National Gran Sasso Laboratories" of the observation of superluminal neutrinos, I briefly recall some basic principles of Relativity Theory and discuss some very stringent positive tests that come from the lack of observation of some effect of the type of Cherenkov effect and indirectly from the OPERA results themselves as stressed by some authors.

1. INTRODUZIONE

Nel settembre del 2011 l'equipe dell'esperimento OPERA, presso i laboratori Laboratori Nazionali del Gran Sasso, sconvolse un po'

(*) Dipartimento di Fisica dell'Università di Milano, Istituto Lombardo, Accademia di Scienze e Lettere, Milan, Italy.
E-mail: Giovanni.Prosperi@mi.infn.it

l'ambiente dei fisici con l'annunzio sensazionale di avere osservato neutrini propagantisi con velocità superiore a quella della luce [1].

L'esperimento coinvolgeva circa duecento fisici appartenenti ad una quarantina di istituzioni internazionali diverse. Il suo scopo principale era lo studio delle oscillazioni di neutrini. La collaborazione aveva realizzato un rivelatore di neutrini in grado di distinguere i tre diversi tipi di neutrini, ν_e, ν_μ, ν_τ (neutrini-e, neutrino-μ, neutrino-τ), disposto in una delle grandi hall dei laboratori sotterranei del gran Sasso, sulla linea di un fascio proveniente dal CERN di Ginevra. Si tratta di un fascio di ν_μ prodotto dal supersincrotrone a protoni (SPS) del CERN. Un breve impulso di protoni, di 400 GeV di energia[1] e della durata di circa 10,5 μs, viene estratto dalla macchina e proiettato entro un cilindro di grafite, dove per collisione con i nuclei di carbonio produce un corrispondente impulso di mesoni π^+. Queste particelle sono successivamente indirizzate verso un tunnel a vuoto di circa 1000 metri di lunghezza diretto verso i laboratori del Gran Sasso lungo il quale decadono in muoni positivi e neutrini ν_μ. I μ^+ presenti nel fascio vengono allontanati tramite un campo magnetico, i protoni e gli altri adroni sono arrestati su un blocco di ferro. Emerge un fascio pulsato quasi puro di neutrini ν_μ con un'energia media di 17,5 GeV che penetra nella crosta terrestre e raggiunge quasi indisturbato (i neutrini interagiscono con le altre particelle molto debolmente) il tunnel del Gran Sasso.

Lo studio dei neutrini provenienti dal sole e vari esperimenti specifici avevano mostrato che durante la loro propagazione libera i neutrini ν_e e ν_μ oscillano, cioè si trasformano alternativamente gli uni negli altri. Vari tipi di argomenti suggerivano che anche il neutrino ν_τ dovesse essere coinvolto in tale oscillazione, ma il fenomeno non era mai stato osservato in maniera diretta. Lo scopo principale dell'esperimento OPERA era quello di confermare che la trasformazione $\nu_\mu \rightarrow \nu_\tau$ avvenisse effettivamente. In questo senso l'esperimento sembra avere avuto pieno successo. Durante tre o quattro anni di operazione quattro neu-

[1] Per i non fisici ricordo che nella Fisica Atomica si usa come unità base per l'energia l'elettronvolt (eV), definito come l'energia acquistata da un elettrone quando è accelerato in una differenza di potenziale di un Volt. Accanto all'elettronvolt nella Fisica, dei Nuclei e delle Particelle sono molto usati i multipli 1MeV=10^6 eV, 1 GeV=10^9 eV, 1 TeV=10^{12} eV.

trini ν_τ sono stati effettivamente osservati. Il numero è più piccolo di quello atteso ma è ritenuto significativo.

L'apparecchiatura pareva prestarsi, tuttavia, anche ad una misura per la prima volta diretta del *tempo di volo* dei neutrini dal luogo di produzione al CERN fino al Gran Sasso. Questo poteva essere ottenuto confrontando il profilo temporale degli impulsi dei protoni all'atto della loro estrazione con quello della rivelazione dei neutrini ν_μ nell'apparecchio. La distanza tra la sorgente al CERN e il rivelatore al Gran Sasso, di circa 730 km, ha potuto essere misurata con i metodi della Geodesia moderna (basati sull'uso dei satelliti) con la precisione di 20 cm, le misure di tempo eseguite con orologi atomici sincronizzati ai satelliti del sistema GPS erano stimate precise a qualche nanosecondo. In questo modo la collaborazione credette di aver stabilito che la velocità c_ν dei neutrini era di 7,5 km/s maggiore di quella della luce, $\delta=(c_\nu^2-c^2)/c^2\cong 5\times 10^{-5}$.

Il risultato era in qualche modo sensazionale perché in contraddizione con i principi base della Teoria Relatività Ristretta e, se fosse stato confermato, avrebbe posto in discussione tutto un quadro della Fisica Moderna e in particolare della Fisica delle Particelle che si riteneva ben stabilito. Esso era tuttavia anche in contrasto con vari altri fatti sperimentali, vecchie osservazioni (1987) su neutrini (di energia però molto più bassa, ~10 MeV) emessi dalla supernova SN1987 [2] e tutta una serie di test indiretti che sembravano confermare la Teoria ben al di là della pretesa violazione riscontrata da OPERA. In particolare negli ultimi anni del '900 S. Coleman e S.L. Glashow avevano mostrato che sotto ragionevoli ipotesi una violazione dell'invarianza di Lorentz avrebbe dovuto manifestarsi in una serie di processi del tipo dell'*effetto Cherenkov*, che non apparivano e la cui assenza permetteva al contrario di stabilire dei limiti molto stringenti sull'entità di tali violazioni [3]. Specificamente, in una nota pubblicata dopo l'annuncio del risultato di OPERA dallo stesso S.L. Glashow e da un altro collaboratore, A.G. Cohen, veniva mostrato che, nell'ipotesi di una violazione dell'entità indicata, i neutrini avrebbero dovuto perdere rapidamente energia per produzione di coppie elettrone-positone e non avrebbero dovuto essere osservati neutrini con un'energia superiore ai 12,5 GeV, contro le stesse osservazioni della collaborazione [4].

In questa situazione, altri gruppi sperimentali che operavano nelle hall del Gran Sasso rivolsero la loro attenzione al problema. La collaborazione ICARUS [5] per prima presentava un'analisi della distribuzione

energetica dei neutrini fino ad allora raccolti col proprio rivelatore, mostrando che questa corrispondeva perfettamente alle caratteristiche del fascio iniziale. Una ricerca diretta delle ipotetiche coppie elettrone-positone dava pure risultato negativo; nessuna traccia appariva del fenomeno previsto da Cohen e Glashow. La misura diretta del tempo di volo eseguita da OPERA fu poi ripetuta dall'equipe del rivelatore BOREXINO, progettato per lo studio dei neutrini terrestri e solari, dal rivelatore LVD e ancora da ICARUS, non confermando l'anomalia, trovando invece una velocità dei neutrini uguale, entro gli errori, a quella della luce, come previsto [5-9]. Nel frattempo la stessa equipe di OPERA si era resa conto di un difetto in un'apparecchiatura sperimentale che introduceva un errore sistematico nella misura dei tempi e ripetuto l'esperimento con grande accuratezza ritrovava anche essa la velocità della luce.

Non solo quindi non si era in presenza di neutrini che superavano quella che era ritenuta una velocità limite, ma al contrario la non osservazione della emissione di coppie $e^- e^+$ permetteva di stabilire un limite superiore sul valore della grandezza δ, di gran lunga più stringente della precisione con cui si sarebbe mai potuto pensare di misurare la velocità in modo diretto; $\delta<4\times10^{-8}$ secondo ICARUS, $\delta<1{,}7\times10^{-11}$, combinando vari risultati secondo Cohen e Glashow. L'esperimento che sembrava aver messo in qualche modo in discussione una teoria ormai universalmente accettata aveva alla fine portato in modo più o meno diretto ad una ulteriore straordinaria conferma della stessa.

Dato l'indubbio interesse della questione, voglio nel seguito prima richiamare brevemente per i non fisici le idee che sono alla base della Teoria della Relatività Ristretta e quale sia il ruolo di questa teoria nella Fisica delle Particelle. Successivamente tenterò di illustrare alcuni dei fatti che hanno portato a una sua nuova straordinaria conferma e in particolare discutere l'argomento di Coehn e Glashow, specificamente formulato in relazione al presunto risultato di OPERA.

2. Principi base della Teoria della relatività ristretta

Riprendendo un'originaria intuizione di Galileo, Newton aveva mostrato che, supposte le sue leggi della Meccanica valide rispetto ad un certo sistema di riferimento, esse dovevano conservare la stessa forma in ogni altro riferimento in moto traslatorio uniforme rispetto al

primo. Nelle presentazioni moderne questo risultato viene comunemente dedotto osservando che le equazioni di moto di Newton sono invarianti sotto una trasformazione di coordinate della forma

$$x'=x-vt, \qquad y'=y, \qquad z'=z\ ,$$

che si suppone colleghi le coordinate di un corpo puntiforme rispetto ad un sistema di riferimento K con quelle di un secondo sistema K' in moto traslatorio uniforme rispetto al primo con velocità v (nella scrittura precedente gli assi dei due sistemi sono ovviamente supposti paralleli e si suppone che l'asse x' scorra sull'asse x).

Alla classe dei sistemi così individuata si dà il nome di *sistemi inerziali*, perché evidentemente in ciascuno di essi vale il principio di inerzia. Il teorema stabilito da Newton si può quindi riesprimere dicendo che le Leggi della Meccanica hanno la stessa forma in tutti i sistemi inerziali o che questi sistemi sono tutti equivalenti per la descrizione del moto. Si parla comunemente anche di principio della Relatività Galileiano e le trasformazioni su riportate sono dette trasformazioni di Galileo, anche se in realtà né Galileo, né Newton, ancora legati all'idea di un moto assoluto, ne avessero completamente compreso la portata.

Nelle seconda metà dell'ottocento, d'altra parte, Maxwell aveva compendiato le leggi dell'Elettromagnetismo nelle sue famose equazioni, che, nel linguaggio di oggi e nel sistema simmetrico di Gauss, possono essere scritte

$$\operatorname{div} \boldsymbol{E} = 4\pi\rho \qquad \operatorname{div} \boldsymbol{H} = 0$$
$$\operatorname{rot} \boldsymbol{H} - \frac{1}{c}\frac{\partial \boldsymbol{E}}{\partial t} = \frac{4\pi}{c}\boldsymbol{j} \qquad \operatorname{rot} \boldsymbol{E} + \frac{1}{c}\frac{\partial \boldsymbol{H}}{\partial t} = 0$$

In queste equazioni compare la costante c che esprime sostanzialmente la relazione tra campo elettrico e campo magnetico. Una conseguenza fondamentale delle stesse è stata la previsione dell'esistenza di onde elettromagnetiche che si propagano con la velocità c. La coincidenza numerica di c con la velocità della luce aveva indotto Maxwell, come è noto, a supporre che la luce fosse una radiazione elettromagnetica, circostanza successivamente confermata nel modo più ampio e a c ci si riferisce comunemente appunto con questo nome.

Poiché la velocità delle onde è espressa da una costante che compare nelle stesse equazioni fondamentali, sembrerebbe che, a differenza delle leggi della Meccanica, queste non possano valere in qualsiasi rife-

rimento inerziale ma richiedano un sistema privilegiato. Poiché le trasformazioni di Galileo implicano l'additività delle velocità, è infatti ovvio che esse non possono essere invarianti rispetto a queste equazioni. Il sistema privilegiato veniva spontaneamente identificato con l'*etere*, ipotetico mezzo elastico, che si supponeva supporto per la propagazione delle onde luminose e che, corrispondentemente, attraverso le sue deformazioni si sperava potesse permettere un'interpretazione meccanica del campo elettrico e del campo magnetico.

Il fallimento di tutti i tentativi di mettere in evidenza un moto rispetto all'etere, la difficoltà di costruirne un modello ragionevole e soprattutto ragioni di carattere estetico (legate ad una disimmetria che si veniva a creare tra le leggi della Meccanica e quelle dell'Elettromagnetismo) indussero, come è noto, Einstein a rifiutare ogni riferimento a questo mezzo e a formulare la sua famosa Teoria. Questa era fondata su due principi fondamentali: il principio di relatività, secondo cui: ***tutte le leggi fisiche debbono avere la stessa forma in ogni riferimento inerziale***, e il principio di invarianza della velocità della luce, secondo cui ***la luce ha la medesima velocità c in ogni direzione e in ogni riferimento inerziale***.

Einstein fece osservare che le trasformazioni di Galileo e la legata legge di addizione delle velocità, che creavano problemi con le leggi dell'Elettromagnetismo, erano in realtà una conseguenza dell'ipotesi dell'esistenza di un tempo assoluto; più precisamente dell'ipotesi che la misura del tempo in cui si verificava un certo evento fosse indipendente dal riferimento. Egli fece notare che in assenza di un segnale che si propagasse con velocità infinita l'ipotesi era ingiustificata, e che solo un confronto tra eventi immediatamente vicini poteva avere un significato operativo. Osservò che, per definire correttamente una misura dei tempi, ogni sistema di riferimento doveva disporre di una rete indipendente di orologi distribuiti in tutto lo spazio, solidali con esso e sincronizzati con qualche criterio. Coerentemente suppose che il tempo attribuito a un evento dovesse esser quello indicato da un orologio disposto nelle sue immediate prossimità e che proprio il principio di invarianza della velocità della luce dovesse fornire il criterio di sincronizzazione. Considerati due sistemi inerziali K e K' ed indicate con (x, y, z, t) le coordinate spaziali e temporali attribuite ad un *evento elementare* da un sistema K (con $x, y, z,$ che specificano il *luogo* e t che specifica il *tempo* a cui l'evento si verifica) e con (x', y', z', t') quelle attribuite allo stesso evento da un altro sistema inerziale K', mostrò che, imponendo i suoi due principi, si ritrovavano in luogo delle trasformazioni di Galileo

(con la stessa disposizione degli assi adottata per le trasformazioni di Galileo) le trasformazioni speciali di Lorentz

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - v^2/c^2}}$$

che erano state originariamente introdotte su base puramente formale e si sapeva che lasciavano invariate le equazioni di Maxwell.

Per $|v| \ll c$ queste trasformazioni si riducono a quelle di Galileo, per $|v| \geq c$, esse invece perdono senso, suggerendo che la velocità della luce debba essere considerata una velocità limite, non superabile da alcun segnale di carattere fisico.

L'esistenza di una velocità limite è importante, per la coerenza della teoria, per un'altra ragione. Se consideriamo due eventi di coordinate temporali t_1 e t_2 rispetto a K e t_1' e t_2' rispetto a K', a seconda della loro disposizione spaziale può accadere, in base alle trasformazioni di Lorentz, che sia $t_1 < t_2$ e $t_1' > t_2'$, cioè che la successione temporale si inverta nei due riferimenti. Questa circostanza, tuttavia, si verifica solo se i punti P_1 e P_2 in cui accadono gli eventi sono a distanza $(\overline{P_1P_2})$ maggiore di $c(t_2 - t_1)$, cioè di quella che può essere percorsa nel tempo $t_2 - t_1$ da un segnale che si propaghi con la velocità c. Se $\overline{P_1P_2} < c(t_2 - t_1)$, quindi, se i due eventi possono essere in relazione causale, la successione temporale risulta la medesima in tutti i sistemi di riferimento. Se c non può essere superata, quindi, non può mai accadere che venga invertito l'effetto con la causa. A questa proprietà, che sembra essenziale per l'interpretazione, si dà il nome di *legge della causalità relativistica.*

Nel contesto è anche notevole, e non molto noto, che trasformazioni del tipo di quelle di Lorentz, e quindi l'esistenza di una velocità limite, possono essere dedotte da ipotesi molto generali, senza alcun riferimento alla velocità della luce. E' sufficiente sostituire il requisito dell'invarianza di quest'ultima con quello che le trasformazioni formino gruppo, cioè che la composizione di due trasformazioni del tipo $K \to K'$ e $K' \to K''$ dia luogo ad una trasformazione $K' \to K''$ della stessa forma delle due precedenti. La costante c che in esse compare è allora una costante arbitraria, una costante di integrazione. Le trasformazioni di Galileo corrispondono al caso limite per $c \to \infty$, quelle di Lorentz propriamente dette alla identificazione di c con la costante che compare nelle equazioni di Maxwell.

Come abbiamo detto con la sostituzione delle trasformazioni di Lorentz a quelle di Galileo, a meno di una appropriata ridefinizione delle variabili di campo, le equazioni di Maxwell restano invariate. Coerentemente con il principio di relatività, quindi, le leggi dell'Elettromagnetismo, supposte valide in un sistema K, restano valide in qualsiasi altro sistema inerziale K'. E' chiaro però che non restano più invariate le equazioni della Meccanica. In obbedienza al principio, queste dovranno essere perciò modificate; in particolare dovranno essere modificate le espressioni dell'energia e del momento di una particella.

Coerentemente con i nostri scopi consideriamo, a questo proposito, l'urto elastico tra due particelle. Secondo l'ordinaria meccanica non relativistica nell'urto il momento e l'energia totali, devono conservarsi, cioè deve aversi

$$\begin{aligned} \boldsymbol{p}_1+\boldsymbol{p}_2&=\boldsymbol{p}'_1+\boldsymbol{p}'_2 \\ W_1+ W_2&=W'_1+W'_2, \end{aligned}$$

dove i termini a sinistra si riferiscono alle particelle prima dell'urto, quelli a destra alle stesse dopo l'urto. In tale meccanica il momento e l'energia della singola particella sono definite in funzione della velocità $\boldsymbol{v_P}$ dalle espressioni

$$\boldsymbol{p} = m\boldsymbol{v_P}\,, \quad W = \frac{1}{2}m\boldsymbol{v}_P^2 = \frac{1}{2m}\boldsymbol{p}^2\,.$$

Con queste definizioni le leggi di conservazione restano invariate sotto trasformazioni di Galileo ma non lo sono più sotto trasformazioni di Lorentz. Una legge di conservazione invariante sotto tali trasformazioni si ottiene se si ridefiniscono momento ed energia nel modo seguente

$$\boldsymbol{p} = \frac{m\boldsymbol{v}_P}{\sqrt{1-v_P^2/c^2}}\,, \qquad W = \frac{mc^2}{\sqrt{1-v_P^2/c^2}}\,.$$

Le componenti p_x, p_y, p_z di $\boldsymbol{p}$ e la quantità W/c, costituiscono allora quello che si chiama un tetravettore nello *spazio-tempo*, cioè si trasformano rispettivamente come le coordinate spaziali e come il tempo secondo le trasformazioni di Lorentz. Essendo queste trasformazioni lineari e omogenee primo e secondo membro delle relazioni di conser-

vazione si comportano nello stesso modo nel passaggio da un sistema di riferimento all'altro e sono pertanto invarianti.

Osserviamo che nel limite di bassa velocità le nuove definizioni si riducono a quelle non relativistiche, e quindi in queste condizioni la meccanica ordinaria mantiene, come deve, la sua validità. Per $v_P=0$ abbiamo tuttavia un'energia a riposo $W_0=mc^2$ univocamente determinata che esprime la famosa legge *dell'equivalenza tra massa e energia*. Per $\boldsymbol{v}_P \to c$ infine energia e momento divengono infinite e questo è un altro modo di guardare al carattere limite della velocità c.

Nella collisione tra particelle elementari ad alte energie, l'uso delle leggi di conservazione relativistiche diviene ovviamente essenziale. In questo caso si presenta tutta une serie di fenomeni nuovi. Nel processo nuove particelle possono essere prodotte e altre tra quelle originariamente presenti scomparire. Per quanto detto sul modo di trasformarsi dell'energia e del momento, tuttavia, le leggi di conservazione possono essere immediatamente generalizzate in modo invariante anche a tali processi. Si può supporre che il numero delle particelle iniziali o finali sia qualsiasi e che le particelle finali siano diverse da quelle iniziali e scrivere in tutta generalità

$$\sum W_i = \sum W_f \,, \qquad \sum \boldsymbol{p}_i = \sum \boldsymbol{p}_f \,,$$

essendo appunto le somme a sinistra estese a tutte le particelle iniziali (indice i) e quelle a destra a tutte quelle finali (indice f).

L'ultima equazione è alla base di tutta la nostra comprensione dei fenomeni che riguardano la fisica dei nuclei e quella delle particelle. Poiché in tali esperimenti quello che si misura in maniera più diretta è il momento e non la velocità è però utile riesprimere l'energia in funzione del momento. Una tale espressione può essere immediatamente ottenuta dalla relazione $W^2-\boldsymbol{p}^2 c^2=m^2 c^4$; da cui

$$W = c\sqrt{m^2c^2 + \boldsymbol{p}^2} \,.$$

Il fatto che la grandezza $W^2-\boldsymbol{p}^2 c^2$ sia uguale ad una costante e quindi indipendente dal riferimento, è legato alla circostanza che essa è l'analogo per un tetravettore del modulo di un vettore ordinario.

La velocità può essere riottenuta in funzione del momento, invertendo la definizione originaria o a partire da $W(\boldsymbol{p})$ usando la prima equazione di Hamilton,

$$\boldsymbol{v}_P = \frac{\partial W}{\partial \boldsymbol{p}} = \frac{\boldsymbol{p}c}{\sqrt{m^2c^2 + \boldsymbol{p}^2}}\,.$$

Da quest'ultima equazione si ritrova ovviamente che è sempre $v_P<c$ e che inversamente $v_P\to c$ per $p\to\infty$. In pratica, appena pc diviene grande rispetto all'energia a riposo mc^2si può supporre $v_P\cong c$.

Si osservi ancora che le espressioni precedenti mantengono senso quando $m=0$. Nella fisica relativistica ha quindi senso considerare particelle di massa nulla; per esse si ha $W=pc$ ed è sempre esattamente $v_P=c$. Una particella di questo tipo è il fotone γ, il quanto del campo elettromagnetico. Fino qualche tempo fa anche ai neutrini era attribuita una massa nulla. Oggi il fenomeno dell'oscillazione ci permette di affermare che tale massa, molto piccola e non completamente nota, è però certamente diversa da zero (dell'ordine di una frazione di eV). Tutte le volte che l'energia complessiva è grande rispetto all'energia a riposo è tuttavia sempre lecito trascurare nell'espressione $W(\boldsymbol{p})$ il termine di massa e questa è la situazione che quasi sempre si presenta nel caso di neutrini.

3. Limiti su possibili violazioni della Teoria

Le conferme della teoria relativistica sono così numerose e i successi, in particolare nello studio del comportamento dei nuclei e delle particelle, così grandi, che eventuali violazioni della stessa devono necessariamente essere piccole. In ogni caso vista la generalità delle considerazioni con cui, come abbiamo visto, esse possono essere dedotte, possiamo sempre ammettere che le equazioni di trasformazione tra un sistema di riferimento e l'altro siano del tipo delle trasformazioni di Lorentz con c velocità limite assoluta di valore prossimo, ma eventualmente di qualcosa maggiore di quello della luce. Quello che eventualmente invece ammeteremo come ipotesi di lavoro è che le equazioni dinamiche siano solo approssimativamente Lorentz invarianti

Con l'intento di stabilire limiti quantitativi alla precisione entro cui la Teoria della Relatività Ristretta potesse considerarsi confermata dai dati, diversi autori fin dal 1986 partirono dall'ipotesi di lavoro che, alle espressioni delle lagrangiane relativistiche relative ai campi associati secondo la Teoria Quantistica ai vari tipi di particelle, andassero aggiunti piccoli termini che violassero l'invarianza di Lorentz. Si fece

però l'ipotesi che dovesse comunque esistere un sistema di riferimento privilegiato $\boldsymbol{F}$ (tendenzialmente identificato con quello del centro di massa della radiazione del fondo cosmico a microonde (CBM)) in cui tali termini fossero essere invarianti per rotazione e per traslazione (che in tale riferimento non ci fossero quindi posizioni o direzioni privilegiate) e si cercò di stabilire quali effetti tali termini potessero avere. Uno studio sistematico in tale spirito è stato in particolare condotto nel 1997-1999, come abbiamo detto, da Coleman e Glashow [3].

Per semplificare le notazioni, in questa sezione supporremo di scegliere l'unità di misura dei tempi, come consueto nella Teoria delle Particelle, in modo che la velocità limite assoluta risulti uguale a l. In tutte le equazioni precedenti dovremo quindi porre c=1.

Sotto le ipotesi fatte sopra, l'unico effetto di termini di violazione dell'invarianza relativistica nella lagrangiana per un certo tipo di particella, può essere quello di aggiungere nel sistema privilegiato al termine m_P^2 nell'espressione di $W^2-\boldsymbol{p}^2$ un ulteriore termine negativo proporzionale a $\boldsymbol{p}^2$. Scriveremo perciò

$$W^2-\boldsymbol{p}^2=m_P^2-\varepsilon_P\boldsymbol{p}^2,$$

con ε_P parametro piccolo, che in generale potrà esser supposto diverso da particella a particella. Equivalentemente l'espressione dell'energia nel sistema $\boldsymbol{F}$ andrà scritta

$$W=\sqrt{m^2+\boldsymbol{p}^2c_P^2} \qquad \text{con} \qquad c_P=\sqrt{1-\varepsilon_P}\,.$$

e per la velocità avremo

$$\boldsymbol{v}_P=\frac{\partial W}{\partial \boldsymbol{p}}=\frac{c_P^2\boldsymbol{p}}{\sqrt{m^2+\boldsymbol{p}^2c_P^2}}\,.$$

Nell'ipotesi c_P rappresenta la velocità limite, sempre in $\boldsymbol{F}$, per la specifica particella P e naturalmente non dovrà essere superiore alla velocità limite assoluta, cioè dovrà essere con le nostre convenzioni $c_P\leq 1$.

Per $c_P<1$ e, in particolare, diverso da particella a particella diventano possibili processi che in una ordinaria teoria relativistica sarebbero vietati dalle leggi di conservazione. L'espressione dell'energia in un sistema generico K può essere ottenuta da quella in $\boldsymbol{F}$ con una trasformazione di Lorentz trattando, come indicato, l'energia come un tempo e le componenti del momento come coordinate spaziali. La forma dell'espressione sarà però più complicata e dipenderà dalla direzione della velocità K rispetto a $\boldsymbol{F}$. Per le considerazione che faremo un riferimento

al sistema $\boldsymbol{F}$ è però sufficiente e poiché, comunque, la Terra ha una velocità rispetto al baricentro della CMB di 370 km/s contro c=300.000 km/s ($v^2/c^2 \sim 10^{-6}$), a molti effetti lo stesso il riferimento terrestre si potrà identificare con $\boldsymbol{F}$.

Per comprendere la natura degli argomenti conviene partire da un effetto reale, l'*effetto Cherenkov*.

Supponiamo che una particella carica si muova in un mezzo con indice di rifrazione maggiore di 1. In tale mezzo la luce si propaga con una velocità c_γ inferiore a quella nel vuoto, possiamo scrivere $c_\gamma<1$. Supposto allora, come di norma c_P=1, la condizione perché nel suo moto la particella possa emettere un fotone è che possa essere soddisfatta la legge di conservazione dell'energia

$$\sqrt{m_P^2 + \boldsymbol{p}^2} = \sqrt{m_P^2 + (\boldsymbol{p} - \boldsymbol{k})^2} + |\boldsymbol{k}| c_{\boldsymbol{\gamma}} \,,$$

cioè che per lo specifico $\boldsymbol{p}$ esista un valore di $\boldsymbol{k}$ che soddisfi l'equazione. Si può verificare che la condizione perché ciò avvenga è che sia

$$v_P = \frac{|\boldsymbol{p}|}{\sqrt{m^2 + \boldsymbol{p}^2}} > c_{\boldsymbol{\gamma}} \,,$$

cioè che la particella abbia una velocità superiore a c_γ. Questa relazione è naturalmente impossibile nel vuoto, dove si suppone c_γ=1, lo è però in un mezzo rinfrangente. Il fenomeno è ben noto e correntemente sfruttato per la rivelazione e la misura dell'energia di particelle cariche, registrando i fotoni emessi (***contatori Cherenkov***).

Supponiamo ora che si abbia anche nel vuoto $c_\gamma<1$ e che per qualche particella carica sia $c_P>c_\gamma$, dovremmo evidentemente attenderci un ***effetto Cherenkov nel vuoto***. In effetti la relazione di conservazione dell'energia per emissione di un fotone si scrive in questo caso

$$\sqrt{m_P^2 + \boldsymbol{p}^2 c_P^2} = \sqrt{m_P^2 + (\boldsymbol{p} - \boldsymbol{k})^2 \, c_P^2} + |\boldsymbol{k}| c_{\boldsymbol{\gamma}}$$

e di nuovo si trova che essa può essere soddisfatta se

$$v_P = \frac{|\boldsymbol{p}| c_{\boldsymbol{P}}^{\boldsymbol{2}}}{\sqrt{m^2 + \boldsymbol{p}^2 c_P^2}} > c_{\boldsymbol{\gamma}} \,.$$

Questa ultima relazione si può anche riesprimere in funzione dell'energia ottenendo che il processo è possibile se l'energia è superiore ad un valore di soglia, precisamente se

$$W > W_{\text{thr}} = \frac{m_P c_P}{\sqrt{c_P^2 - c_\gamma^2}}.$$

In questo caso la particella carica dovrebbe nella sua propagazione nel vuoto perdere progressivamente energia, fino a portarsi al valore W_{thr}. Nel caso di un protone p, m_p=938 MeV e, poiché c_P e c_γ devono comunque essere sempre essere pensati molto prossimi a 1, risulta

$$W_{\text{thr}} = \frac{663}{\sqrt{c_P - c_\gamma}}\ \text{MeV}.$$

Poiché nei raggi cosmici si osservano protoni di provenienza extra galattica di energia fino a $1{,}25\times10^{14}$ MeV, W_{thr}, non può essere inferiore a tale valore. Si ha quindi la limitazione

$$c_P - c_\gamma < 3\cdot10^{-23}.$$

Similmente, poiché si osservano elettroni di energia fino a 500 GeV=5×10^{11} eV, deve essere

$$c_e - c_\gamma < 5\cdot10^{-13}.$$

Inversamente, se supponiamo $c_e < c_\gamma$. Il decadimento

$$\gamma \rightarrow e^- + e^-$$

diviene possibile, se

$$pc_\gamma > 2\sqrt{m_e^2 + (\frac{p}{2})^2 c_e^2}\ ,$$

ovvero

$$W_\gamma > \frac{2m_e}{\sqrt{c_\gamma^2 - c_e^2}}\ .$$

Poiché di nuovo nella radiazione cosmica si osservano fotoni di energia fino a 20 TeV=$2\cdot10^7$ MeV, si ottiene la limitazione opposta

$$c_\gamma - c_e < 1{,}25\cdot10^{-15}.$$

Veniamo infine al caso degli ipotetici neutrini superluminali.

Poiché, per quanto ora visto, $-10^{-13}<c_\gamma-c_e<10^{-15}$, se fosse $c_\nu-c_\gamma=\frac{\delta}{2}\cong 2{,}5\times10^{-5}$, secondo l'originaria misura di OPERA, dovrebbe anche essere $c_\nu-c_e\cong 2{,}5\times10^{-5}$. Ma per $c_\nu>c_e$, se

$$W_\nu > W_{\mathrm{thr}} = \frac{2m_\nu c_\nu}{\sqrt{c_\nu^2 - c_e^2}},$$

diviene tra altri possibile il processo di tipo *debole*

$$\nu_\mu \rightarrow \nu_\mu + e^+ + e^-.$$

Anche attribuendo alla massa del neutrino uno dei valori più alti possibili come $m_\nu=0{,}5$ eV, per il valore di δ dato da OPERA si ottiene $W_{\mathrm{thr}}\cong 140$ MeV, quindi un valore molto inferiore all'energia dei neutrini del fascio (energia media=17,5 GeV).

In queste condizioni, con un calcolo della probabilità della reazione nel contesto del cosiddetto modello standard, Cohen a Glashow hanno mostrato [3] che i neutrini perderebbero rapidamente energia fino a portarsi ad un valore limite, praticamente indipendente dal valore dell'energia iniziale, dato dalla relazione

$$W_T = \left(\frac{5}{14}\frac{G_F}{192\pi^3}\delta^3 L\right)^{-\frac{1}{5}},$$

essendo $G_F=1{,}17\times10^{-5}$ GeV^{-2} la costante delle interazioni deboli ed L la distanza percorsa. Ponendo $L=730$ km e $\delta=5\times10^{-5}$, si trova $W_T=12{,}5$ GeV. OPERA, non avrebbero dovuto osservare, come invece faceva, neutrini di energia superiore a tale valore e avrebbero dovuto essere osservate coppie elettrone-positone che ICARO ha invece potuto escludere.

Combinando questi risultati con altri dello stesso tipo, tra cui alcuni di esperimenti specificamente dedicati all'osservazione dello spettro di energia di neutrini che avevano attraversato tutta la Terra (Super- Kamiokande) o spessori dell'ordine di 500 km (Ice- Cube), Cohen e Glashow sono giunti a stabilire per il valore di δ una serie coerente di limitazioni, la più forte delle quali è come già indicato $\delta<1{,}7\times10^{-11}$.

BIBLIOGRAFIA

[1] OPERA collaboration, T. ADAMS *et al.*, "Measurement of the neutrino velocity with the OPERA detector in CNGS beam", *JHEP* **1210** (2012), 093; **arXiv:1109.4897** [hep-ex].

[2] M. J LONGO, "Tests of Relativity from SN1987a", *Phys. Rev*. **D36** (1987), 3276.

[3] S. R. COLEMAN and S. L. GALSHOW, "High energy tests of special relativity", *Phys. Rev.* **D59** (1999), 116008; "Cosmic ray and neutrino tests of special relativity", *Phys. Lett.* **B59** (1997), 249-252.

[4] A.G. COHEN and S.L. GLASHOW, "New Constraints on Neutrino Velocities", *Phys.Rev.Lett.* **107** (2011), 181803; **arXiv:1109.6562v1** [hep-ph].

[5] ICARUS collaboration, M. ANTONELLO *et al.*, "The ICARUS experiment and the neutrino velocity measurement", *Nucl.Phys.Proc.Suppl.* **237-238** (2013), 190-192; "A search for the analogue to Chernkov radiation by high energy neutrinos at superluminal speeds in ICARUS", *Phys.Lett.* **B711** (2012), 270-275; "Precision measurement of the neutrino velocity with the ICARUS detector in the CNGS beam", *JHEP* **1211** (2012), 049.

[6] BOREXINO collaboration, P. ALVAREZ SANCHEZ *et al.*, "Measurement of CNGS muon neutrino speed with BOREXINO", *Phys. Lett.* **B716** (2012), 401.

[7] LVD collaboration, N. YU. AGANOVA *et al.*, "Measurement of the velocity of neutrinos from the CNGS beam with Large Volume Detector", *Phys.Rev.Lett***. 109** (2012) 070801.

[8] BOREXINO, ICARUS and LVD_collaboration (Sergio Bertolucci (CERN) for the collaboration), "Neutrino speed: a report on the $\nu\mu$ speed measurements of the BOREXINO, ICARUS and LVD experiments with the CNGS beam", *Nucl.Phys.Proc.Suppl.* **235-236** (2013), 289-295.

[9] LVD and OPERA Collaborations, N.Yu. Agafonova, "Determination of a time-shift in the OPERA set-up using high energy horizontal muons in the LVD and OPERA detectors", *Eur.Phys.J.Plus* **127** (2012), 7.

COME CATTURARE NEL TEMPO GLI EVENTI PRIMARI DI PROCESSI NATURALI: I LASER AD IMPULSI ULTRABREVI

Nota del s.c. SANDRO DE SILVESTRI e del m.e. ORAZIO SVELTO (*)

(Adunanza del 15 novembre 2012)

SUNTO. – Il presente lavoro presenta una breve rassegna delle moderne tecnologie laser che consentono di generare impulsi di luce della durata fino a "pochi cicli ottici" su ampi intervalli spettrali, che permettono di filmare l'evoluzione di processi dinamici straordinariamente brevi. In particolare verranno illustrati i risultati ottenuti dallo studio degli eventi primari di due processi importanti in natura che coinvolgono l'assorbimento della luce, quali la visione e la fotosintesi. Questi eventi avvengono in tempi molto rapidi e innescano, nei due casi, l'intero processo. Lo studio di questi eventi è importante sia per ragioni di carattere fondamentale sia al fine di progettare sistemi molecolari artificiali, di elevata efficienza, da impiegare per nuovi dispositivi nel campo della visione artificiale e del fotovoltaico.

***

ABSTRACT. – This paper presents a brief review of the modern laser technologies that enable to generate light pulses lasting "a few optical cycles" and tunable over large spectral ranges, thus allowing to follow the evolution of extraordinarily fast dynamical processes. In particular, we will describe the results obtained from the studies of the primary events of two important processes in nature involving the absorption of light, namely vision and photosynthesis. These events last for extremely short times and trigger the corresponding overall processes. The study of these events is important both for fundamental reasons and for the design of artificial molecular systems

(*) Dipartimento di Fisica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy.
E-mail: orazio.svelto@polimi.it ; sandro.desilvestri@polimi.it

of high efficiency to be used for new devices in the field of artificial vision and of photovoltaics.

1. Introduzione

L'osservazione e lo studio dell'evoluzione temporale di un processo naturale ha sempre presentato aspetti di grande interesse scientifico ed applicativo: infatti, si ha così la possibilità non solo di far luce sui relativi meccanismi elementari ma anche di intervenire e guidare l'evoluzione stessa. Tuttavia, dall'invenzione dei primi orologi meccanici, avvenuta nel XVI secolo, fino al XIX secolo, la nostra capacità investigativa era limitata a quei processi che avvenivano su scale temporali compatibili con la nostra percezione sensoriale diretta. Nel caso dell'occhio, la persistenza delle immagini sulla retina (≈ 0.1 s) impedisce, infatti, di studiare fenomeni più rapidi. A partire dal XIX secolo, diverse evoluzioni tecnologiche hanno permesso di ridurre drasticamente gli intervalli temporali misurabili.

Utilizzando diverse macchine fotografiche lungo il percorso di un cavallo al galoppo, dotate di otturatori meccanici molto rapidi (circa 1 millisecondo) e con un ingegnoso sistema di sincronizzazione, E. Muybridge riuscì nel 1878 ad analizzare le varie fasi del galoppo. Dimostrò, in questo modo, che, per un breve istante invisibile all'occhio umano, l'animale teneva tutti e quattro gli zoccoli sollevati dal suolo. Il quesito se ci fosse un determinato instante in cui un cavallo, durante il suo galoppo, tenesse le sue quattro zampe staccate dal suolo appassionava infatti parecchi esperti del settore di quel tempo. L'interesse era tale che L. Stanford, il fondatore negli anni successivi dell'Università di Stanford, istituì un grosso premio per chi fosse riuscito a risolvere questo quesito. Muybridge, con le sue ingegnose misure, riuscì ad aggiudicarselo.

La "fotografia a flash" utilizza lampi di luce molto brevi per illuminare un oggetto, consentendo di congelarne una fase del movimento: usando una serie di flash nel tempo che illuminano sequenzialmente lo stesso oggetto e combinando i singoli fotogrammi in una sorta di "moviola", è possibile ottenere una riproduzione completa del moto. Il corrispondente strumento, detto stroboscopio e inventato nel 1930, portò la risoluzione temporale ad alcuni microsecondi: risalgono a questo periodo

le famose fotografie del proiettile che attraversa una mela ad opera di H. Edgerton. Un tempo di esposizione dell'ordine del microsecondo rappresenta il limite ottenibile con le sorgenti di luce ordinarie.

Un radicale miglioramento della risoluzione temporale è stato reso possibile dall'invenzione del laser, nel 1960. Qualche anno dopo questa invenzione e con una tecnica denominata *mode-locking*, fu infatti possibile generare, da un laser, impulsi di luce di alcuni picosecondi (1 ps=10^{-12} s, il prefisso *pico* deriva dalla parola italiana *piccolo*). Attualmente, le sorgenti laser sono in grado di generare direttamente impulsi luminosi con durate sino ad alcune decine di femtosecondi (1 fs=10^{-3} ps=10^{-15} s, il prefisso *femto* deriva dal danese o norvegese *femten* che significa *quindici*).[1] Inoltre, il loro utilizzo, combinato con tecniche basate su processi ottici non lineari, ha consentito di produrre impulsi ancora più brevi, da pochi femtosecondi,[2] sino ad arrivare attualmente ad impulsi della durata di un centinaio di attosecondi (1 as=10^{-3} fs=10^{-18} s, il prefisso *atto* deriva dal danese *atten* che significa *diciotto*).[3] Tali impulsi di luce consentono di estendere oltremisura la nostra capacità di osservare in tempo reale la dinamica di processi naturali.

Esistono in natura eventi tanto rapidi da richiedere un'elevata risoluzione temporale? La risposta è affermativa per numerosi fenomeni a livello atomico e molecolare. Quando una corrente elettrica fluisce in un filo di rame, gli elettroni urtano contro gli atomi del reticolo mediamente ogni 10 fs. Per leggere questo articolo, i recettori della retina (molecole di rodopsina) ripetono per circa 10 miliardi di volte al secondo una foto-reazione il cui evento primario è concluso in soli 200 fs. Numerosi processi di rilassamento dell'energia assorbita da parte delle molecole avvengono su scale dalle decine alle centinaia di femtosecondi. I moti dei nuclei e degli elettroni all'interno di una molecola avvengono su una scale dei tempi comprese rispettivamente tra un centinaio di femtosecondi e pochi attosecondi. In termini di meccanica quantistica, si può stabilire un'analogia fra l'evoluzione temporale di una opportuna sovrapposizione coerente di stati stazionari di una data particella (es. un nucleo o un elettrone) e il moto classico di quella particella. Supponiamo ad esempio che una particella si trovi in uno stato non stazionario dato dalla sovrapposizione coerente di uno primo stato di energia E_1 e di un secondo stato di energia E_2 ($E_2 > E_1$). Tale sovrapposizione costituisce un pacchetto d'onda. Risolvendo l'equazione di Schrödinger per la funzione d'onda della particella, si ricava che la corrispondente distribuzione di probabilità nello spazio oscilla con un periodo $T = h / (E_2 - E_1)$, dove h è la costante

di Planck (h=6.63×10^{-34} J s). La variazione della posizione del centro di massa del pacchetto d'onda costituisce l'analogo quantistico del moto classico della particella in considerazione. Pertanto, quanto maggiore è la differenza in energia, $\Delta E=E_2-E_1$, fra i due stati coinvolti, tanto più veloce è il moto della particella. Ad esempio, nel caso dei moti vibrazionali degli atomi all'interno di una molecola, la differenza di energia fra i livelli vibrazionali è dell'ordine del millielettronvolt (1 eV=1.6 × 10^{-19} J) e pertanto il corrispondente moto avviene su scale temporali che vanno dalle decine alle centinaia di femtosecondi. Il moto degli elettroni in nanostrutture o all'interno di una molecola avviene invece su scale temporali ancora più brevi, dalle decine di femtosecondi a pochi attosecondi. Infine, dinamiche ancora più veloci avvengono all'interno del nucleo, dove le differenze tra i livelli energetici sono dell'ordine di 10^6 eV: in questo caso il tempo caratteristico per il movimento dei nucleoni all'interno del nucleo risulta dell'ordine di alcuni zeptosecondi (1 zs=10^{-3} as=10^{-21} s; il prefisso *zepto* deriva dal francese *sept*, sette, che richiama il fatto che 10^{-21} può essere scritto come la settima potenza di un millesimo, 1/1000^7). Lo studio dei fenomeni ultrabrevi è dunque un argomento interdisciplinare, che riguarda sia le scienze di base (fisica, chimica, biologia) sia le discipline applicate (elettronica, fotonica).

Nel seguito verranno illustrati i risultati ottenuti dallo studio di due processi importanti in natura, che coinvolgono l'assorbimento della luce, come la visione e la fotosintesi, determinanti per l'esistenza del mondo animale e vegetale, rispettivamente. Tali processi sono innescati da eventi primari molto rapidi nel tempo. Questo è un modo per la natura di garantirsi la massima efficienza nello sviluppo del processo stesso in quanto la brevità dell'evento primario riduce drasticamente la competizione con altri processi dissipativi. Per poter catturare nel tempo questi eventi è necessario disporre di impulsi estremamente brevi, cioè caratterizzati da una durata di alcuni femtosecondi, e accordabili, cioè in grado di ricoprire un ampio intervallo spettrale nel visibile-ultravioletto in modo da poter entrare in risonanza con gli stati elettronici delle molecole coinvolte in questo processo. La tecnica ottica generalmente usata per studiare questi processi dinamici estremamente veloci è detta di "pompa e sonda". Un primo impulso, detto di "pompa", con il suo spettro posizionato nella regione di assorbimento del sistema coinvolto eccita il sistema stesso e dà inizio al fenomeno che si vuole studiare (reazione chimica, rilassamento energetico, cambiamento strutturale). Un secondo impulso, detto di "sonda", sincronizza-

to con il primo e opportunamente ritardato, rivela l'evoluzione del sistema. La dinamica del sistema viene misurata tramite la variazione della trasmissione dell'impulso di sonda. A ciascun ritardo tra pompa e sonda corrisponde un "fotogramma"; anche qui, combinando i singoli fotogrammi si può ottenere una sorta di film dell'evoluzione temporale del sistema. Si noti che lo studio dei processi di visione e fotosintesi è importante non solo per ragioni di carattere fondamentale ma anche per il fatto che tali studi potranno servire a progettare sistemi molecolari artificiali, di elevata efficienza, da impiegare per nuovi dispositivi nel campo della visione artificiale e del fotovoltaico.

Generazione di impulsi di luce ultrabrevi "accordabili"

Impulsi ultrabrevi di pochi cicli ottici nel visibile (corrispondenti ad una durata inferiore alla decina di femtosecondi) vengono prodotti e amplificati partendo da una sorgente laser dotata di impulsi più lunghi pari a un centinaio di femtosecondi (generalmente una laser a Titanio in zaffiro che opera alla lunghezza d'onda di 800 nm). Lo schema di principio dell'intero sistema, basato sulla combinazione dei processi di "amplificazione parametrica" e "compressione ottica",[4] è mostrato in *Fig. 1.*

Dapprima, l'impulso proveniente dalla sorgente laser viene inviato in un opportuno sistema che produce, mediante processi nonlineari, una modulazione della fase dell'impulso stesso (automodulazione di fase, *Fig. 1*). All'uscita di questo modulatore, l'impulso di luce è quindi caratterizzato da uno spettro più largo rispetto a quello dell'impulso in ingresso. Se l'allargamento è sufficientemente elevato da ricoprire l'intero spettro del visibile, questi impulsi sono detti a luce bianca poiché di questo colore appare il corrispondente fascio all'occhio umano. Tali impulsi tuttavia sono generalmente di bassa energia (decine di nanojoule) rendendo così necessaria una loro amplificazione ai fini dell'utilizzo in un successivo apparato di misura. L'impulso viene pertanto inviato ad un successivo stadio di amplificazione (amplificatore parametrico, *Fig. 1*), dove guadagna energia a spese di un intenso impulso di luce laser violetta, ottenuto dalla sorgente di partenza mediante generazione di seconda armonica. Il processo di amplificazione avviene facendo interagire i due impulsi all'interno di un opportuno cristallo, dove l'impulso violetto trasferisce energia all'impulso di luce bianca mediante un processo ottico non-lineare noto con il nome di "amplificazione parametrica": l'energia

dell'impulso di luce bianca può così aumentare in maniera notevole (anche di un fattore 10^4). Da notare inoltre che, se solo una parte dello spettro dell'impulso di luce bianca proveniente dall'automodulatore viene selezionata per il suo invio all'amplificatore, il processo di amplificazione genererà impulsi energetici solo in quella regione spettrale. Cambiando il centro dello spettro selezionato è possibile così ottenere impulsi di luce sufficientemente energetici e accordabili in frequenza.

L'impulso amplificato viene quindi inviato ad un opportuno compressore (*Fig. 1*) il quale non fa altro che convertire la modulazione di fase dell'impulso in modulazione di ampiezza. In uscita dal compressore è allora possibile ottenere impulsi di luce che, sulla base del noto teorema di Fourier, siano pari all'inverso della banda dell'impulso. Poiché questa banda era stata notevolmente allargata mediante l'automodulazione di fase, in uscita da questo sistema si ottengono impulsi notevolmente più brevi di quelli emessi dalla sorgente laser originaria e sufficientemente energetici per le applicazioni. Nel caso di compressione ottimale, si possono così ottenere impulsi con durata pari a pochi femtosecondi (corrispondenti nello spettro visibile a pochi cicli ottici dell'onda portante). Inoltre, come spiegato in precedenza, questi impulsi possono essere accordati, col loro centro-banda, in tutta la regione del visibile o anche nell'infrarosso.

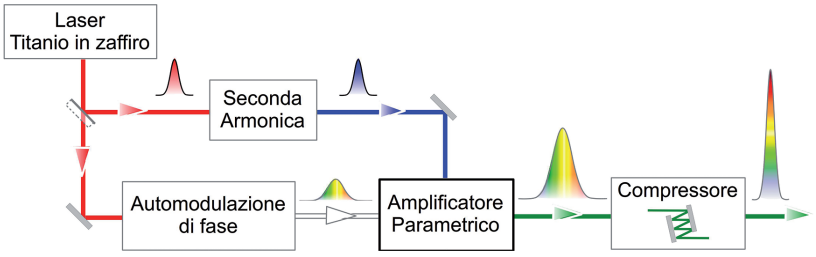


Fig. 1 – Schema di principio per la generazione di impulsi ultrabrevi e accordabili.

Il processo primario nella visione

La superficie sensibile dell'occhio, la retina, è costituita da fotorecettori (bastoncelli e coni) il cui compito è quello di trasformare in impulsi elettrici le informazioni ricevute dalle reazioni fotochimiche che vengono attivate dalla radiazione luminosa. Questi impulsi vengono

poi inviati ai neuroni retinici per una prima elaborazione del segnale visivo. Coni e bastoncelli hanno una diversa sensibilità spettrale e diverso adattamento al buio. I bastoncelli sono circa 1000 volte più sensibili alla luce dei coni e molto più numerosi (rispettivamente 120 milioni contro 6 milioni). In maniera approssimativa, si può dire che di notte vediamo con i bastoncelli e di giorno con i coni. Coni e bastoncelli contengono rispettivamente la iodopsina e la rodopsina, che sono tasche proteiche, costituite da diverse forme dell'opsina (una proteina) al cui interno è presente un cromoforo detto retinale, la cui struttura chimica è quella di un'aldeide della vitamina A, responsabile dell'assorbimento della luce. L'evento primario della visione consiste in una reazione di foto-isomerizzazione del retinale. Al buio il retinale è in conformazione 11-*cis*, in cui cioè un doppio legame in posizione n=11 della catena $[(C\text{-}H)=]_n$ che costituisce la molecola di retinale è in posizione ripiegata (*Fig. 2b*). Se ora il retinale assorbe un fotone esso può diventare retinale *tutto-trans*, in cui cioè lo stesso legame passa a una forma distesa (*trans*) (*Fig. 2d*). Questa piccola modifica strutturale del retinale comporta la scissione del complesso retinale-opsina ed è in grado di cambiare la conformazione sterica della tasca proteica della opsina. La tasca proteica così attivata dà luogo ad una serie di ulteriori reazioni che avviano la formazione e la trasmissione di un segnale elettrico. Poiché la molecola di retinale, come detto, è presente sia nei coni che nei bastoncelli, il meccanismo di attivazione è del tutto simile per entrambi. La diversa sensibilità e la diversa regione spettrale di funzionamento sono essenzialmente determinate, nei due casi, da due diverse conformazioni del legame opsina-retinale che determinano la regione spettrale di assorbimento del retinale stesso. In ambedue i casi, un enzima specifico produce poi la riconversione del retinale nell'isomero 11-*cis* e con questo la sua ricombinazione con l'opsina; il pigmento visivo viene così continuamente rigenerato.

Lo studio della dinamica della conversione del cromoforo retinale 11-*cis* alla sua forma *tutto-trans* è stato oggetto di diversi lavori sia teorici che sperimentali, in particolare nella rodopsina. L'elevata resa quantica (65%) del processo di trasformazione della rodopsina, che è la ragione dell'elevatissima sensibilità dell'occhio umano, suggeriva infatti la presenza di una fotoreazione insolitamente veloce e unidirezionale: infatti, solo con una reazione molto veloce il processo di fotoisomerizzazione poteva prevalere rispetto agli altri inevitabili processi di dissipazione dell'energia assorbita, disponibili per la stessa molecola. La presenza di una

reazione molto veloce nella variazione di conformazione del retinale è stata generalmente attribuita ad una intersezione tra la superficie di energia potenziale dello stato eccitato del retinale in conformazione 11-*cis* con la superficie di energia potenziale dello stato fondamentale della molecola nello stato *tutto-trans*. L'intersezione viene detta "intersezione conica" poiché, nello spazio tridimensionale delle superfici di potenziale, essa produce un cono. Come i "buchi neri" sono delle singolarità che mettono probabilmente in collegamento due diversi spazi del mondo interstellare, così le intersezioni coniche possono essere viste come "buchi neri della chimica", singolarità che mettono in collegamento due stati elettronici diversi della materia. Inoltre, al pari dei buchi neri interstellari, i buchi neri della chimica erano previsti solo teoricamente e risultavano di difficile osservazione sperimentale. Ottenere prove sperimentali dirette del coinvolgimento di una intersezione conica era infatti impegnativo: l'intervallo di energia tra gli stati elettronici della molecola che reagisce cambia in modo notevole su una scala temporale ultrabreve, il che richiede metodi di osservazione che combinano alta risoluzione temporale e osservazione su un'ampia finestra spettrale. Avendo a disposizione impulsi di luce ultrabrevi e accordabili dal vicino infrarosso al visibile con una risoluzione temporale di circa 10-fs è stato possibile seguire in modo completo tutta la dinamica temporale di trasformazione del retinale attraverso l'intersezione conica, dimostrando la fondatezza delle previsioni teoriche e, in particolare, una scala dei tempi insolitamente veloce.

La tecnica sperimentale di "pompa e sonda" utilizza un impulso di pompa della durata di 10 fs alla lunghezza d'onda di 500 nm, che viene assorbito dallo stato fondamentale del retinale in configurazione *cis*. Lo stato dinamico del retinale è poi monitorato mediante impulsi di sonda, opportunamente ritardati rispetto all'impulso di pompa, che coprono un intervallo spettrale 500-750 nm (7 fs) e 820-1020 nm (13 fs) rispettivamente. In *Fig. 2a* viene mostrata la reazione di isomerizzazione così come viene direttamente visualizzata dalla misura di "pompa e sonda". A seguito dell'assorbimento dell'impulso di pompa (eccitazione, fase 1), il retinale si porta sulla superficie di energia potenziale corrispondente al primo stato eccitato, dove ha inizio un processo di rilassamento verso il corrispondente minimo di energia potenziale. Tale processo è stato seguito in tempo reale analizzando il retinale eccitato con impulsi di sonda a diversi ritardi in tempo nella regione spettrale 800-1000 nm. Essendo il retinale in stato eccitato questi impulsi di

sonda provocano un processo di emissione stimolata nel retinale, con conseguente amplificazione dell'impulso di sonda stesso dopo l'attraversamento del retinale. A ritardi crescenti dell'impulso di sonda rispetto a quello di eccitazione, l'amplificazione avviene a lunghezze d'onda sempre più elevate. Questo è indice dell'esistenza di un processo di rilassamento del retinale verso il minimo della curva di energia potenziale dello stato eccitato (*Fig. 2a*). Utilizzando impulsi di sonda ancora più ritardati nel tempo si è osservato invece un assorbimento del retinale a lunghezze d'onda via via decrescenti nella regione spettrale 800-500 nm. Tale fenomeno è dovuto al fatto che, a causa dell'intersezione conica della curva che descrive lo stato eccitato del retinale 11-*cis* con quella dello stato fondamentale del retinale *all-trans*, (*Fig. 2a*, fase 2), il retinale si è ora portato in quest'ultimo stato. L'assorbimento (assorbimento fotoindotto) a lunghezze d'onda via via decrescenti dell'impulso di sonda deriva dal fatto che il retinale in configurazione *all-trans* si porta, col tempo, verso il minimo di potenziale del suo stato fondamentale (*Fig. 2a*, fase 3). L'intero processo di isomerizzazione, dalla fase 1 alla fase 3 si completa in 200 fs come riportato nel lavoro [5] con uno studio di immagini spettrali intervallate di un tempo incredibilmente breve (circa 10 fs) e su un ampio intervallo spettrale (500-1000 nm). Ciò ha consentito di dimostrare sperimentalmente, per la prima volta, l'esistenza di una intersezione conica e di osservare la corrispondente dinamica ultrarapida.

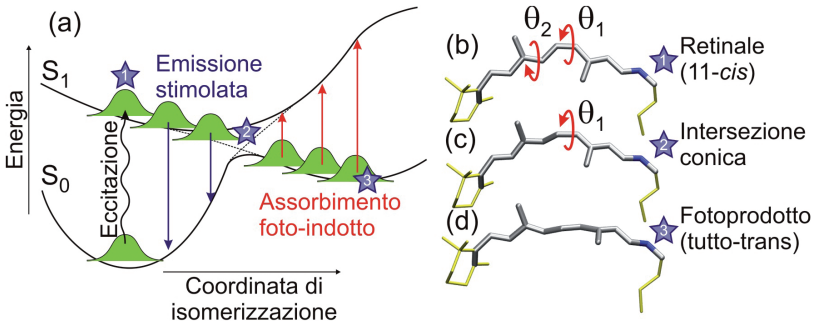


Fig. 2 – (a) Misura della dinamica di isomerizzazione del retinale lungo la coordinata di isomerizzazione: l'eccitazione è indotta dall'impulso di pompa mentre l'emissione stimolata e l'assorbimento sono indotti dagli impulsi di sonda ritardati nel tempo (fasi 1-3); (b-d) trasformazione della struttura del retinale da cis *a* tutto-trans *durante le tre fasi (θ_1 e θ_2 rappresentano gli angoli di rotazione a cui sono soggetti i legami durante la trasformazione).*

Il processo primario nella fotosintesi

In questo paragrafo viene presentato lo studio di alcuni processi ultrabrevi che coinvolgono una molecola coniugata, il beta-carotene, una catena di atomi di carbonio e idrogeno di eccezionale importanza in fisica, chimica e biologia. In generale le catene coniugate di atomi di carbonio dette carotenoidi sono presenti dovunque nel nostro universo (si trovano anche nello spazio interstellare), danno il colore a molti fiori, frutti o ortaggi comuni, e svolgono funzioni vitali nella fotosintesi delle piante e dei batteri. Nei sistemi fotosintetici la funzione primaria dei carotenoidi è di antenna per la radiazione solare. Queste molecole hanno infatti un forte assorbimento nella regione del visibile (blu-verde) e trasferiscono poi l'energia raccolta dal sole ad altre molecole (clorofille) per continuare la fotoreazione. Esse pertanto contribuiscono in maniera decisiva all'efficienza del processo di raccolta di energia solare da parte degli organismi fotosintetici. Il complesso antenna (denominato LH2) è costituita da 9 unità fotosintetiche identiche, ciascuna costituita da un carotenoide e da tre clorofille (BChl, *Fig. 3*).

Un esempio importante di sistema fotosintetico è rappresentato da alcune alghe, quali i dinoflagellati, che si trovano nelle profondità degli oceani, assorbono una rilevante frazione della luce solare ivi presente e costituiscono un importante anello della catena alimentare; tali alghe utilizzano in prevalenza i loro carotenoidi che, come detto in precedenza, assorbono la componente blu-verde della luce solare che è quella che risulta trasmessa più efficientemente dall'acqua. I complessi antenna di queste alghe sono dei sistemi modello della reazione di fotosintesi e quindi vengono molto studiati. Utilizzando uno di questi complessi antenna e la tecnica di "pompa e sonda", è stato possibile filmare l'evoluzione del trasferimento dell'energia assorbita dalle molecole di beta-carotene verso le clorofille.[6] Un primo impulso di pompa viene assorbito dal beta-carotene e porta la molecola dallo stato fondamentale S_0 allo stato eccitato S_2 (*Fig. 3*). Da qui ha inizio una serie di trasformazioni riassunte schematicamente nella citata figura: un rapidissimo passaggio dallo stato eccitato S_2 allo stato Q_x della clorofilla con una costante di tempo di circa 50 fs, un rilassamento interno verso lo stato S_1 in 100 fs seguito da un trasferimento verso lo stato eccitato Q_y della clorofilla (con un tempo di trasferimento di 3 ps) oppure, in parallelo, da un rilassamento interno del carotenoide verso il suo stato fondamentale S_0 (con un tempo di decadimento di 5 ps). Queste trasformazioni

vengono visualizzate inviando impulsi di sonda opportunamente ritardati in tempo che vengono assorbiti dagli stati eccitati del beta-carotene. Confrontando misure di "pompa e sonda" del beta-carotene in soluzione e del beta-carotene legato al complesso antenna è possibile determinare gli effetti sui tempi di vita degli stati eccitati dovuti al trasferimento di energia verso le clorofille (e ottenere così la mappa dei tempi riportata in *Fig. 3*) e valutare l'efficienza di tale trasferimento. In particolare, è stato per la prima volta evidenziato che il trasferimento dallo stato eccitato S_2 del carotenoide allo stato eccitato Q_x della clorofilla avviene in un tempo veramente molto rapido (50 fs) il che rende conto della elevata efficienza del trasferimento stesso e quindi dell'importanza della presenza del carotenoide nel sistema antenna.

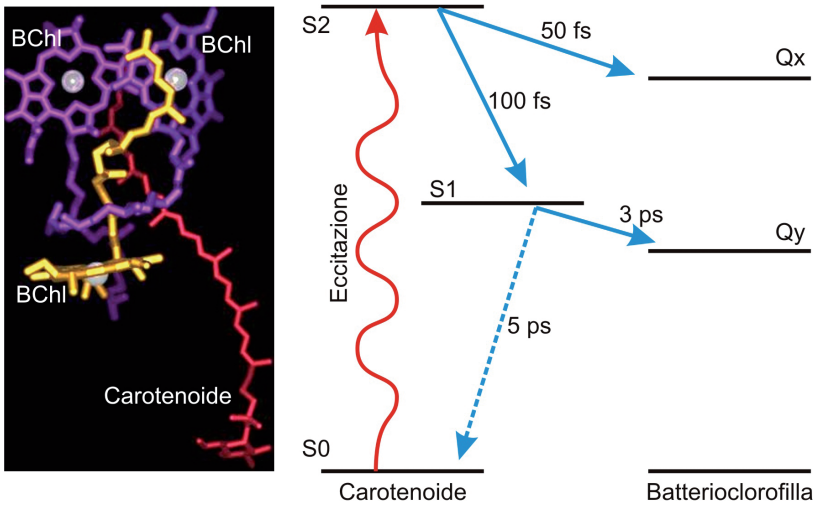


Fig. 3 – A sinistra, uno dei nove elementi del complesso antenna LH2 costituito da un carotenoide e da tre batteriocloroflille (BChl); a destra, schema dei livelli elettronici del carotenoide e della batteriocloroflilla e relativi tempi di rilassamento e trasferimento di energia.

Conclusioni

In questa breve rassegna, abbiamo visto come le moderne tecnologie ottiche consentano di generare impulsi di luce della durata fino a "pochi cicli ottici" e che possono ricoprire ampi intervalli spettrali. Questi impulsi, unitamente a particolari tecniche di misura "pompa e

sonda" consentono di filmare l'evoluzione di processi dinamici straordinariamente brevi. La conoscenza profonda dei meccanismi alla base dei sensori naturali di luce ci aiuterà, in futuro, a progettare congegni artificiali altrettanto efficienti. Le applicazioni sono molteplici, e vanno da nuovi tipi di memorie ottiche, a motori molecolari azionati dalla luce che alimentano nano-dispositivi, fino ad arrivare, in prospettiva, a congegni artificiali fotosensibili che riproducano il comportamento della retina. Ovviamente le possibili applicazioni non si limitano ai soli casi riportati, ma coprono una più vasta gamma di settori dalla fisica della materia, alla chimica e alla biologia. Si tratta del frutto di una evoluzione tecnologica che vede impegnati numerosi gruppi di ricerca in campo internazionale e che pone l'Italia tra i paesi più avanzati in questo settore. In particolare presso il Dipartimento di Fisica del Politecnico di Milano opera una infrastruttura laser, il "Centre for Ultrafast Science and Biomedical Optics" (CUSBO), riconosciuta dall'Unione Europea. Ad essa possono accedere gruppi di ricerca italiani e stranieri che intendono utilizzare la strumentazione avanzata e le tecniche sofisticate di misura, ivi disponibili e in parte descritte nel presente contributo.

BIBLIOGRAFIA

[1] O. Svelto, *"Principles of lasers"*, 5th ed., Springer (2010).

[2] M. Nisoli, S. De Silvestri, O. Svelto, R. Szipöcs, K. Ferencz, Ch. Spielmann, S. Sartania, F. Krausz, *"Compression of high-energy laser pulses below 5 fs"*, Opt. Lett. 22 (1997) 522-524.

[3] G. Sansone, E. Benedetti, F. Calegari, C. Vozzi, L. Avaldi, R. Flammini, L. Poletto, P. Villoresi, C. Altucci, R. Velotta, S. Stagira, S. De Silvestri, M. Nisoli, *"Isolated single-cycle attosecond pulses"*, Science 314 (2006) 443-446.

[4] D. Brida, C. Manzoni, G. Cirmi, M. Marangoni, S. Bonora, P. Villoresi, S. De Silvestri and G. Cerullo, *"Few-optical-cycle pulses tunable from the visible to the mid-infrared by optical parametric amplifiers"*, J. Opt. 12 (2010) 013001 (13pp).

[5] D. Polli, P. Altoè, O. Weingart, K.M. Spillame, C. Manzoni, D. Brida, G. Tomasello, G. Orlandi, P. Kukura, R.A. Mathies, M. Garavelli, G. Cerullo, *"Conical intersection dynamics of the primary photoisomerization event in vision"*, Nature 467 (2010) 440-443.

[6] D. Polli, G. Cerullo, G. Lanzani, S. De Silvestri, H. Hashimoto, R.J. Cogdell, *"Carotenoid-Bacteriochlorophyll energy transfer in LH2 complexes studied with 10-fs time resolution",* Biophys. J. 90 (2006) 2486-2497.

Istituto Lombardo (Rend. Scienze) 146, 251-277 (2012)

ON THE REPRESENTATION OF MAPS BY LIE TRANSFORMS

Nota del s.c. ANTONIO GIORGILLI (*)

(Adunanza del 22 marzo 2012)

SUNTO. Si riconsidera il problema della rappresentazione di una mappa in una forma adatta all'applicazione dei metodi di forma normale. Si mostra che ricorrendo ai metodi delle serie di Lie e delle trasformate di Lie si può costruire in modo diretto un algoritmo di normalizzazione. Si discute brevemente l'applicazione alla mappa di Schröder–Siegel e alla mappa standard di Chirikov, estendendole al caso di dimensione generica.

***

ABSTRACT. The problem of representing a class of maps in a form suited for application of normal form methods is revisited. It is shown that using the methods of Lie series and of Lie transform a normal form algorithm is constructed in a straightforward manner. The examples of the Schröder–Siegel map and of the Chirikov standard map are included, with extension to arbitrary dimension.

1. INTRODUCTION

Surface transformation as a tool for studying the flow of a system of differential equations have been introduced by Poincaré ([20], Vol. III, ch. XXXIII) and deeply investigated by Birkhoff [2]. The corresponding method of *Poincaré sections* has become classical, and has been widely used also for numerical explorations. A special interesting case is that of

(*) Dipartimento di Matematica, Via Saldini 50, 20133 Milano, Italy.
E-mail: antonio.giorgilli@unimi.it

a periodic flow, for which the corresponding surface transformation is the flow at time T of the period. A natural question is whether a given map may be represented by the flow at time T of a differential system. Positive answers to this question exist for symplectic maps that are perturbations of integrable ones (see, e.g., [17, 18, 22, 13] and the references therein).

In the present paper I will reconsider the problem of giving a suitable representation of a class of maps with a method that is somehow connected to, but does not coincide with, the interpolation by a periodic flow. I will rather use the formalism of Lie transforms, which allows us to extend to maps the techniques (e.g., calculation of normal forms) that are available for differential equations.

The basic tools exploited here are not new: a short historical account is given at the beginning of Section 2. However, most practical applications of Lie series and Lie transform methods are related to differential equations. Typical subjects are numerical integration and coordinate transformations in perturbation theory. In the latter framework, in particular, these methods have proven to be very effective also in investigating the convergence or the asymptotic properties of perturbation series involving small divisors (see, e.g., [7, 8, 9] and references therein) and in devising effective methods for perturbation expansions via algebraic manipulation on computers (see, e.g., [10] and references therein).

It is a well known fact, however, that transporting the analytical methods of normal form theory from differential equations to maps is not straightforward. A common remark is that the case of differential equations is easier to deal with, which justifies the attempts to interpolate a map with the Poincaré section of a flow.

A natural question is whether one can write a map in such a form that transporting the normal form methods that work fine for flows becomes straightforward. Answering this question in general is a major task, of course. However, if one considers a class of maps which are perturbation of integrable ones, then the question can be positively answered.

One may consider as a basic example the Schröder–Siegel problem of iteration of analytic functions, extending it to many dimensions. However, the method developed here applies also to other cases, e.g., perturbations of integrable symplectic maps as considered in the papers quoted at the beginning of this section. As interesting models one may

consider: the quadratic map investigated by Hénon [14] and its generalization in higher dimension; the twist map of an invariant annulus investigated by Poincaré [21], Birkhoff [1] and Moser [19]; the well know standard map.

In all these cases the map may be represented as a composition of two maps: an integrable one (e.g., a linear one) and a near the identity perturbation. This is a trivial well known fact, of course. However, representing the integrable map as a Lie series and the perturbation as a Lie transform allows one to implement the normal form theory as a straightforward extension of the methods used for differential equations, provided a suitable formula for the composition of Lie transforms is available. This is what I'm going to illustrate.

The paper is organized as follows. In Section 2 a short account of the methods based on Lie series and Lie transforms is given, including the representation of a near the identity map and a composition formula. In Section 3 the main proposition on the representation of a perturbation of an integrable map is proven. In Section 4 it is shown how to construct a normal form algorithm for the map. The actual construction is worked out for two typical examples. A technical appendix follows.

The theory is developed at a formal level. Some hints on quantitative applications to particular models are included at the end of Section 4.

2. Basic tools

The concept of Lie serie goes back to Sophus Lie. The use of Lie series in various problems has been widely investigated by Gröbner in a series of papers after 1957 [11]. An accurate exposition with particular emphasis on applications to numerical integration can be found in Gröbner's book [12]. The starting point, already found in Newton's work, is to express the solution of an holomorphic system of differential equations as a series expansion in time. The basic idea is to use the representation via power series as a one parameter near the identity map written in an explicit and useful form.

Lie transform may be considered as a generalization of Lie series, in a sense that will be clarified later. Its usefulness as a tool in perturbation theory has been emphasized by Hori [16] and Deprit [3], who however paid attention in particular to Hamiltonian systems. The un-

derlying idea of Deprit's work is to generate a one parameter family of near the identity coordinate transformations using the flow of a *non autonomous* system. It is quite common in the milieu of Celestial Mechanics to call *Lie Transform* the algorithms proposed by Hori and Deprit, thus reserving the name *Lie series* to the algorithm based on the flow of an *autonomous* system; thus I follow the tradition.

Actually, several explicit algorithms for Lie transform have been proposed by many authors. A list of references may be found in Henrard's paper [15]. A similar algorithm for Hamiltonian systems has been introduced on a purely algebraic basis in [5]. In this paper I will follow the latter exposition, reformulating it for vector fields. A rigorous treatment may be found in [6], which has been prompted by the need of extending the contents of Gröbner's book to a more general context.

I recall here the definitions and the properties of Lie series and Lie transform working at a formal level, including what is needed in order to develop Lie methods for maps. I will omit most of the proofs, that can be found elsewhere. Furthermore, I will forget about the origin of Lie series as solutions of a system of differential equations, thus paying particular attention to the algebraic aspect.

2.1 Lie derivatives

Let $\mathcal{D} \subset \mathbb{C}^n$ be an open domain endowed with coordinates $x = (x_1, \dots, x_n)$, and let $X(x) = (X_1, \dots, X_n)$ be a complex holomorphic vector field on $\mathcal{D}$. Let ϕ_X^t denote the time–t flow generated by X.

The Lie derivative of a holomorphic function $f(x)$ at the point x under the flow ϕ_X^t is the new function

$$L_X f = \frac{\mathrm{d}}{\mathrm{d}t}\big(\phi_X^t f\big)\Big|_{t=0}$$

where $\big(\phi_X^t f\big)(x) = \big(f \circ \phi_X^t\big)(x)$.

Similarly, the Lie derivative of a holomorphic vector field v is the new vector field

$$L_X v = \frac{\mathrm{d}}{\mathrm{d}t}\big(\phi_X^t v\big)\Big|_{t=0}$$

where $\big(\phi_X^t v\big)(x) = \big(\mathrm{d}\phi_X^t\big)^{-1}\big(v \circ \phi_X^t\big)(x)$. It is well known that one has

$$L_X v = \{X, v\} \,,$$

where $\{X, v\}$ is the commutator between the vector fields X and v.

The Lie derivative is a linear operator mapping the space of holomorphic functions (respectively holomorphic vector fields) into itself. It is also immediate to check that the property

$$L_{\alpha X + \beta Y} = \alpha L_X + \beta L_Y$$

holds true, where X, Y are vector fields and α, β are complex numbers. Further useful properties are the following. For two functions f, g the Leibniz rule applies, namely

$$\begin{aligned} L_X(fg) &= f L_X g + g L_X f \,, \\ L_X^s(fg) &= \sum_{j=0}^{s} \binom{s}{j} \big(L_X^j f\big)\,\big(L_X^{s-j} g\big), \quad s \geq 1 \,. \end{aligned}$$

For two vector fields v, w one has

$$L_X\{v, w\} = \big\{L_X v, w\big\} + \big\{v, L_X w\big\} \,.$$

Finally, denoting by $[L_X, L_Y] = L_X L_Y - L_Y L_X$ the commutator between the Lie derivatives with respect to the vector fieds X, Y one has

$$[L_X, L_Y] = L_{\{X,Y\}} \,.$$

The latter two properties are just different writings of Jacobi's identity for the commutator between vector fields, namely $\{X, \{v, w\}\} + \{v, \{w, X\}\} + \{w, \{X, v\}\} = 0$. In particular, if the vector fields X, Y do commute, namely if $\{X, Y\} = 0$ then we have $[L_X, L_Y] = 0$.

It will also be useful to write the explicit expression of the Lie derivatives in coordinates. For a function f one has

$$L_X f = \sum_{j=1}^{n} X_j \frac{\partial}{\partial x_j} f \,.$$

For a vector field one gets the expression of the commutator, namely

$$\big(L_X v\big)_j = \sum_{l=1}^{n} \left(X_l \frac{\partial v_j}{\partial x_l} - v_l \frac{\partial X_j}{\partial x_l} \right) ,$$

where the l.h.s. is the j–th component of the vector field $L_X v$.

2.2 Lie series and Lie transform

Let again $\mathcal{D} \subset \mathbb{C}^n$ be an open domain, and let $X = (X_1, \dots, X_n)$ be a holomorphic vector field. The Lie series operator is defined as

$$\exp(L_X) = \sum_{s \geq 0} \frac{1}{s!} L_X^s$$

A family of near the identity transformations depending on a parameter ε may be constructed as

$$y = \exp\big(\varepsilon L_X\big) x \tag{1}$$

or, in explicit form for the coordinates,

$$y_j = \exp(\varepsilon L_X) x_j = x_j + \varepsilon X_j(x) + \frac{\varepsilon^2}{2} L_X X_j(x) + \dots , \quad j = 1, \dots, n .$$

The Lie transform is introduced as follows. Let $X = \{X_1, X_2, \dots\}$ be a sequence of holomorphic vector fields, that I will call the *generating sequence*; here the lower index labels the element of the sequence, not the component of the field in coordinates, that will be denoted, e.g., by $X_{1,j}$. The Lie transform operator is defined as

$$T_X = \sum_{s \geq 0} E_s^X ,$$

where the sequence E_s^X of linear operators in recursively defined as

$$E_0^X = 1 , \quad E_s^X = \sum_{j=1}^{s} \frac{j}{s} L_{X_j} E_{s-j}^X .$$

The superscript in E^X is introduced in order to specify which sequence of vector fields is intended. However, I will remove it when unnecessary. By letting the sequence to have only one vector field different from zero, e.g., $X = \{0, \dots, 0, X_k, 0, \dots\}$ it is easily seen that one gets $T_X = \exp\bigl(L_{X_k}\bigr)$.

Writing $\varepsilon^s X_s$ in place of X_s a one parameter family of near the identity transformations may be defined as

$$y = T_X x \,, \tag{2}$$

i.e., in coordinates,

$$y_j = x_j + \varepsilon X_{1,j}(x) + \varepsilon^2 \left[\frac{1}{2} L_{X_1} X_{1,j}(x) + X_{2,j}(x)\right] + \dots \,, \quad j = 1, \dots, n \,.$$

Here I used the ε expansion in order to make clear the connection with classical methods based on power expansion in a small parameter. However, in many cases it is convenient to just consider the vector field X_s to be ``small of order s'' in some appropriate sense (e.g., using a norm) the order being implicit in the label of the field. Adding the powers of ε the reader will easily check that every term in the definition of the operator E^X_s carries a factor ε^s, so that E^X_s is of order s. An equivalent procedure is to find the order of E^X_s as the sum of the indices of $L_{X_j} E^X_{s-j}$, which is s indeed. In the rest of the paper I will remove the parameter, unless it has a particular meaning. If the reader gets confused, he or she may just rewrite a formula by adding the powers of ε, check that everything is put in the correct order, and then set $\varepsilon = 1$.

The Lie series and Lie transform are linear operators acting on the space of holomorphic functions and of holomorphic vector fields on the domain $\mathcal{D}$. They preserve products between functions and commutators between vector fields, i.e., if f, g are functions and v, w are vector fields then one has

$$T_X(fg) = T_X f \cdot T_X g \,, \quad T_X\{v, w\} = \bigl\{T_X v, T_X w\bigr\} \,.$$

Here, replacing T_X with $\exp\bigl(L_X\bigr)$ gives the corresponding property for Lie series. Moreover both operators are invertible. The inverse of $\exp\bigl(L_X\bigr)$ is $\exp\bigl(-L_X\bigr)$, which is a natural fact if one recalls the origin of

Lie series as a solution of an autonomous system of differential equations. The inverse of T_X takes the form

$$\big(T_X\big)^{-1} = \sum_{s\ge 0} G_j^X \ ,$$

$$G_0^X = 1 \ , \quad G_s^X = -\sum_{j=1}^{s} \frac{j}{s} G_{s-j}^X L_{X_j} \ .$$

I come now to a remarkable property which justifies the usefulness of Lie methods in perturbation theory. I will adopt the name *exchange theorem* introduced by Gröbner. Let f be a function and v be a vector field. *Consider the near the identity transformation* (2) *(or* (1) *for Lie series) and denote by* J *the differential of the transformation* (2) *(or* (1)*), namely, in coordinates, the jacobian matrix with elements* $J_{j,k} = \frac{\partial y_j}{\partial x_k}$. *Then one has*

$$f(y)\big|_{y=T_X x} = \big(T_X f\big)(x) \ , \quad \mathsf{J}^{-1} v(y)\Big|_{y=T_X x} = \big(T_X v\big)(x) \ . \qquad (3)$$

This result should be interpreted as follows. On the l.h.s. of the equalities there is the transformed function (resp. vector field) calculated via the usual method of substitution of variables. The r.h.s. is the transformed function (resp. vector field) via the Lie transform, where the variables are renamed as x. The claim is that both operations give the same result. The remarkable fact is that if one uses expansions order by order, then the r.h.s. gives it in a straightforward way for the transformed function (resp. vector field) in terms of Lie derivatives, thus requiring only operations that are easily performed, e.g., via algebraic manipulation on computers. Obtaining the same result via substitution of variables is a definitely longer process, unless one stops the expansion at very low order. This claim may appear a little obscure, but it will turn out to be immediately evident if one writes the expansions as power series in a parameter.

The proof of the identities (3) may be worked out using the algebraic properties of T_X. For, a holomorphic function (or vector field) may be expanded in power series of the variables, and by exploiting the linearity and the preservation of product the operator T_X can be moved

from the variables (the substitution) to the whole function. This justifies the name ``exchange theorem'', since the symbol of the Lie transform operator is exchanged with the symbol of the function.

2.3 Representation of a near the identity transformation

Let now a near the identity transformation be given in the form

$$y_j = x_j + \varepsilon\varphi_{1,j}(x) + \varepsilon^2\varphi_{2,j}(x) + \dots\ , \quad j = 1, \dots, n\ , \tag{4}$$

where $\varphi_{1,j}(x)$, $\varphi_{2,j}(x), \dots$ are holomorphic functions.

Proposition 2.1. *Any transformation of the form* (4) *may be written as a Lie transform of the coordinates* $y = T_X x$ *with the generating sequence*

$$X_{1,j} = \varphi_{1,j}, \quad X_{r,j} = \varphi_{r,j} - \sum_{k=1}^{r-1} \frac{k}{r} L_{X_k} E_{r-k} x_j, \quad j = 1, \dots, n, \quad r > 2\,.$$

In view of this proposition it may appear that Lie transform is more general and more attractive than Lie series, since the latter can not represent any near the identity transformation. However, a similar result may be obtained if one makes use of the *composition of Lie series*.

Let again $\{X_1, X_2, \dots\}$ be a sequence of holomorphic vector fields. Consider the sequence of transformations $\{S_X^{(0)}, S_X^{(1)}, S_X^{(2)}, \dots\}$ recursively defined as

$$S_X^{(0)} = 1\ , \quad S_X^{(r)} = \exp\bigl(L_{X_r}\bigr) \circ S_X^{(r-1)}\ . \tag{5}$$

We may well consider in formal sense the limit

$$S_X = \dots \circ \exp\bigl(L_{X_r}\bigr) \circ \dots \circ \exp\bigl(L_{X_2}\bigr) \circ \exp\bigl(L_{X_1}\bigr) \tag{6}$$

as an operator obtained by composition of Lie series. A noticeable fact, pointed out by Fassò [4], is that one has

$$S_X = \sum_{s \geq 0} : E_s^X : \tag{7}$$

where : E_s^X : denotes a reordering of E_s^X in the following sense: a composition of r Lie derivatives $L_{X_{k_1}} \circ L_{X_{k_2}} \circ \ldots \circ L_{X_{k_r}}$ is reordered as

$$: L_{X_{k_1}} \circ L_{X_{k_2}} \circ \ldots \circ L_{k_r} : \; = \; L_{X_{\sigma_1(k)}} \circ L_{X_{\sigma_2(k)}} \circ \ldots \circ L_{X_{\sigma_r(k)}}$$

where $\sigma(k) = \{\sigma_1(k), \ldots, \sigma_r(k)\}$ is any permutation of $k = \{k_1, \ldots, k_r\}$ such that $\sigma_1(k) \geq \sigma_2(k) \geq \ldots \geq \sigma_r(k)$. E.g., $: L_{X_1} L_{X_2} : = L_{X_2} L_{X_1}$.

Proposition 2.2. *Any transformation of the form* (4) *may be represented via a composition of Lie series of the form* $y = S_X x$ *with a generating sequence* $\{X_1, X_2, \ldots\}$ *that can be explicitely determined with a recursive procedure.*

As the reader will notice, an explicit expression for the vector fields is missing in the statement. Such an expression may be produced exploiting (7), but it turns out to be quite useless, because it has a non recursive form. However, a recursive procedure for determining $X_1, X_2, \ldots$ may be easily constructed by trying the first steps. Setting $X_{1,j} = \varphi_{1,j}$ one has

$$y_j - \exp\bigl(L_{X_1}\bigr) x_j = \varphi_{2,j} - \frac{1}{2} L_{X_1} X_{1,j} + \ldots$$

where the dots denote terms at least of third order. Setting $X_{2,j} = \varphi_{2,j} - \frac{1}{2} L_{X_1} X_{1,j}$ one gets that $y_j - \exp\bigl(L_{X_2}\bigr) \circ \exp\bigl(L_{X_1}\bigr) x_j$ starts with terms at least of third order, which are used in order to determine X_3, and so on. Such a procedure is easily implemented, e.g., via computer algebra, and it is in fact also the scheme of proof of the proposition. I add just a remark concerning the actual use of propositions 2.1 and 2.2. Apparently, it seems unreasonable to do extra work in order to calculate the generating sequence of the transformation. However, if one wants to transform either a function or a vector field, pushing the calculation at high orders, then using the exchange theorem turns out to be definitely more effective than performing a substitution.

2.4 Composition formulæ

I begin with a formula for the commutation between two Lie transforms. *Let* X, Y *be generating sequences. Then one has*

$$T_X \circ T_Y = T_W \circ T_X \,, \tag{8}$$

with the generating sequence $W = \{T_X Y_1, T_X Y_2, \ldots\}$. Here, the Lie trasform T_X may be replaced by $\exp\bigl(L_X\bigr)$ in case X is a vector field, and similarly for Y.

Generally speaking the formula above does not seem very interesting because the vector field W turns out to be itself a series. However, it is useful in some case, typically when one is able to put W in a manageable form by explicitly calculating the sum of the series. It will be used later, in Section 3.2.

It is more interesting to observe that since $y = T_X \circ T_Y x$ is a near the identity transformation then in view of proposition 2.1 there exists a generating sequence Z such that $y = T_Z x$. The following proposition gives an explicit expression for Z.

Proposition 2.3. *Let* X, Y *be generating sequences. Then one has* $T_X \circ T_Y = T_Z$ *where* Z *is the generating sequence recursively defined as*

$$Z_1 = X_1 + Y_1 \,, \quad Z_s = X_s + Y_s + \sum_{j=1}^{s-1} \frac{j}{s} E^X_{s-j} Y_j \,. \tag{9}$$

The proof requires a long sequence of tedious calculations that can hardly be found in previous papers. Thus I include it in appendix A. Composition formulæ for any combination of Lie series and Lie transforms are easily obtained by suitably elaborating formula (9). E.g., if X, Y are vector fields then one has $\exp\bigl(L_X\bigr) \circ \exp\bigl(L_Y\bigr) = T_W$, where $W = \{W_1, W_2, \ldots\}$ is the generating sequence

$$W_1 = X + Y \,, \quad W_s = \frac{1}{s!} L_X^{s-1} Y \,.$$

The latter formula reminds the well known Baker–Campbell–Hausdorff composition of exponentials. The difference is that the result is expressed as a Lie transform instead of an exponential.

3. Representation of a map by a composition of Lie transforms

I shall consider two cases. The first one is a map in a neighbourhood of an equilibrium, which may be expanded in Taylor series as

$$z' = \Lambda z + v_1(z) + v_2(z) + \ldots \,, \quad z \in \mathbb{C}^n \tag{10}$$

where Λ is a $n \times n$ complex matrix and $v_s(z)$ is a homogeneous polynomial af degree $s+1$. The second example is a real analytic map

$$\begin{aligned} \varphi' &= \varphi + \omega(I) + \varepsilon f_1(\varphi, I) + \varepsilon^2 f_2(\varphi, I) + \dots \\ I' &= I + \varepsilon g_1(\varphi, I) + \varepsilon^2 g_2(\varphi, I) + \dots \end{aligned} \tag{11}$$

where $(\varphi, I) \in \mathbb{T}^n \times \mathcal{G}$, with $\mathcal{G} \subset \mathbb{R}^m$ open, and ε is a small perturbation parameter.

3.1 The unperturbed map

By unperturbed map I mean here (as usual) either the linear part of (10) or the map (11) with $\varepsilon = 0$, which is a Kronecker map on a family of invariant tori parameterized by the actions I.

Let me start with the linear part $z' = \Lambda z$ of the map (10). Let $\Lambda = e^{\mathsf{A}}$ with a complex $n \times n$ matrix A. For the vector field $\mathsf{A}z$ one has $L_{\mathsf{A}z} z = \mathsf{A}z\,, \dots,\ L^s_{\mathsf{A}z} z = \mathsf{A}^s z$, and so also $\exp\bigl(L_{\mathsf{A}z}\bigr) z = \sum_{s\geq 0} \frac{1}{s!} \mathsf{A}^s z = e^{\mathsf{A}} z$. Using the exchange theorem we may transform a function $f(z')$ and a vector field $v(z')$ as

$$\begin{aligned} \exp\bigl(L_{\mathsf{A}z}\bigr) f(z) &= f(z')\Big|_{z'=e^{\mathsf{A}}z} &&= f(\Lambda z)\,, \\ \exp\bigl(L_{\mathsf{A}z}\bigr) v(z) &= e^{-\mathsf{A}} v(z')\Big|_{z'=\exp(L_{\mathsf{A}z})z} &&= \Lambda^{-1} v(\Lambda z)\,. \end{aligned} \tag{12}$$

The simplest case occurs when $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix.

I point out that here the exchange theorem is used only in order to represent a transformation of a function and of a vector field as the action of a Lie series operator. Of course, in this case everybody would perform the transformations via a direct substitution as in the last member of the formulæ above, since this is actually the sum of the Lie series in closed form and requires no further expansion.

I come now to considering the map (11), that for $\varepsilon = 0$ writes

$$\varphi' = \varphi + \omega(I)\,, \quad I' = I.$$

Introducing the $(n+m)$–dimensional vector field $\Omega = \big(\omega(I), 0\big)$ one immediately gets

$$L_\Omega(\varphi, I) = \big(\omega(I), 0\big) \,, \quad L^s_\Omega\big(\omega(I), 0\big) = (0, 0) \quad \text{for } s > 1 \,.$$

Thus, the map may be rewritten as $(\varphi', I') = \exp\big(L_\Omega\big)(\varphi, I)$. Using again the exchange theorem as above we may transform a function $f(\varphi, I)$ or a $(n+m)$–dimensional vector field $v(\varphi, I)$ as

$$\begin{aligned} \exp\big(L_\Omega\big) f(\varphi, I) &= f(\varphi', I')\Big|_{(\varphi', I') = \exp(L_\Omega)(\varphi, I)} = f\big(\varphi + \omega(I), I\big) \,, \\ \exp\big(L_\Omega\big) v(\varphi, I) &= \mathsf{J}^{-1} v(\varphi', I')\Big|_{(\varphi', I') = \exp(L_\Omega)(\varphi, I)} \\ &= \mathsf{J}^{-1} v\big(\varphi + \omega(I), I\big) \,, \end{aligned} \tag{13}$$

where J and its inverse J^{-1} are the jacobian block matrices

$$\mathsf{J} = \begin{pmatrix} \mathsf{I}_n & \mathsf{B} \\ 0 & \mathsf{I}_m \end{pmatrix} \,, \quad \mathsf{J}^{-1} = \begin{pmatrix} \mathsf{I}_n & -\mathsf{B} \\ 0 & \mathsf{I}_m \end{pmatrix} \,, \quad \mathsf{B} = \left\{ \frac{\partial \omega_j}{\partial I_l} \right\}_{1 \le j \le n,\, 1 \le l \le m} \,,$$

I_s denoting the $s \times s$ identity matrix.

3.2 The full map

The maps (10) or (11) considered in the last section are just interesting examples. In more general terms one can consider an unperturbed map $x' = f_0(x)$ which has some nice properties and can be represented as a Lie transform $x' = \exp\big(L_X\big)$ with an appropriately defined vector field X. In view of the particular role of the latter map I will denote $\mathsf{R} = \exp\big(L_X\big)$.

Then one considers a perturbed map

$$x' = \mathsf{R}x + f_1(x) + f_2(x) + \dots \tag{14}$$

where $f_s(x)$ is of order s in some reasonable sense. E.g., $f_s(x)$ should be replaced by $v_s(z)$ in (10), and by $\varepsilon f_s(\varphi, I)$ in (11), with the obvious change of the symbols for the coordinates.

Proposition 3.1. *Consider the map* (14) *where* R *is a Lie series operator. Then there exist generating sequences* $V(x) = \{V_1(x), V_2(x), \ldots\}$ *and* $W(x) = \{W_1(x), W_2(x), \ldots\}$ *of vector fields with* $W_s = \mathsf{R}V_s$ *such that one has both*

$$x' = \mathsf{R} \circ T_V x \quad \textit{and} \quad x' = T_W \circ \mathsf{R}\, x \tag{15}$$

Proof. Using the linearity of the Lie series operator R rewrite the map (15) as

$$x' = \mathsf{R}\big(x + \mathsf{R}^{-1}\,(f_1 + f_2 + \ldots)\big)\,.$$

This is a representation of the map as the composition of two operations, namely

$$\widetilde{\varphi}(x) = x + \mathsf{R}^{-1}\,(f_1 + f_2 + \ldots)\,, \quad x' = \mathsf{R}\widetilde{\varphi}(x)\,.$$

In view of proposition 2.1 one may determine the generating sequence $V(x) = \{V_1(x), V_2(x), \ldots\}$ such that $\tilde{\varphi}(x) = T_V x$. Applying R to both members and using linearity one readily gets $x' = \mathsf{R} \circ T_V x$, namely the first of (15). Using the identity (8) of proposition 2.3, which clearly applies also to the Lie series operator R, one gets $\mathsf{R} \circ T_V = T_W \circ \mathsf{R}$ with W as in the statement, which gives the second of (15). ∎

3.3 Conjugating maps

I come now to the following question. *Let two maps*

$$x' = T_W \circ \mathsf{R}\, x\,, \quad y' = T_Z \circ \mathsf{R}\, y \tag{16}$$

be given, where R *is an invertible Lie series operator and* $W = \{W_1, W_2, \ldots\}$, $Z = \{Z_1, Z_2, \ldots\}$ *are generating sequences. To find whether the maps are conjugated by a holomorphic near the identity transformation*

$$y = x + \varphi_1(x) + \varphi_2(x) + \ldots\,. \tag{17}$$

Using the same operator R in both maps means only that the unperturbed maps are trivially conjugated. In view of propositions 2.1 and 2.2 it is natural to consider the transformation (17) as generated by either a Lie transform or a Lie series.

Proposition 3.2. *Let $X = \{X_1, X_2, \ldots\}$ be a generating sequence of the near the identity transformation $y = T_X x$. Then the maps* (16) *are conjugated if*

$$T_W \circ T_{\mathsf{R}X} = T_X \circ T_Z \ . \tag{18}$$

More explicitly, the following relations must be satisfied:

$$\begin{aligned} &\mathsf{D}X_1 = Z_1 - W_1 \ , \quad \mathsf{D} = \mathsf{R} - 1 \\ &\mathsf{D}X_s = Z_s - W_s + \sum_{j=1}^{s-1} \frac{j}{s} \big(E^X_{s-j} Z_j - E^W_{s-j} \mathsf{R} X_j \big) \ , \quad s > 1 \ . \end{aligned} \tag{19}$$

Proof. In the l.h.s. of the map $y' = T_Z \circ \mathsf{R}\, y$ set $y' = T_X x'$ and then substitute $x' = T_W \circ \mathsf{R}\, x$. This must be the same as substituting $y = T_X x$ in the r.h.s. of the map. Thus the identity

$$T_X x' \Big|_{x' = T_W \circ \mathsf{R}\, x} = T_Z \circ \mathsf{R}\, y \Big|_{y = T_X x}$$

must be true, and by the exchange theorem this gives $T_W \circ \mathsf{R} \circ T_X = T_X \circ T_Z \circ \mathsf{R}$. In view of (8) one has $\mathsf{R} \circ T_X = T_{\mathsf{R}X} \circ \mathsf{R}$, so that (18) is readily found in view of the invertibility of R. Apply now proposition 2.3 to both members of the latter equality. By the first of (9) we get $W_1 + \mathsf{R}X_1 = X_1 + Z_1$, namely the first of (19). For $s > 1$ we get

$$W_s + \mathsf{R}X_s + \sum_{j=1}^{s-1} \frac{j}{s} E^W_{s-j} \mathsf{R} X_j = X_s + Z_s + \sum_{j=1}^{s-1} \frac{j}{s} E^X_{s-j} Z_j \ ,$$

from which the second of (19) readily follows. ∎

A similar result for Lie series holds true if in (18) one replaces the Lie transform T_X with the infinite composition of Lie series S_X as given by (6). However, (19) must be restated in a more elaborated manner, proceeding step by step. Let me say that the maps (18) are conjugated up to order r in case there exists a finite generating sequence $X = \{X_1, \ldots, X_r\}$ such that the transformation $y = S^{(r)} x$ makes the difference between the maps to be of order higher than r, i.e.,

$$S^{(r)}_X x' \Big|_{x' = T_W \circ \mathsf{R}\, x} - T_Z \circ \mathsf{R}\, y \Big|_{y = S^{(r)}_X x} = \mathcal{O}(r+1) \ ,$$

where $S^{(r)}$ is defined by (5). The maps are trivially conjugated up to order r if the generating sequences W, Z coincide up to order r, i.e., $W_1 = Z_1, \dots, W_r = Z_r$. For this implies $T_W x - T_Z x = O(r+1)$.

Proposition 3.3. *Let the generating sequences of the maps* (16) *coincide up to order* $r-1$ *and let* X_r *be a vector field of order* r *generating the near the identity transformation* $y = \exp\bigl(L_{X_r} x\bigr)$. *Then the maps are conjugated up to order* r *if*

$$T_W \circ \exp\bigl(L_{\mathsf{R}X_r}\bigr) = \exp\bigl(L_{X_r}\bigr) \circ T_Z \ . \tag{20}$$

More explicitly, the following relations must be satisfied:

$$\begin{aligned} &\mathsf{D}X_r = Z_r - W_r \ , \quad \mathsf{D} = \mathsf{R} - 1 \ ; \\ &Z_s - W_s = \frac{r}{s} E^W_{s-r} \mathsf{R} X_r - \sum_{j=1}^{\lfloor (s-1)/r \rfloor} \frac{s-jr}{s \cdot j!} L^j_{X_r} Z_{s-jr} \qquad \text{for } s > r \ . \end{aligned} \tag{21}$$

The proof of (20) is a straightforward adaptation of that of proposition 3.2 using a generating sequence X with all elements zero except X_r. For, in this case one has $T_X = \exp\bigl(L_{X_r}\bigr)$, so that $E^X_s = 0$ if s is not a multiple of r. Removing from (19) all vanishing term, with some patience one obtains the second of (21).

4. Normal form algorithm

A standard and useful tool in perturbation theory is the construction of a normal form for either a map or a system of differential equations. This is a classical and widely investigated problem, so I will limit the discussion to indicating how the known methods may be revisited within the scheme of representation of maps presented in this paper. Constructing a normalization algorithm for maps is indeed an easy matter in view of the results of Section 3.3. Actually, two different algorithms may be devised, the first one based on Lie transform, the second one based on composition of Lie series.

4.1 A general formulation

Let me start with the Lie transform. The key point is that formula (18) must be considered as an equation for the generating sequence X, which

must be so determined that the transformed sequence Z has some nice property that characterizes it as being in normal form. Look now at equations (19), which are just a rewriting of (18) order by order. The reader will immediately see that for every $s > 1$ one has to solve recursively an equation of the form

$$\mathsf{D}X_s + Z_s = \Psi_s \tag{22}$$

where Ψ_s is known, since it is determined by $X_1, \ldots, X_{s-1}$ and $W_1, \ldots, W_{s-1}$, which are known. Thus, the problem is only that the prescription that Z has a normal form should be imposed so that eq. (22) may be solved for X_s and Z_s. This is the standard problem in normal form theory for both maps and differential equations. If the process can be worked out, at least formally, then the generating sequence X produces a coordinate transformation $y = T_X x$, with inverse $x = T_X^{-1} y$, such that the map in the new coordinates writes $y' = T_Z \circ \mathsf{R}\, y$.

The algorithm based on composition of Lie series appears to be more elaborated, since it requires using proposition 3.3 as an iteration step. Precisely, one constructs an infinite sequence $\{W^{(r)}\}_{r\geq 0}$ of generating sequences, with $W^{(0)} = W$, and a generating sequence $X = \{X_1, X_2, \ldots\}$ such that for every $r > 0$ the generating sequences $W^{(r-1)}$ and $W^{(r)}$ coincide up to order $r-1$, and $Z_1 = W_1^{(r)}, \ldots, Z_r = W_r^{(r)}$ are in normal form and do not change with the next iteration. Proposition 3.3 shows how to determine X_r and $W_r^{(r)} = Z_r$ by solving the first of (21) with $W_r^{(r-1)}$ in place of W_r. Then the whole generating sequence $W^{(r)}$ is constructed as given by the second of (21), putting $W^{(r-1)}$ in place of W and $W^{(r)}$ in place of Z. The normal form is thus determined step by step.

At first sight, the reader may think that this is a too complicated process. However, by implementing the algorithm using algebraic manipulation he or she will realize that there is no substantial increase of complexity with respect to the algorithm based on a single Lie transform, and that in some cases the composition of Lie series may even be more effective.

4.2 Back to examples

Let me illustrate how a normal form may be constructed for the examples (10) and (11). Actually, this means that I should explain how to characterize the normal form and how to solve eq. (22).

Assume that the matrix Λ in (10) has a diagonal form, namely $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, and write $\lambda_j = e^{\mu_j + i\omega_j}$. This is a n–dimensional version of the problem of iteration of analytic maps investigated by Schröder [23], who gave the formal solution for the case $n = 1$. Clearly one has

$$\mathsf{R} = \exp\bigl(L_{\Omega x}\bigr) \,, \quad \Omega = \mathrm{diag}(\mu_1 + i\omega_1, \ldots, \mu_n + i\omega_n) \,,$$

and applying the method of proposition 3.1 the generating sequence W may be determined so that W_s is a homogeneous polynomial of degree $s+1$. Following Schröder one tries to conjugate the map to its linear part. This means that the normal form of the map should be $\zeta' = \Lambda\zeta$, which in the Lie transform representation means that one wants $Z = \{0, 0, \ldots\}$, the null sequence. Thus, according to (22), the generating sequence is determined by solving for X_s the equation

$$\mathsf{D} X_s = \Psi_s \,, \quad \mathsf{D} = \exp\bigl(L_{\Omega x}\bigr) - 1 \,, \tag{23}$$

where $\Psi_s(x_1, \ldots, x_n)$ is a homogeneous polynomial of degree $s + 1$. The relevant property is that D is diagonal on the basis of monomials $x^k \mathsf{e}_j = x_1^{k_1} \cdot \ldots \cdot x_n^{k_n} \mathsf{e}_j$, where $(\mathsf{e}_1, \ldots, \mathsf{e}_n)$ is the canonical basis of $\mathbb{C}^n$. For, in view of the second of (12) one has

$$\mathsf{D}\, x^k \mathsf{e}_j = \bigl(e^{\langle k, \mu + i\omega\rangle - \mu_j - i\omega_j} - 1\bigr)\, x^k \mathsf{e}_j \,.$$

Determining the vector field X_s is an easy matter if none of the eigenvalues of D is zero. For, writing the homogeneous polynomial vector field as

$$\Psi_s = \sum_{j=1}^{n} \mathsf{e}_j \sum_{|k|=s+1} \psi_{j,k} x^k$$

the solution of (23) is readily found to be

$$X_s = \sum_{j=1}^{n} \mathsf{e}_j \sum_{k} \frac{\psi_{j,k}}{e^{\langle k, \mu + i\omega\rangle - \mu_j - i\omega_j} - 1} x^k \,.$$

Thus the map may be formally linearized if the *nonresonance condition*

$$e^{\langle k, \mu + i\omega\rangle - \mu_j - i\omega_j} \neq 1 \text{ for } k \in \mathbb{Z}_+^n \,, |k| > 1 \text{ and } j = 1, \ldots, n$$

is fulfilled. In the case $n = 1$ considered by Schröder a resonance may occur only if λ is a root of the unity.

If the resonance condition is violated then a different definition of normal form must be introduced. Precisely, the space $\mathcal{P}^s$ of homogeneous polynomials of any degree s splits into two complementary subspaces

$$\mathcal{N}^s = \mathsf{D}^{-1}\{0\} , \quad \mathcal{R}^s = \mathsf{D}(\mathcal{P}^s) ,$$

namely the kernel of D and the image of $\mathcal{P}^s$ through D. For, D maps $\mathcal{P}^s$ onto itself, so that both $\mathcal{N}^s$ and $\mathcal{R}^s$ are subspaces of $\mathcal{P}^s$, and D is diagonal. Then the operator D may be uniquely inverted on $\mathcal{R}^s$. Thus eq. (22) may be solved by splitting $\Psi_s = \Psi_s^{\mathcal{N}} + \Psi_s^{\mathcal{R}}$, with obvious meaning of the superscripts, and setting

$$Z_s = \Psi_s^{\mathcal{N}} , \qquad X_s = \mathsf{D}^{-1}\Psi_s^{\mathcal{R}} , \quad X_s \in \mathcal{R}^s .$$

The latter condition makes the solution unique. The procedure thus described is a standard one in normal form theory. Different solutions may be considered, of course, depending on what one is looking for.

The convergence of the tranformation to normal form in the non resonant case has been proved by Siegel [24] for the case $n = 1$ under the additional hypothesis that λ satisfies a diophantine condition. The work of Siegel represents a milestone for the problem of convergence of perturbation series with small divisors.

Let me now come to the model (11). In view of the particular form of the map it is convenient to represent the vector fields separating, so to say, the φ component from the I component by writing $\begin{pmatrix} X \\ Y \end{pmatrix}$ in place of X, where $X(\varphi, I)$ and $Y(\varphi, I)$ are a n–dimensional and a m–dimensional vector function, respectively. The explicit form of the commutator is written as

$$\begin{aligned} &\left\{ \begin{pmatrix} X \\ Y \end{pmatrix}, \begin{pmatrix} V \\ W \end{pmatrix} \right\} \\ &\quad = \begin{pmatrix} \displaystyle\sum_{l=1}^{n} \left(X_l \frac{\partial V}{\partial \varphi_l} - V_l \frac{\partial X}{\partial \varphi_l} \right) + \sum_{l=1}^{m} \left(Y_l \frac{\partial V}{\partial I_l} - W_l \frac{\partial X}{\partial I_l} \right) \\ \displaystyle\sum_{l=1}^{n} \left(X_l \frac{\partial W}{\partial \varphi_l} - V_l \frac{\partial Y}{\partial \varphi_l} \right) + \sum_{l=1}^{m} \left(Y_l \frac{\partial W}{\partial I_l} - W_l \frac{\partial Y}{\partial I_l} \right) \end{pmatrix} \end{aligned} \tag{24}$$

Recall also that in this case with the notation above one has

$$\mathsf{D} = \exp\bigl(L_\Omega\bigr) - 1\,, \quad \Omega = \begin{pmatrix} \omega(I) \\ 0 \end{pmatrix}\,.$$

In the r.h.s. of (22) we may expand $\Psi_s(\varphi, I)$ in Fourier series as

$$\Psi_s = \begin{pmatrix} \sum_{k\in\mathbb{Z}^m} \alpha_k(I) e^{i\langle k,\varphi\rangle} \\ \sum_{k\in\mathbb{Z}^m} \beta_k(I) e^{i\langle k,\varphi\rangle} \end{pmatrix}$$

with known coefficients $\alpha_k(I)$ and $\beta_k(I)$. Use a similar expansion for

$$X_s = \begin{pmatrix} \sum_{k\in\mathbb{Z}^m} c_k(I) e^{i\langle k,\varphi\rangle} \\ \sum_{k\in\mathbb{Z}^m} d_k(I) e^{i\langle k,\varphi\rangle} \end{pmatrix}\,. \tag{25}$$

In view of (13) the action of the operator D is given by

$$\mathsf{D}X_s = \begin{pmatrix} \sum_{k\in\mathbb{Z}^n} \bigl(e^{i\langle k,\omega(I)\rangle} - 1\bigr) c_k(I) e^{i\langle k,\varphi\rangle} - \sum_{k\in\mathbb{Z}^n} e^{i\langle k,\omega(I)\rangle} \mathsf{B} d_k(I) e^{i\langle k,\varphi\rangle} \\ \sum_{k\in\mathbb{Z}^n} \bigl(e^{i\langle k,\omega(I)\rangle} - 1\bigr) d_k(I) e^{i\langle k,\varphi\rangle} \end{pmatrix}.$$

Thus, one would be tempted to solve eq. (22) by setting

$$Z_s = \begin{pmatrix} \alpha_0(I) \\ \beta_0(I) \end{pmatrix}\,,$$

$$X_s = \begin{pmatrix} \sum_{k\in\mathbb{Z}^n\setminus\{0\}} \left[\dfrac{\alpha_k(I)}{e^{i\langle k,\omega(I)\rangle} - 1} + \dfrac{e^{i\langle k,\omega(I)\rangle} \mathsf{B}\beta_k(I)}{(e^{i\langle k,\omega(I)\rangle} - 1)^2} \right] e^{i\langle k,\varphi\rangle} \\ \sum_{k\in\mathbb{Z}^n\setminus\{0\}} \dfrac{\beta_k(I)}{e^{i\langle k,\omega(I)\rangle} - 1} e^{i\langle k,\varphi\rangle} \end{pmatrix}.$$

However, some denominator could vanish at some point $I \in \mathcal{G}$, the action's domain, or at least become very small. This is indeed the classical problem of small divisors in Celestial Mechanics, which was well known, e.g., to Lagrange and Laplace.

Assume for a moment that no divisor actually vanishes. This is true, e.g., if $\omega \in \mathbb{R}^n$ is a constant vector and the non resonance condition $e^{i\langle k,\omega\rangle} \neq 1$ for $k \in \mathbb{Z}^m \setminus \{0\}$ is satisfied. Then the generating sequence $Z(I)$ is independent of the angle variables φ. Thus one has

$$L_{Z_s(I)}\varphi = Z_s(I)\ , \quad L^r_{Z_s}\varphi = 0 \text{ for } r > 1\ ,$$

and the map in normal form is written as

$$\varphi' = \varphi + \omega'(I)\ , \quad I' = I + g'(I)$$

where $\omega'(I)$ and $g'(I)$ are determined via the normalization process. The dynamics of the actions I is separated from that of the angles φ.

All this is formal, of course. Making a rigorous statement, as is known, is a definitely more complicated matter. If the angles $\omega(I)$ of the unperturbed map do depend on the actions I then the formal construction above is expected to fail due to the presence of zero divisors, and an analog of Poincaré's theorem on non existence of holomorphic first integrals for Hamiltonian systems of differential equations applies. However one can prove, possibly with some extra condition on $\omega(I)$, that KAM theory applies, thus showing the existence of a big set on invariant tori carrying a Kronecker map with strongly non resonant angles. On the other hand, it should also be possible to prove a theorem of Nekhoroshev's type on exponential stability. If the angles ω are constant, then some normal form may be constructed, possibly taking into account the resonances, but the series so constructed are expected to be divergent, unless one looks for a Kolmogorov's normal form on an invariant torus. However, all this matter goes behind the limits of the present note, which deals only with formal aspects.

Appendix A: Proof of proposition 2.3

Let me state a preliminary identity. If X is a generating sequence and V a vector field then one has

$$E^X_s L_V = \sum_{j=0}^{s} L_{E^X_j V} E^X_{s-j} \ . \tag{26}$$

The proof is worked out by induction, since the equality is trivial for $s = 0$ and moreover for $s = 1$ it is just Jacobi's identity for commutators. Here is the complete calculation for $s > 1$. The notation is made simpler by writing E_j in place of E^X_j, since there is no confusion.

$$\begin{aligned}
E_s L_V &= \sum_{m=1}^{s} \frac{m}{s} L_{X_m} E_{s-m} L_V \\
&= \sum_{m=1}^{s} \frac{m}{s} L_{X_m} \sum_{j=0}^{s-m} L_{E_j V} E_{s-m-j} \\
&= \sum_{m=1}^{s} \sum_{j=0}^{s-m} \frac{m}{s} \Big(L_{L_{X_m} E_j V} + L_{E_j V} L_{X_m} \Big) E_{s-m-j} \\
&= \sum_{m=1}^{s} \sum_{j=0}^{s-m} \frac{m}{s} \Big(L_{L_{X_m} E_{s-j-m} V} E_j + L_{E_j V} L_{X_m} E_{s-j-m} \Big) \\
&= \sum_{j=0}^{s-1} \frac{s-j}{s} \sum_{m=1}^{s-j} \frac{m}{s-j} \Big(L_{L_{X_m} E_{s-j-m} V} E_j + L_{E_j V} L_{X_m} E_{s-j-m} \Big) \\
&= \sum_{j=0}^{s-1} \frac{s-j}{s} \Big(L_{E_{s-j} V} E_j + L_{E_j V} E_{s-j} \Big) \\
&= \sum_{j=1}^{s} \frac{j}{s} L_{E_j V} E_{s-j} + \sum_{j=0}^{s-1} \frac{s-j}{s} L_{E_j V} E_{s-j} = \sum_{j=0}^{s} L_{E_j V} E_{s-j} \ .
\end{aligned}$$

Jacobi's identity for commutators written as $L_X L_w - L_w L_X = L_{L_X w}$ is used in order to obtain the third equality.

Coming to the proof of (9) of proposition 2.3, by definition of Lie transform one has

$$T_X \circ T_Y = \left(\sum_{l\geq 0} E_l^X\right)\left(\sum_{k\geq 0} E_k^Y\right) = \sum_{s\geq 0}\sum_{m=0}^{s} E_m^X \ E_{s-m}^Y \,.$$

On the other hand for the generating sequence Z defined as in (9) one has

$$E_s^Z = \sum_{l=1}^{s} \frac{l}{s} L_{X_l+Y_l} E_{s-l}^Z + \sum_{l=2}^{s} \frac{l}{s} \sum_{m=1}^{l-1} \frac{m}{l} L_{E_{l-m}^X Y_m} E_{s-l}^Z \,. \tag{27}$$

Thus it is enough to check that

$$E_s^Z = \sum_{m=0}^{s} E_m^X \ E_{s-m}^Y \,, \quad s \geq 0 \,. \tag{28}$$

I proceed by induction. For $s = 0, 1$ the equality is true. For $s > 1$ calculate

$$\begin{aligned}
&\sum_{m=0}^{s} E_m^X \ E_{s-m}^Y \\
&= \sum_{m=0}^{s-1} \frac{s-m}{s} \left(E_{s-m}^X E_m^Y \ + E_m^X \ E_{s-m}^Y\right) \\
&= \sum_{m=0}^{s-1} \frac{s-m}{s} \sum_{l=1}^{s-m} \frac{l}{s-m} \left(L_{X_l} E_{s-m-l}^X E_m^Y \ + E_m^X \ L_{Y_l} E_{s-m-l}^Y\right) \\
&= \sum_{m=0}^{s-1} \sum_{l=1}^{s-m} \frac{l}{s} \left(L_{X_l} E_{s-m-l}^X E_m^Y \ + \sum_{k=0}^{m} L_{E_k^X Y_l} E_{m-k}^X E_{s-m-l}^Y\right) \\
&= \sum_{l=1}^{s} \sum_{m=0}^{s-l} \frac{l}{s} \left(L_{X_l} E_{s-m-l}^X E_m^Y \ + L_{Y_l} E_m^X \ E_{s-m-l}^Y\right) \\
&\quad + \sum_{l=1}^{s-1} \sum_{m=1}^{s-l} \sum_{k=1}^{m} \frac{l}{s} L_{E_k^X Y_l} E_{m-k}^X E_{s-m-l}^Y \,.
\end{aligned}$$

The identity (26) is used in order to obtain the third line. The first double sum in the latter expression is further elaborated as

$$\sum_{l=1}^{s} \frac{l}{s} \left(L_{X_l} \sum_{m=0}^{s-l} E^X_{s-l-m} E^Y_m + L_{Y_l} \sum_{m=0}^{s-l} E^X_m E^Y_{s-l-m} \right)$$
$$= \sum_{l=1}^{s} \frac{l}{s} L_{X_l+Y_l} \sum_{m=0}^{s-l} E^X_m E^Y_{s-l-m} = \sum_{l=1}^{s} \frac{l}{s} L_{X_l+Y_l} E^Z_{s-l} \,.$$

The latter expression coincides with the first sum in the r.h.s. of (27). The triple sum is further elaborated as

$$\sum_{l=1}^{s-1} \sum_{k=1}^{s-l} \frac{l}{s} L_{E^X_k Y_l} \sum_{h=0}^{s-l-k} E^X_h E^Y_{s-l-k-h} = \sum_{l=1}^{s-1} \sum_{k=1}^{s-l} \frac{l}{s} L_{E^X_k Y_l} E^Z_{s-l-k}$$
$$= \sum_{l=1}^{s-1} \sum_{m=l+1}^{s} \frac{l}{s} L_{E^X_{m-l} Y_l} E^Z_{s-m} = \sum_{m=2}^{s} \sum_{l=1}^{m-1} \frac{l}{s} L_{E^X_{m-l} Y_l} E^Z_{s-m} \,.$$

Here the induction hypothesis is used in the first step. The last expression coincides with the second double sum in the r.h.s. of (27). Thus, the right member of (28) equals the last member of (27), and this concludes the proof.

REFERENCES

[1] G.D. Birkhoff: *Proof of Poincar's Geometric Theorem*, Transactions of the American Mathematical Society 14, 14–22 (1913).

[2] G.D. Birkhoff: *Surface transformations and their dynamical applications*, Acta Mathematica 43, 1–119 (1920).

[3] A. Deprit: *Canonical transformations depending on a small parameter*, Cel. Mech. 1, 12–30 (1969).

[4] F. Fassò: *On a relation among Lie series*, Cel. Mech. 46, 113–118 (1989).

[5] A. Giorgilli, L. Galgani: *Formal integrals for an autonomous Hamiltonian system near an equilibrium point*, Cel. Mech. 17 (1978), 267–280.

[6] A. Giorgilli, L. Galgani: *Rigorous estimates for the series expansions of Hamiltonian perturbation theory*, Cel. Mech. 37, 95–112 (1985).

[7] A. Giorgilli, U. Locatelli: *Kolmogorov theorem and classical perturbation theory*, ZAMP 48, 220–261 (1997).

[8] A. Giorgilli, U. Locatelli: *A classical self–contained proof of Kolmogorov's theorem on invariant tori*, in Proceedings of the NATO ASI school ``Hamiltonian systems with three or more degrees of freedom'', C. Simó ed., NATO ASI series C Math. Phys. Sci., Vol. 533, Kluwer Academic Publishers, Dordrecht–Boston–London, 72–89 (1999).

[9] A. Giorgilli, S. Marmi: *Improved estimates for the convergence radius in the Poincaré–Siegel problem*, Discrete and Continuous Dynamical Systems series S 3, 601–621 (2010).

[10] A. Giorgilli, M. Sansottera: *Methods of algebraic manipulation in perturbation theory*, in *Chaos, Diffusion and Non-integrability in Hamiltonian Systems - Applications to Astronomy*, Proceedings of the 3rd La Plata International School on Astronomy and Geophysics, P.M. Cincotta, C.M. Giordano and C. Efthymiopoulos eds., Universidad Nacional de La Plata and Asociacion Argentina de Astronomia Publishers, La Plata, Argentina (2012).

[11] W. Gröbner: *Nuovi contributi alla teoria dei sistemi di equazioni differenziali nel campo analitico*, Atti Accad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Nat. 23 375–379 (1957).

[12] W. Gröbner: *Die Lie–Reihen und Ihre Anwendungen*, VEB Deutscher Verlag der Wissenschaften, Mathematische Monographien, 3 (1960). Italian translation: *Serie di Lie e loro applicazioni*, Ed. Cremonese, Roma, 1973

[13] A. Haro: *The primitive function of an exact symplectomorphism*, Nonlinearity 13, 1483–1500 (2000).

[14] M. Hénon: *Numerical Study of Quadratic Area-Preserving Mappings*, Quart. Appl. Math. 27, 291–312 (1969).

[15] J. Henrard, J. Roels: *Equivalence for Lie transforms*, Cel. Mech. 10, 497–512 (1974).

[16] G. Hori: *Theory of general perturbations with unspecified canonical variables*, Publ. Astron. Soc. Japan, 18, 287–296 (1966).

[17] S. Kuksin: *On the Inclusion of an Analytic Symplectomorphism Close to an Integrable One into a Hamiltonian Flow*, Russian Journal of Mathematical Physics 1, 191–207 (1993).

[18] S. Kuksin, J. Pöschel: *On the inclusion of analytic symplectic maps in analytic Hamiltonian flows and its applications*, in: S. Kuksin, V. Lazutkin, J Pöschel (Eds.), *Seminar on dynamical systems*, Birkhäuser, Basel, 96–116 (1994)

[19] J.K. Moser: *On invariant curves of area–preserving mappings of an annulus*, Nachr. Akad. Wiss. Gött., II Math. Phys. Kl. 1962, 1–20 (1962).

[20] Poincaré, H.: *Les méthodes nouvelles de la mécanique céleste*, Gauthier–Villars, Paris (1892).

[21] H. Poincaré: *Sur un théorème de géométrie*, Rendiconti del Circolo Matematico di Palermo 33, 375–407 (1912).

[22] A.V. Pronin, D.V. Treschev: *On the inclusion of analytic maps into analytic flows*, Regular and Chaotic Dynamics 2, 14–24 (1997).

[23] E. Schröder: *Über iterierte Functionen*, Math. Ann. 3, 296–322 (1871).

[24] Siegel, C.L.: *Iterations of analytic functions*, Annals of Math. 43, 607–612 (1942).

NEURODEGENERATIVE DISEASES AND TAU: PATHOLOGICAL EVENTS AND MOLECULAR TARGETS

PIERFAUSTO SENECI (*)

Nota presentata dal m.e. Stefano Maiorana
(Adunanza del 29 novembre 2012)

SUNTO. – La proteina neuronale tau, e le malattie neurodegenerative da essa causate (taupatie), sono il cuore di questo contributo. Tau è descritta come proteina strutturalmente disordinata, attraverso i suoi domini e le sue sei isoforme. Le mutazioni di tau note, e le loro conseguenze sulla funzionalità di tau (distacco dai microtubuli, aggregazione di tau in oligomeri solubili e poi in aggregati insolubili) sono elencate ad introdurre le taupatie ad esse riferibili. E' descritta la natura e la frequenza di modifiche post-traslazionali della proteina, spiegando il loro impatto – quando la regolazione fisiologica è alterata – sulla conformazione di tau e sulla sua localizzazione, seguite dall'aggregazione patologica. Nel dettaglio, sono descritte la fosforilazione (regolatore negativo – inibizione necessaria), la glicosilazione con N-acetilglucosammina e l'isomerizzazione cis-trans dell'epitopo pT231-P232 (regolatori positive – stimolazione necessaria), spiegandone la connessione con le anomalie biochimiche di tau, e lo sfruttamento come target molecolari per l'identificazione di piccole molecole capaci di prevenire o bloccare l'avanzamento delle taupatie.

***

ABSTRACT. – The neuronal protein tau, and the related neurodegenerative diseases named tauopathies, are thoroughly described here. Tau is characterized in terms of its disorganized structure, of its isoforms, and of its structural domains. Tau mutations,

(*) Dipartimento di Chimica, Università Statale di Milano, Via Golgi 19, 20133 Milan, Italy.
E-mail: pierfausto.seneci@unimi.it

and their consequences on the functions of tau (detachment from microtubules, aggregation into soluble oligomers and insoluble aggregates), are mentioned to introduce a set of ≈30 tauopathies. The pattern of post-translational modifications of tau is described, to explain its impact – when disregulated – on the folding and mislocalization of tau, followed by its aggregation. In details, tau hyper-phosphorylation (negative regulator, to be inhibited), N-Acetylglucosamine glycosylation and pT231-P232 *cis*-trans isomerization (positive regulator, to be stimulated) are described in terms of their connection to tau biochemical abnormalities, and of their exploitation as tauopathy-directed molecular targets.

1. Epidemiology

If one looks at life expectancy, we seem to be moving in the right direction [1]. Life expectancy at birth did not significantly vary from the Neolithic (≈20 years) to Rome (between 20 and 30 years), to medieval Britain (≈30 years) and even to early 20th century Britain (≈31 years). Improved sanitary conditions, disease prevention (*e.g.*, vaccinations) and treatment (*e.g.*, antibiotics) have significantly increased life expectancy, up to the 67.2 years value in 2010 [2]. Thus, the average lifespan is steadily increasing.

Alzheimer's Disease International (ADI) estimates in its 2013 report [3] that there are more than 35 million people with dementia worldwide as of 2010, that the number will double by 2030 and triplicate to 115 million by 2050. Numbers further increase when one takes into account all neurodegenerative diseases (NDDs).

Treatment strategies for NDDs are inadequate. Limited benefits come from compensation for neuronal loss by increasing levels of corresponding neurotransmitters in the central nervous system (CNS), without directly slowing or halting neurodegeneration. Such symptomatic therapies offer temporary relief, but the ultimate NDD outcome does not change. Is it good that our life expectancy steadily increases, if existing treatments for NDD/ageing diseases only treat the symptoms, rather than addressing the cause and eradicating, or at least halting disease progression?

Pharmaceutical research and development (R&D) dealing with CNS, and in particular with NDDs is extremely risky and expensive. There's no question about that, but other costs should also be considered to fully evaluate the financials of NDDs. ADI estimates that for 2010 the global cost of neurodegeneration, including medical costs and

cost of formal (*e.g.*, nursing homes and skilled nurses) and informal (*e.g.*, relatives) care, exceeds $600 billion (about 1% of world gross domestic product), with disproportionatetly high costs in wealthy countries [4]. The cost of providing care for AD patients in the US is ≈$200 billion per year in 2012, projected to grow to $1.1 trillion per year by 2050 [5]. A detailed analysis [6] considers the global cost of so-called "diseases of the brain" in Europe at € 798 billion in 2010 (37% direct healthcare costs, 23% direct non-medical costs, 40% indirect costs associated with patients' production losses). Isn't the conservative figure of ≈$2 trillion – what we should spend in 2050 in global care for NDD patients – enough to stimulate public funding agencies and the public opinion to steadily invest in R&D?

1.1 Molecular Mechanisms

NDDs are a heterogenic set of diseases, and multiple therapeutic intervention strategies can be conceived. Disease-modifying pathways, that should prove beneficial in the treatment of several NDDs, include oxidative [7] and nitrosative [8] stress, endoplasmic reticulum (ER) stress [9], mitochondrial injuries [10], impaired protein degradation [11], chaperone mis-functioning [12], inflammatory responses [13] and heavy metal accumulation in the brain [14].

Aggregation-prone neuronal proteins are the core of NDDs. The same protein aggregate may determine the insurgence of many NDDs. Conversely, a single NDD may entail the simultaneous presence of more than one protein aggregate. For example, extra-cellular senile plaques (SPs) in the AD brain consist of β-*amyloid/A*β [15], a family of amyloidogenic peptides resulting from the cleavage by β- and γ-secretase of the amyloid precursor protein (APP) [16]. Intra-cytoplasmic protein inclusions in familial and sporadic PD, in dementia with Lewy bodies (DLB) and in multiple system atrophy (MSA) contain α-*synuclein* [17], an aggregation-prone small protein found predominantly in neuronal tissue. Intra-neuronal protein inclusions in nine polyglutamine repeat (polyQ) diseases [18], such as HD, contain polyQ-containing proteins such as *polyQ-huntingtin* [19]. Therapeutically relevant aggregation-prone proteins/NDD couples also include *superoxide dismutase-1* (*SOD-1* [20])/amyotrophic lateral sclerosis (ALS); *TAR DNA-binding protein 43* (*TDP-43* [21])/ALS; *fused in sarcoma* (*FUS* [22])/ALS; and the *prion protein* (*PrP*, [23])/prion disease. A thorough

description of disease-modifying approaches targeted against >600 known NDDs, and consequently focused onto the physio-pathological features of these aggregation-prone neuronal proteins, would largely exceed the length of any review. We rather focus on the protein *tau* and on a set of NDDs named *tauopathies*.

2. Tau and Tauopathies

2.1 Tau: Main Features

Tau is a highly soluble microtubule-associated protein (MAP) discovered in 1975 [24] that promotes microtubule (MT) assembly. Tau is mostly expressed in neurons in general, and axons in particular, ensuring their structural integrity [25]. The almost total absence of secondary and tertiary structural elements in tau makes it an intrinsically disordered protein (IDP) [12]. The switch of IDPs between folding states dynamically regulates their interaction with multiple partners through high specificity/low affinity interactions, and modulates many cellular processes and signalling pathways [26]. As to tau, tau-MT binding-unbinding events control the stabilization or destabilization of MT segments to regulate neuritic growth and promote axonal transport [27].

Protein aggregation [28] is influenced by a number of physico-chemical and biological factors in humans. The transition from a soluble, functional protein to mis-folded, aggregation-prone species, and eventually to ordered or disordered aggregates, is promoted by protein-dependent (structural features, genetic or post-traslational modifications/PTMs) and protein-independent factors (cellular, environmental stimuli) [29]. As to *structural features*, IDPs sample a larger conformational space, including several metastable, aggregation-prone conformations. Proteinopathy-prone IDPs include Aβ [30], α-synuclein [12] and tau [31]. As to *genetic modifications*, point mutations may significantly increase the aggregation tendency of a protein (*i.e.* the P301L mutation in tau [32]). As to *PTMs*, they cause aggregation and proteinopathies by influencing, *inter alia*, the solubility and the conformational stability of proteins (i.e the hyper-phosphorylation (HP) of tau [33]). The *aging process* is the most common cause of protein aggregation [34], so much that the latter is a diagnostic marker for the former process in living organisms [35]. Aging-dependent protein aggregation

is a slow, continuous process determined by subtle chronic changes in cellular components, that impact on pathology-unrelated and -related proteins.

The single copy human tau gene *MAPT*, composed by 16 exons, is located on chromosome 17 [36]. Alternative splicing produces up to 30 tau isoforms [37], six of which are expressed in CNS (*Fig. 1*). The longer human brain tau isoform (2N4R, 441 aminoacids-AAs) contains a basic C-terminal domain (AAs 244-441), including four MT-binding repeats (MTBRs); a basic middle domain (AAs 151-243), containing two proline-rich regions (PR), and an N-terminal domain (AAs 1-150).

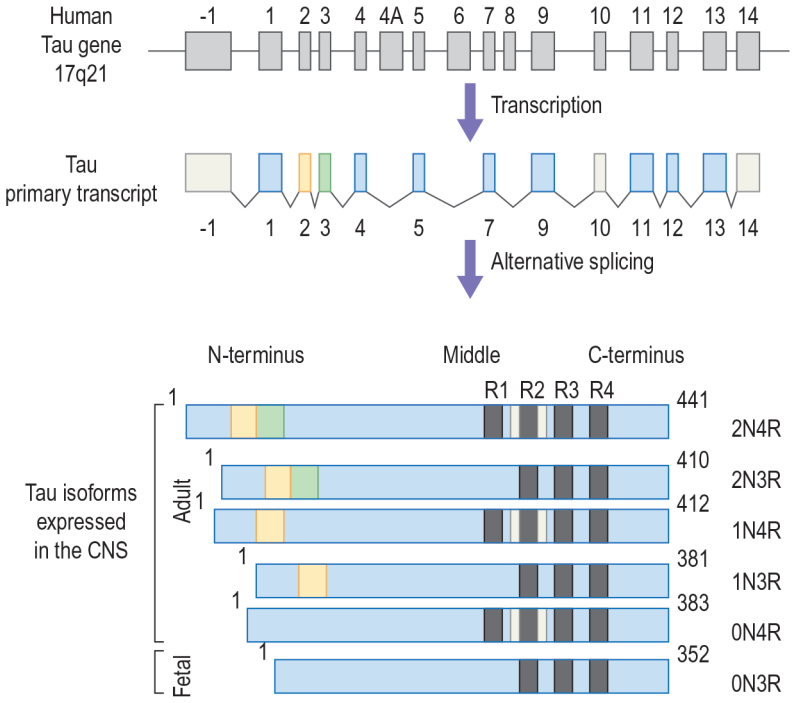


Fig. 1.

Tau isoforms result from alternative splicing of exons 2, 3 (N-terminal domain, 29 AAs in each domain, isoforms 2N, 1N and 0N) and 10 (C-terminal domain, 31 AAs, isoforms 4R and 3R). Six CNS isoforms are observed because exon 3 is expressed only in presence of exon 2 [38]. Mutations in *MAPT*, resulting in familial tauopathies, are

known. Missense mutations cause aminoacid variations in tau, while silent mutations switch the physiological 4R:3R isoform ratio [39]. Exon 10 contains an MTBR, so that 4R isoforms show stronger binding to MTs than 3R isoforms. Adult tau shows a ≈1:1 4R:3R ratio, a compromise between strong MT cohesion to secure neuronal integrity, and morphological plasticity needed by dynamic MT-tau complexes. An abnormal 4R:3R ratio in adult tau pools is invariably associated with MT dysfunctions and tau protein aggregation [40].

As it happens for other amyloidogenic proteins, insoluble tau aggregates/neurofibrillary tangles (NFTs) were long believed to be the major determinants of neurotoxic effects in tauopathies [41]. Now it is accepted that, while NFTs are presumably involved in tau toxicity, they may even represent an endogenous rescuing mechanism to eliminate neurotoxic, soluble oligomeric species from the neuronal environment [42].

Conversely, soluble tau oligomers are neurotoxic in a number of preclinical environments [43,44], are found in the extra-cellular space [45], and propagate from cell to cell in cellular [46] and *in vivo* models [47]. Neurotoxic tau oligomers are observed in patients suffering from AD [48,49] or frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17) [50]. They are observed at high concentrations in the frontal cortex of early AD patients (Braak stage 1) before NFT formation, suggesting usefulness as pre-symptomatic diagnostics [51].

2.2 Tauopathies: Biochemical and Clinical Description

Each NDD showing intracellular accumulation of filamentous tau inclusions into NFTs, and disease-dependent brain dysfunctions, is defined as a *tauopathy* [52,53]. The identification of at least fifty-seven tau mutations on chromosome 17 in patients affected by frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17) [54] proves that dysfunctions in tau cause tauopathies/NDDs without any other neuropathological factor.

At least 27 tauopathies (*Tab. 1*) are known, including largely diffuse and extremely rare NDDs. Tau aggregates are either the only neuronal abnormality, or one of the main disease factors, or secondary inclusions associated with other main pathologies. Tauopathies are divided in familial and sporadic NDDs, and in five classes depending on the neuropathological observation of tauopathy-specific tau aggregates [55].

Tab. 1. – List of characterized tauopathies.

Tauopathy	Inclusions	Origin[d]	Class
AD	Main/Aβ	S,F[d]	I
Amyotrophic lateral sclerosis/parkinsonism-dementia complex of Guam (ALS/PDC)	Main/TDP-43	S	I
Argyrophilic grain disease (AGD)	Tau only	S	II
Corticobasal degeneration (CBD)	Tau only	S,F	II
Dementia pugilistica/Chronic traumatic encephalopathy (CTE)	Main/TDP-43, Aβ	S	I
Diffuse neurofibrillary tangles with calcification (DNTC)	Main/TDP-43, α-syn	S	I
Down's syndrome (DS)	Main/Aβ[a]	S	I
Familial British dementia (FBD)	Main/Aβ-like, TDP-43	F[e]	I
Familial Danish dementia (FDD)	Main/Aβ-like	F[e]	I
Frontal lobe dementia, non-AD non-Pick	Lack of tau	S	0
Frontotemporal dementia and parkinsonism linked to chromosome 17 caused by MAPT mutations	Main/TDP-43, FUS[b]	F	I, II, III
Frontotemporal lobar degeneration caused by C9ORF72 mutations (FTLD-C9)	Support[c]/TDP-43, p62	F[e]	I
Gerstmann-Sträussler-Scheinker disease (GSS)	Support[c]/prion	S	I
Globular glial tauopathies (GGT)/White matter tauopathy with globular glial inclusions	Tau only	S	II
Guadeloupean parkinsonism with dementia (Gd-PDC)	Tau only	S	I
Guadeloupean PSP	Tau only	S	II
Multiple system atrophy (MSA)	Support[c]/α-syn	S	I
Myotonic dystrophy (DM)	Tau only[a]	F[e]	IV
Neurodegeneration with brain iron accumulation (NBIA)/Hallevorden-Spatz disease/Pantothenate kinase-associated neurodegeneration (PKAN)	Support[c]/TDP-43, α-syn	F[e]	I
Neurofibrillary tangle-predominant dementia (NFTPD)	Tau only	S	I
Niemann-Pick disease, type C (NPC)	Tau only[a]	F[e]	I
Pick's disease (PiD)	Tau only	S	III
Postencephalitic parkinsonism (PEP)	Main/Aβ	S	I
Prion protein cerebral amyloid angiopathy (PrPCAA)	Support[c]/prion	F[e]	I
Progressive supranuclear palsy (PSP)	Tau only	S	II
SLC9A6-related mental retardation	Tau only[a]	S	II
Subacute sclerosing panencephalitis (SSPE)	Tau only[a]	S	I

[a], the main disease cause is not the proteinopathy/tauopathy; [b], each pathology is typical of FTDP-17 sub-classes, no overlapping proteinopathies; [c], tau pathology has lesser importance than the other proteinopathies; [d], S=sporadic, F=familial; [e], other proteins than tau are mutated.

A single ***class 0 tauopathy*** is characterized by a lack of neuropathological hallmarks of tau [55]. ***Class I tauopathies*** (eighteen) are defined by the presence in insoluble HP tau aggregates of three major tau bands at 60, 64 and 69 kDa, and of a less intense band at 72-74 kDa [55]. The 60 kDa band corresponds to the 0N3R isoform; the 64 kDa band corresponds to the 1N3R and 0N4R isoforms; the 69 kDa band corresponds to the 2N3R and 1N4R isoforms; and the 72-74 kDa band corresponds to the longest 2N4R tau isoform. ***Class II tauopathies*** (seven) are defined by the presence in insoluble HP tau aggregates of two major tau bands at 64 and 69 kDa, and of a less intense band at 72-74 kDa [55]. This profile corresponds to aggregates composed in large part by 4R tau isoforms, as class II tau aggregates are heavily stained by exon 10-specific antibodies [53]. A single ***class III tauopathy*** is defined by the presence in insoluble HP tau aggregates of two major tau bands at 60 and 64 kDa, and of a less intense band at 69 kDa [55]. This corresponds to aggregates composed in large part by 3R tau isoforms, as class II tau aggregates do not stain in presence of exon 10-specific antibodies. A single ***class IV tauopathy*** is defined by the presence in insoluble HP tau aggregates of a single major tau band at 60 kDa, and of two less intense bands at 64 and 69 kDa [55]. This profile corresponds to aggregates composed in large part by the smallest tau isoform 0N3R [53].

3. Tau and Post-Translational Modifications (PTMs)

The basic-polar nature of tau supports its interaction with acidic MTs, and favors *PTMs* on tau [56,57]. PTM patterns, and phosphorylation in particular, heavily influence the conformational stability, the interaction network, and the physico-chemical properties (including aggregation propensity) of tau [56].

3.1 Phosphorylation: Kinases, Phosphatases, GSK-3β

Phosphorylation of tau has a strong impact on its functions [58]. Out of 85 Ser, Thr and Tyr residues, more than 30 are mostly phosphorylated in non-diseased brains, around 15 in both physiological and pathological conditions, and almost 30 are phosphorylated in AD brains [56]. Tau varies its phosphorylation state depending on its local-

ization [59] and on developmental stage, as fetal human tau is more phosphorylated than adult tau [60]. Adult HP tau is a marker for tau aggregation, and a risk factor for tauopathies [61]. HP fetal tau is highly soluble and perfectly functional [62].The HP pattern in human brain tissues from AD patients is different from human fetal tau. Namely, residues S202, T212, S214, T217, T231, S262, S396, S404 and S422 are linked to adult HP [63] as early, intermediate or late stage, tauopathy-specific hyper-phosphorylated epitopes [64,65].

An updated, online-available Table [66] provides the phosphorylation specificity for 33 putative tau kinases. In pathological conditions some of them may increase their activity on tau, and/or that the activity of their phosphatase counterparts is reduced. Targeting a *decrease of tau HP* is an assessed drug discovery approach against tauopathies, as several tau kinase inhibitors have entered clinical trials [67].

In particular, *glycogen synthase kinase 3 beta* (*GSK-3β* [68] phosphorylates *in vitro* up to 40 Ser and Thr residues [69], and is, together with CK1, the most efficient kinase in phosphorylating tau. All but one of phosphorylated residues found in control human brains, and the majority (27 out of 45) of tauopathy-specific, HP epitopes of tau found in brains from AD patients are phosphorylated *in vitro* by GSK-3β [70]. Treatment of cultured cortical neurons with small molecule GSK-3β inhibitors in a model of tau hyper-phosphorylation induced by okadaic acid (OA, a protein phosphatase 2A inhibitor) [71] reverse the OA-induced pathological tau pattern without any toxicity, *i.e.*, without inhibition of essential, physiological tau phosphorylation [71]. GSK-3β also phosphorylates the Thr668 residue on the amyloid precursor protein (APP), hinting to a potential role for GSK-3β inhibitors as neuroprotective-Aβ anti-aggregant agents [72]. *In vivo* validation of GSK-3β as a tauopathy-AD target is copious [73]. Transgenic mice bearing conditionally expressed GSK-3β as such [74], or crossed with tauopathies-inducing mutated tau isoforms [75], are known. They are considered suitable tauopathy models to test GSK-3β inhibitors.

GSK-3β is considered a valuable therapeutic target since early '80s, and a steady flow of patents and papers dealing with hundreds of compound classes as GSK-3β inhibitors started in the last decade [76], and were recently reviewed [77,78].

Fig. 2 depicts the impact of phosphorylation and of other PTMs on the dynamic equilibrium between native tau (folded, soluble species) and HP tau (aggregation-prone species). Pro-aggregation

PTMs acting on HP-tau and promoting the formation of insoluble tau aggregates are also shown.

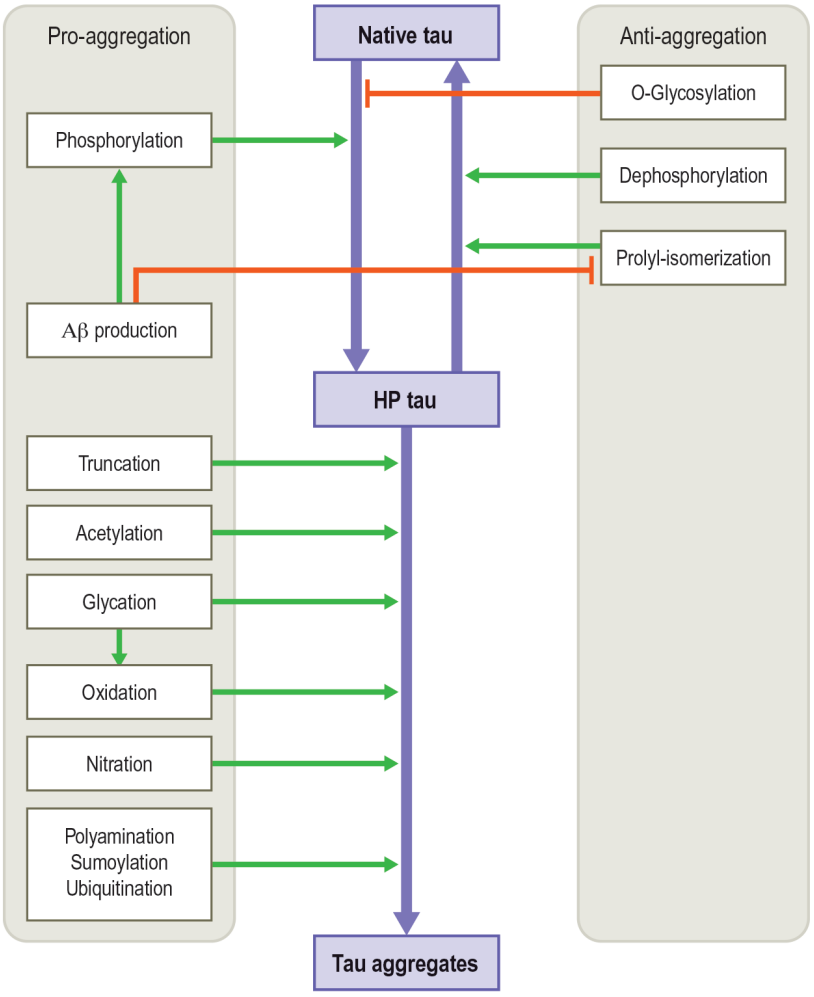


Fig. 2.

3.2 O-GlcNAcylation: Transferases, Hydrolases, OGA

O-glycosylation with N-acetylglucosamine (*O-GlcNAcylation*) is essential for the regulation of proteins in physiological and pathological

states [79]. O-GlcNAcylation negatively regulates tau phosphorylation [80], prevents tau aggregation and the risk for tauopathies. A decreased level of tau O-GlcNAcylation is observed in brains from AD patients, as is correlation between tau hypo-GlcNAcylation and HP [81]. Thus, an *increase of O-GlcNAcylation* on tau is a sound therapeutic goal.

OGT [82], an O-GlcNAc transferase, introduces an N-acetylglucosamine moiety on up to 11 Ser and Thr residues of tau, and on more than 500 other protein substrates. OGA [83], an O-GlcNAc hydrolase, removes GlcNAc from the same residues/substrates in a dynamic equilibrium. Several papers [84,85] review the mechanism of action, the substrate specificity and the main properties of OGT and OGA. Either an increase of OGT activity, or a decrease of OGA activity could lead to an increase of tau O-GlcNAcylation. The latter is a more achievable goal, as structure-based drug design of enzyme inhibitors is well documented in literature [86]. Thus, OGA has become the target of drug discovery projects.

Most known OGA inhibitors are OGlcNAc mimics, as elucidation of the substrate-assisted enzymatic mechanism [87] of OGA permits their rational design-structural optimization-fine tuning. The availability of X-ray structures of human OGA homologues complexed with inhibitors [88,89] provides inspiration for the rational drug design of OGA-targeted inhibitors with selectivity against structurally related hydrolases [83,90], as recently reviewed [91]. *In vitro* [92] and *in vivo* [93] data using a selective OGA inhibitor did not alter glucose homeostasis, and hinted to a cleaner activity profile for such molecules. Thus, the rationale for OGA inhibitors as PTM-correcting agents in neurodegeneration is strong.

3.3 Peptidyl-Prolyl Isomerization: Isomerases, Phosphatases, Pin1

Cis and *trans* conformations of the X-Pro amide bond strongly influence the folding of Pro-rich protein regions [94]. *Peptidyl prolyl isomerization* of one or more X-Pro amides in a protein may influence the accessibility of a domain to PTM enzymes, with functional consequences [95]. The middle region of tau contains two Pro-rich domains and 21 Pro residues. The inhibition of X-Pro amide bond isomerization on the pT231-P232 epitope is a pathology-related event which appears in the early phases of tau pathologies [96]. The phosphorylated epitope causes major conformational changes in tau, leading to the phosphory-

lation of other pathological epitopes [97]. Its use as a diagnostic tool to observe the progression of AD in cerebrospinal fluid (CSF) is suggested, due to a direct correlation between increased CSF levels and AD severity [98]. Thus, *pThr-Pro cis-trans isomerization* of tau is a meaningful target against tauopathies.

The parvulin PPIase *Pin1* (peptidyl-prolyl *cis/trans* isomerase NIMA interacting-1) [99] has specific PPIase activity on more than 50 protein sequences containing the pSer/pThr-Pro phosphorylated dipeptide [100]. It regulates key cellular events involved in physiological [101] and pathological processes [102]. Tau is the first Pin1 substrate identified in neurons [103]. Pin1 targets-isomerizes the *cis* pT231-P232 amide bond on tau, promoting its dephosphorylation catalyzed by *trans*-specific phosphatase PP2A [104].

A Pin1-centered intervention in tauopathies should restore its functions in diseased brain areas. Small molecule Pin1 enhancers are not yet reported, and can not be easily conceived. Pin1 modulators, *i.e.* small molecules acting on Pin1 regulators, could provide the same result. A more druggable set of targets relies on PTMs of Pin1, and in particular on the phosphorylation of two negative regulatory epitopes – Ser16 and Ser71 – respectively by polo-like-kinase-1 (PLK1) [105] and death-associated protein kinase 1 (DAPK1) [106].

Tau acetylation reduces the affinity of tau for MTs and has a pro-aggregation effect [107]. It is observed in most patients suffering by tauopathies [108], while acetylated tau is not observed in primary cultured neurons [109]. Acetylation mostly takes place on Lys274 and Lys280 residues in the MTBR [109]. The latter epitope is strongly associated to sporadic AD and other tauopathies [110].

The *proteolytic cleavage of tau* [111] is promoted by its IDP nature, that ensures the access of large tau regions to brain proteases [112]. Epitope-specific phosphorylation promotes tau processing [113], although inhibition of tau cleavage by phosphorylation of a specific residue is also observed [114]. Tau is cleaved *in vitro* by trypsin, chymotrypsin, and endogenous proteases [115]. Tissue samples from NDD patients contain tau fragments generated by caspases [116] and calpains [117]. In particular, caspases cleave tau at Asp421 in AD patients [118], producing the aggregation-prone/neurotoxic tau C fragment [119].

Tau glycation [120], nitration [121], polyamination [122], sumoylation [123] and oxidation [124] have effects on tau functions, mis-fold-

ing and aggregation, and may in future be fully validated as disease-modifying therapeutic options against tauopathies.

4. Conclusions

The prevention-remediation of tbiochemical alterations of tau is a therapeutically relevant target against tauopathies. While the restoration of physiological PTM tau profiles is an assessed route, new ones aim to refold tau and prevent its aggregation (chaperone intervention [125]), or to dispose of misfolded soluble oligomers and insoluble aggregates of tau (ubiquitin-proteasome degradation [126], autophagy [127], aggrephagy [128]). Although still largely unclarified, it is reasonable to assume that they will in some years generate a flow of clinical candidates against tauopathies.

BIBLIOGRAPHY

[1] http://en.wikipedia.org/wiki/Life_expectancy
[2] https://www.cia.gov/library/publications/the-world-factbook/rankorder/2102rank.html
[3] http://www.alz.co.uk/research/WorldAlzheimerReport2013ExecutiveSummary. pdf
[4] http://www.alz.co.uk/research/world-report
[5] http://www.alz.org/downloads/facts_figures_2012.pdf
[6] Gustavsson A., Svensson M., Jacobi F., *et al.* Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol*., **2011**, 21, 718-79.
[7] *Oxidative Stress and Neurodegenerative Disorders*. Qureshi G.A., Parvez S.H., Eds. Elsevier, Amsterdam, The Netherlands, **2007**.
[8] Pacher P., Beckman J.S., Liaudet L. Nitric oxide and peroxynitrite in health and disease. *Physiol. Rev.*, **2007**, 87, 315-424.
[9] Salminen A., Kauppinen A., Suuronen T., Kaarniranta K., Ojala J. ER stress in Alzheimer's disease: a novel neuronal trigger for inflammation and Alzheimer's pathology. *J. Neuroinflamm*., **2009**, 6, 41.
[10] Moreira P.I., Zhu X., Wang X., *et al.* Mitochondria: a therapeutic target in neurodegeneration. *Biochim. Biophys. Acta*, **2010**, 1802, 212-220.
[11] Matsuda N., Tanaka K. Does impairment of the ubiquitin-proteasome system or the autophagy-lysosome pathway predispose individuals to neurodegenerative disorders such as Parkinson's disease? *J. Alzheimers Dis*., **2010**, 19:1-9.
[12] Uversky V.N. Flexible nets of malleable guardians: Intrinsically disordered chaperones in neurodegenerative diseases. *Chem. Rev*., **2011**, 111, 1134-1166.

[13] Amor S., Puentes F., Baker D., Van Der Valk P. Inflammation in neurodegenerative diseases. *Immunology*, **2010**, 129, 154-169.
[14] Bush A.I. Metals and neuroscience. *Curr. Opin. Chem. Biol*., **2000**, 4, 184-191.
[15] Hardy J., Selkoe D.J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **2002***,* 297, 353-356.
[16] Haass C., De Strooper B. The presenilins in Alzheimer's disease-proteolysis holds the key. *Science*, **1999***,* 286, 916-919.
[17] Goedert M. α-synuclein and neurodegenerative diseases. *Nat. Rev. Neurosci*., **2001***,* 2, 492-501.
[18] Zoghbi H.Y., Orr H.T. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci*., **2000**, **23**, 217-247.
[19] Young A.B. Huntingtin in health and disease. *J. Clin. Invest*., **2003**, 111, 299-302.
[20] Rosen D.R., Siddique T., Patterson D., *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature*, **1993**, 362, 59-62.
[21] Neumann M., Sampathu D.M., Kwong L.K., *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, **2006**, 314, 130-133.
[22] Kwiatkowski T.J. Jr., Bosco D.A., Leclerc A.L., *et al.* Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science*, **2009**, 323, 1205-1208.
[23] Prusiner S.B., Scott M.R., DeArmond S.J., Cohen F.E. Prion protein biology. *Cell*, **1998**, 93, 337-348.
[24] Weingarten M.D., Lockwood A.H., Hwo S.Y., Kirschner M.W. A protein factor essential for microtubule assembly. *Proc. Natl. Acad. Sci. U.S.A.,* **1975***,* 72, 1858-1862.
[25] Hirokawa N. Microtubule organization and dynamics dependent on microtubule-associated proteins. *Curr. Opin. Cell Biol*., **1994**, 6, 74-81.
[26] Dyson H.J., Wright P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.,* **2005**, 6, 197-208.
[27] Zheng S., Chen Y., Donahue C.P., Wolfe M.S., Varani G. Structural basis for stabilization of the tau pre-mRNA splicing regulatory element by novantrone (mitoxantrone). *Chem. Biol*., **2009**, 16, 557-566.
[28] Chiti F., Dobson C.M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem*., **2006**, 75, 333-366.
[29] Uversky V.N. Mysterious oligomerization of the amyloidogenic proteins. *FEBS J.*, **2010**, 277, 2940-2953.
[30] Rosenman D.J., Connors C.R., Chen W., Wang C., García A.E. Aβ monomers transiently sample oligomer and fibril-like configurations: Ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.,* **2013**, 425, 3338-3359.
[31] Elbaum-Garfinkle S., Rhoades E. Identification of an aggregation-prone structure of tau. *J. Am. Chem. Soc*., **2012**, 134, 16607-16613.
[32] Terwel D., Lasrado R., Snauwaert J., *et al.* Changed conformation of mutant tau-

P301L underlies the moribund tauopathy, absent in progressive, nonlethal axonopathy of tau-4R/2N transgenic mice. *J. Biol. Chem*., **2005**, 280, 3963-3973.
[33] Vassar R., Bennett B.D., Babu-Khan S., *et al.* β- Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science*, **1999**, 286, 735-741.
[34] David D.C. Aging and the aggregating proteome. *Front. Genet*., **2012**, 3, 347.
[35] David D.C., Ollikainen N., Trinidad J.C., *et al.* Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS Biol*., **2010**, e1000450.
[36] Neve R.L., Harris P., Kosik K.S., Kurnit D.M., Donlon T.A. Identification of cDNA clones for the human microtubule-associated protein tau and chromosomal localization of the genes for tau and microtubule-associated protein 2. *Brain Res*., **1986**, 387, 271-280.
[37] Andreadis A. Tau gene alternative splicing: Expression pattern, regulation and modulation of function in normal brain and neurodegenerative disease. *Biochem. Biophys. Acta*, **2005**, 1739, 91-103.
[38] Andreadis A., Brown W.M., Kosik K.S. Structure and novel exons of the human *T* gene. *Biochemistry*, **1992**, 31, 10626-10633.
[39] Goedert M., Spillantini M.G. Pathogenesis of the tauopathies. *J. Mol. Neurosci.*, **2011**, 45, 425-431.
[40] Crespo-Biel N., Theunis C., Van Leuven F. Protein tau: Prime cause of synaptic and neuronal degeneration in Alzheimer's disease. *Int. J. Alzheim. Dis*., **2012**, 251426.
[41] Arriagada P.V., Growdon J.H., Hedley-Whyte E.T., Hyman B.T. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology*, **1992**, 42, 631-639.
[42] Spires-Jones T.L., Kopeikina K.J., Koffie R.M., de Calignon A., Hyman B.T. Are tangles as toxic as they look? *J. Mol. Neurosci*., **2011**, 45, 438-444.
[43] Berger Z., Roder H., Hanna A., *et al.* Accumulation of pathological tau species and memory loss in a conditional model of tauopathy. *J. Neurosci*., **2007**, 27, 3650-3662.
[44] Wittmann C.W., Wszolek M.F., Shulman J.M., *et al.* Tauopathy in *drosophila*: neurodegeneration without neurofibrillary tangles. *Science*, **2001**, 293, 711-714.
[45] Lasagna-Reeves C.A., Castillo-Carranza D.L., Sengupta U., *et al.* Identification of oligomers at early stages of tau aggregation in Alzheimer's disease. *FASEB J*., **2012**, 26, 1946-1959.
[46] Michel C.H., Kumar S., Pinotsi D., *et al.* Extracellular monomeric tau protein is sufficient to initiate the spread of tau protein pathology. *J. Biol. Chem*., **2014**, 289, 956-967.
[47] de Calignon A., Polydoro M., Suárez-Calvet M., *et al.* Propagation of tau pathology in a model of early Alzheimer's disease. *Neuron*, **2012**, 73, 685-697.
[48] Ladiwala A.R.A., Dordick J.S., Tessier P.M. Aromatic small molecules remodel toxic soluble oligomers of amyloid β through three independent pathways. *J. Biol. Chem*., **2011**, 286, 3209-3218.
[49] Mori T., Rezai-Zadeh K., Koyama N., *et al.* Tannic acid is a natural β-secretase

inhibitor that prevents cognitive impairment and mitigates Alzheimer-like pathology in transgenic mice. *J. Biol. Chem*., **2012**, 287, 6912-6927.
[50] Chang E., Congdon E.E., Honson N.S., Duff K.E., Kuret J. Structure-activity relationship of cyanine tau aggregation inhibitors. *J. Med. Chem*., **2009**, 52, 3539-3547.
[51] Song J.M., DiBattista A.M., Sung Y.M., *et al.* A tetra(ethylene glycol) derivative of benzothiazole aniline ameliorates dendritic spine density and cognitive function in a mouse model of Alzheimer's disease. *Exp. Neurol*., **2014**, 252, 105-113.
[52] Buee L., Bussie T., Buee-Scherrerb V., Delacourte A., Hof P.R. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res. Rev*., **2000**, 33, 95-130.
[53] Lee V.M.-J., Goedert M., Trojanowski J.Q. Neurodegenerative tauopathies. *Annu. Rev. Neurosci*., **2001**, 24, 1121-1159.
[54] Spillantini M.G., Goedert M. Tau pathology and neurodegeneration. *Lancet Neurol*., **2013**, 12, 609-622.
[55] Sergeant N., Delacourte A., Buee L. Tau protein as a differential biomarker of tauopathies. *Biochim. Biophys. Acta*, **2005**, 1739, 179-197.
[56] Martin L., Latypova X., Terro F. Post-translational modifications of tau protein: Implications for Alzheimer's disease. *Neurochem. Int*., **2011**, 58, 458-471.
[57] Perry G., Zhu X., Smith M.A., Sorensen A., Avila J., Wang J.-Z., Xia Y.-Y., Grundke-Iqbal I., Iqbal K. Abnormal hyperphosphorylation of tau: Sites, regulation, and molecular mechanism of neurofibrillary degeneration. *J. Alzheimer's Dis*., **2013**, 33, S123-S139.
[58] Lee V.M. Regulation of tau phosphorylation in Alzheimer's disease. *Ann. N.Y. Acad. Sci*., **1996**, 777, 107-113.
[59] Riederer B.M., Binder L.I. Differential distribution of tau proteins in developing rat cerebellum. *Brain Res. Bull*., **1994**, 33, 155-161.
[60] Brion J.-P., Octave J.N., Couck A.M. Distribution of the phosphorylated microtubule-associated protein tau in developing cortical neurons. *Neuroscience*, **1994**, 63, 895-909.
[61] Badiola N., Suarez-Calvet M., Lleo A. Tau phosphorylation and aggregation as a therapeutic target in tauopathies. *CNS Neurol. Dis. Drug Targets*, **2010**, 9, 727-740.
[62] Hof P.R., Simic G. Human fetal tau protein isoform: Possibilities for Alzheimer's disease treatment. *Int. J. Biochem. Cell Biol*., **2012**, 44, 1290-1294.
[63] Yu Y., Run X., Liang Z., *et al.* Developmental regulation of tau phosphorylation, tau kinases, and tau phosphatases. *J. Neurochem*., **2009**, 1480-1494.
[64] Su B., Wang X., Drew K.L., *et al.* Physiological regulation of tau phosphorylation during hibernation. *J. Neurochem*., **2008**, 105, 2098-2108.
[65] Bertrand J., Plouffe V., Senechal P., Leclerc N. The pattern of human tau phosphorylation is the result of priming and feedback events in primary hippocampal neurons. *Neuroscience*, **2010**, 168, 323-334.
[66] http://cnr.iop.kcl.ac.uk/hangerlab/tautable

[67] Savage M.J., Gingrich D.E. Advances in the development of kinase inhibitor therapeutics for Alzheimer's disease. *Drug Devel. Res*., **2009**, 70, 125-144.
[68] Mondragon-Rodriguez S., Perry G., Zhu X., Moreira P.I., Williams S. Glycogen synthase kinase 3: a point of integration in Alzheimer's disease and a therapeutic target? *Int. J. Alzheimer's Dis*., **2012**, 276803.
[69] Hanger D.P., Byers H.L., Wray S., *et al.* Novel phosphorylation sites in tau from Alzheimer brain support a role for casein kinase 1 in disease pathogenesis. *J. Biol. Chem.*, **2007**, 282, 23645-23654.
[70] Hanger D.P., Noble W. Functional implications of glycogen synthase kinase-3-mediated tau phosphorylation. *Int. J. Alzheim. Dis*., **2011**, 352805.
[71] Martin L., Page G., Terro F. Tau phosphorylation and neuronal apoptosis induced by the blockade of PP2A preferentially involve GSK-3β. *Neurochem. Int*., **2011**, 59, 235-250.
[72] Aplin A.E., Gibb G.M., Jacobsen J.S., Gallo J.M., Anderson B.H. *In vitro* phosphorylation of the cytoplasmic domain of the amyloid precursor protein by glycogen synthase kinase-3β. *J. Neurochem*., **1996**, 67, 699-707.
[73] Gomes-Sintes R., Hernandez F., Lucas J.J., Avila J. GSK-3 mouse models to study neuronal apoptosis and neurodegeneration. *Front. Molec. Neurosci*., **2011**, 4, 45.
[74] Engel T., Hernandez F., Avila J., Lucas J.J. Full reversal of Alzheimer's disease-like phenotype in a mouse model with conditional overexpression of glycogen synthase kinase-3. *J. Neurosci.,* **2006**, 26, 5083-5090.
[75] Terwel D., Muyllaert D., Dewachter I., *et al.* Amyloid activates GSK-3beta to aggravate neuronal tauopathy in bigenic mice. *Am. J. Pathol ,* **2008**, 172, 786-798.
[76] Phukan S., Babu V.S., Kannoji A., Hariharan R., Balaji V.N. GSK-3beta: role in therapeutic landscape and development of modulators. *Br. J. Pharmacol.*, **2010**, 160, 1-19.
[77] Cohen P., Goedert M. GSK3 inhibitors: Development and therapeutic potential. *Nat. Rev. Drug Discov*., **2004**, 3, 479-487.
[78] Eldar-Finkelman H., Martinez A. GSK-3 inhibitors: preclinical and clinical focus on CNS. *Front. Molec. Neurosci*., **2011**, 4, 32.
[79] Hanover J.A., Krause M.W., Love D.C. Bittersweet memories: Linking metabolism to epigenetics through O-GlcNacylation. *Nat. Rev. Mol. Cell Biol*., **2012**, 13, 312-321.
[80] Lazarus B.D., Love D.C., Hanover J.A. O-GlcNAc cycling: Implications for neurodegenerative disorders. *Int. J. Biochem. Cell Biol*., **2009**, 41, 2134-2146.
[81] Liu Y., Liu F., Grundke-Iqbal I., Iqbal K., Gong Cheng-X. Brain glucose transporters, O-GlcNAcylation and phosphorylation of tau in diabetes and Alzheimer's disease. *J. Neurochem*., **2009**, 111, 242-249.
[82] Haltiwanger R.S., Holt G.D., Hart G.W. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-*N*-acetylglucosamine:peptide beta-*N*-acetylglucosaminyltransferase. *J. Biol. Chem.,* **1990***,* 265, 2563-2568.
[83] Comtesse N., Maldener E., Meese E. Identification of a nuclear variant of

MGEA5, a cytoplasmic hyaluronidase and a beta-*N*-acetylglucosaminidase. *Biochem. Biophys. Res. Commun.,* **2001***,* 283, 634-640.

[84] Gloster T.M., Vocadlo D.J. Mechanism, structure, and inhibition of O-GlcNAc processing enzymes. *Curr. Signal Transduct, Ther.,* **2010***,* 5*,* 74-91.

[85] Shen D.L., Gloster T.M., Yuzwa S.A., Vocadlo D.J. Insights into O-linked N-Acetylglucosamine (O-GlcNAc) processing and dynamics through kinetic analysis of OGlcNAc transferase and O-GlcNACase activity on protein substrates. *J. Biol. Chem*., **2012**, 287, 15395-15408.

[86] Kuntz I.D. Structure-based drug design: past, present and future. *Solvay Pharm. Conf*., **2009**, 9, 23-27.

[87] Vocadlo D.J., Davies G.J. Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol*., **2008**, 12, 539-555.

[88] Dennis R.J., Taylor E.J., Macauley M.S., *et al.* Structure and mechanism of a bacterial beta-glucosaminidase having *O*-GlcNAcase activity. *Nat. Struct. Mol. Biol*., **2006**, 13, 365-371.

[89] He Y., Bubb A.K., Stubbs K.A., Gloster T.M., Davies G.J. Inhibition of a bacterial O-GlcNAcase homologue by lactone and lactam derivatives: structural, kinetic and thermodynamic analyses. *Amino Acids*, **2011**, 40, 829-839.

[90] Lameira J., Nahum Alves C., Tunon I., Marti S., Moliner V. Enzyme molecular mechanism as a starting point to design new inhibitors: a theoretical study of O-GlcNAcase. *J. Phys. Chem. B*, **2011**, 8, 6764-6775.

[91] Macauley M.S., Vocadlo D.J. Increasing O-GlcNAc levels: An overview of small-molecule inhibitors of O-GlcNAcase. *Biochim. Biophys. Acta*, **2010**, 1800, 74-91.

[92] Macauley M.S., He Y., Gloster T.M., *et al.* Inhibition of O-GlcNAcase using a potent and cell-permeable inhibitor does not induce insulin resistance in 3T3-L1 adipocytes. *Chem. Biol.*, **2010**, 17, 937-948.

[93] Macauley M.S., Shan X., Yuzwa S.A., Gloster T.M., Vocadlo D.J. Elevation of global O-GlcNAc in rodents using a selective O-GlcNAcase inhibitor does not cause insulin resistance or perturb glucohomeostasis. *Chem. Biol*., **2010**, 17, 949-958.

[94] Wedemeyer W.J., Welker E., Scheraga H.A. Proline cis-trans isomerization and protein folding. *Biochemistry*, **2002**, 41, 14637-14644.

[95] Zhou X.Z., Lu P.-J., Wulf G., Lu K.P. Phosphorylation-dependent prolyl isomerization. A novel signaling regulatory mechanism. *Cell. Mol. Life Sci*., **1999**, 56, 788-806.

[96] Hampel H., Buerger K., Kohnken R., *et al.* Tracking of Alzheimer's disease progression with cerebrospinal fluid tau protein phosphorylated at threonine 231, *Ann. Neurol*., **2001**, 49, 545-546.

[97] Luna-Munoz J., Chavez-Macias L., Garcia-Sierra F., Mena R. Earliest stages of tau conformational changes are related to the appearance of a sequence of specific phosphodependent tau epitopes in Alzheimer's disease. *J. Alzheimer's Dis.,* **2007***,* 12, 365-375.

[98] Kohnken R., Buerger K., Zinkowski R., *et al.* Detection of tau phosphorylated at

threonine 231 in cerebrospinal fluid of Alzheimer's disease patients. *Neurosci. Lett*., **2000**, 287, 187-190.
[99] Lu K.P., Hanes S.D., Hunter T. A human peptidylprolyl isomerase essential for regulation of mitosis. *Nature,* **1996***,* 380, 544-547.
[100] Lu K.P., Zhou X.-Z. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat. Rev. Mol. Cell. Biol*., **2007**, 8, 904-916.
[101] Pearson G., Robinson F., Beers Gibson T., *et al.* Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr. Rev.,* **2001***,* 22, 153C-183C.
[102] Balastik M., Lim J., Pastorino L., Lu K.P. Pin1 in Alzheimer's disease: multiple substrates, one regulatory mechanisms? *Biophys. Biochem. Acta,* **2007***,* 1772, 422-429.
[103] Lu P.J., Wulf G., Zhou X.Z., Davies P., Lu K.P. The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature,* **1999***,* 399, 784-788.
[104] Landrieu I., Smet-Nocca C., Amniai L., *et al.* Molecular implication of PP2A and Pin1 in the Alzheimer's disease specific hyperphosphorylation of tau. *PLoS One*, **2011**, 6, e21521.
[105] Bulbarelli A., Lonati E., Cazzaniga E., Gregori M., Masserini M. Pin1 affects tau phosphorylation in response to Aβ oligomers. *Mol. Cell. Neurosci*., **2009**, 42, 75-80.
[106] Bialik S., Kimchi A. The death-associated protein kinases: structure, function, and beyond. *Annu. Rev. Biochem*., **2006**, 75, 189-210.
[107] Min S.W., Cho S.H., Zhou Y., *et al.* Acetylation of tau inhibits its degradation and contributes to tauopathy. *Neuron*, **2010**, 67, 953-966.
[108] Cohen T.J., Guo J.L., Hurtado D.E., *et al.* The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nat. Commun*., **2011**, 2, 252.
[109] Cohen T.J., Friedmann D., Hwang A.W., Marmorstein R., Lee V.M.Y. The microtubule-associated tau protein has intrinsic acetyltransferase activity. *Nat. Struct. Mol. Biol*., **2013**, 20, 756-762.
[110] Irwin D.J., Cohen T.J., Grossman M., *et al.* Acetylated tau, a novel pathological signature in Alzheimer's disease and other tauopathies. *Brain*, **2012**, 135, 807-818.
[111] Hanger D.P., Wray S. Tau cleavage and tau aggregation in neurodegenerative disease. *Biochem. Soc. Trans*., **2010**, 38, 1016-1020.
[112] Proteases in the brain. Volume 3 in "*Proteases in biological diseases*". **2005**. Editors: Lendeckel U., Hooper N.M. Springer Science and Business Media, Inc., New York. 383 pages.
[113] Johnson G.V.W. Tau phosphorylation and proteolysis: Insights and perspectives. *J. Alzheim. Dis*., **2006**, 9, 243-250.
[114] Guillozet-Bongaards A.L., Cahill M.E., Cryns V.L., *et al.* Pseudophosphorylation of tau at serine 422 inhibits caspase cleavage: *in vitro* evidence and implications for tangle formation. *J. Neurochem*., **2006**, 97, 1005-1014.
[115] Wang I., Garg S., Mandelkow E.-M., Mandelkow E. Proteolytic processing of tau. *Biochem. Soc. Trans*., **2010**, 38, 955-961.

[116] Rohn T.T. The role of caspases in Alzheimer's disease; potential novel therapeutic opportunities. *Apoptosis*, **2010**, 15, 1403-1409.
[117] Getz G.S. Calpain inhibition as a potential treatment of Alzheimer's disease. *Am. J. Physiol.*, **2012**, 181, 388-391.
[118] Basurto-Islas G., Luna-Munoz J., Guillozet-Bongaards A.L., *et al.* Accumulation of aspartic acid 421- and glutamic acid 391-cleaved tau in neurofibrillary tangles correlates with progression in Alzheimer's disease. *J. Neurophathol. Exp. Neurol.*, **2008**, 67, 470-483.
[119] Chung C.W., Song Y.H., Kim I.K., *et al.* Proapoptotic effects of tau cleavage product generated by caspase-3. *Neurobiol. Dis.*, **2001**, 8, 162-172.
[120] Nacharaju P., Ko L., Yen S.H. Characterization of *in vitro* glycation sites of tau. *J. Neurochem.*, **1997**, 69, 1709-1719.
[121] Horiguchi T., Uryu K., Giasson B.I., *et al.* Nitration of tau protein is linked to neurodegeneration in tauopathies. *Am. J. Pathol.*, **2003**, 163, 1021-1031.
[122] Halverson R.A., Lewis J., Frausto S., Hutton M., Muma N.A. Tau protein is cross-linked by transglutaminase in P301L tau transgenic mice. *J. Neurosci.*, **2005**, 25, 1226-1233.
[123] Takahashi K., Ishida M., Komano H., Takahashi H. SUMO-1 immunoreactivity co-localizes with phospho-Tau in APP transgenic mice but not in mutant Tau transgenic mice. *Neurosci. Lett.*, **2008**, 441, 90-93.
[124] Schweers O., Mandelkow E.M., Biernat J., Mandelkow E. Oxidation of cysteine-322 in the repeat domain of microtubule-associated protein tau controls the *in vitro* assembly of paired helical filaments. *Proc. Natl. Acad. Sci. USA*, **1995**, 92, 8463-8467.
[125] Bukau B., Weissman J., Horwich A. Molecular chaperones and protein quality control. *Cell,* **2006**, 125, 443-451.
[126] Weissman A.M., Shabek N., Ciechanover A. The predator becomes the prey: regulating the ubiquitin system by ubiquitylation and degradation. *Nat. Rev. Mol. Cell Biol.*, **2011**, 12, 605-620.
[127] He C., Klionsky D.J. Regulation mechanisms and signaling pathways of autophagy. *Annu. Rev. Genet.*, **2009**, 43, 67-93.
[128] Lamark T., Johansen T. Aggrephagy: Selective disposal of protein aggregates by macroautophagy. *Int. J. Cell Biol.*, **2012**, 736905.

PERCHÉ IL GUSTO DEL CAFFÈ È COSÌ BUONO

MARIA ELISABETTA BRENNA (*), CLAUDIO FUGANTI (*)

Nota presentata dal m.e. Claudio Fuganti
(Adunanza del 20 dicembre 2012)

SUNTO. – Il caffè contiene un numero molto elevato di composti volatili che contribuiscono al caratteristico aroma. Si tratta di 'aromi termici' generati durante la tostatura. Nessuno di questi possiede le caratteristiche di 'impact flavour', per cui il gusto del caffè dipende dalla presenza contemporanea di un numero elevato di componenti. Tuttavia, almeno quattro di questi sono utili per dare un 'buon'caffè.

***

ABSTRACT. – Coffee contains a huge amount of volatile components which contribute to the characteristic flavour. These materials are generated during the heating process. No one of these materials can be considered the 'impact flavour'. Accordingly, the whole flavour depends upon the presence of an elevated number of components. However, at least four of the materials of are essential for a 'good' coffee.

L'albero del caffè appartiene alla famiglia delle Rubiaceae, ordine Gentianales, 80 specie nel genere *Coffea*. Di queste, solo *C. arabica* e *C. canephora* var. *robusta* sono commercialmente importanti, con produzione annua di 6 milioni ton/anno (*Fig. 1*). Nella *Fig. 2* è riportato un grafico della produzione (anno 2001) dei maggiori Paesi. Nella *Fig. 3* è riportato un grafico del consumo pro-capite kg/anno per i maggiori paesi consumatori (anno 2000). Nella *Fig. 4* c'è la composizione tipica del caffè grezzo per le

(*) Dipartimento di Chimica del Politecnico di Milano, Italy.
E-mail: mariaelisabetta.brenna@polimi.it

due varietà arabica e robusta. Nella *Fig. 5* sono riportate la formula di struttura dei componenti chimici più rilevanti, che tuttavia non contribuiscono all'aroma dell'infuso acquoso derivante dai chicchi di caffè tostati.

Fig. 1.

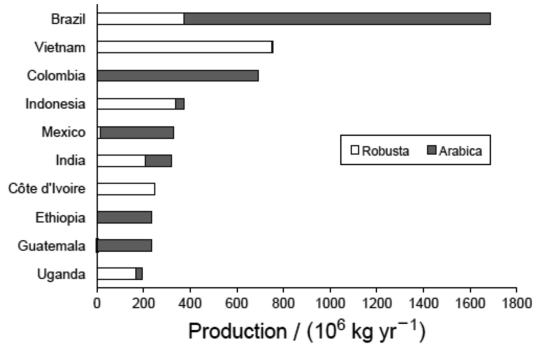


Fig. 2.

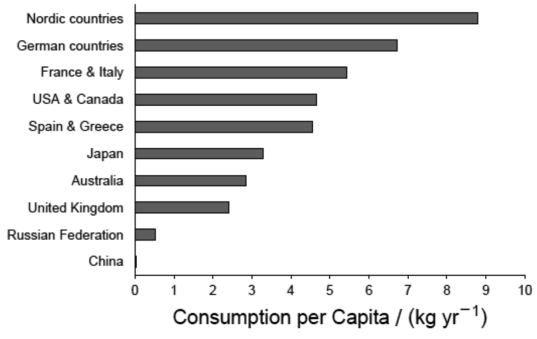


Fig. 3.

Composizione tipica del caffè grezzo

componenti	*arabica (%)*	*robusta (%)*
Caffeina	0.9 - 1.2	1.6 - 2.4
Minerali	3.4 - 4.2	4.0 - 4.5
Lipidi	12.2 - 18.0	9.0 - 13.0
Trigonellina	1.0 - 1.2	0.6 - 0.75
Proteine	11.0 - 13.0	11.0 - 13.0
Acidi alifatici	1.5 - 2.0	1.5 - 2.0
Acidi clorogenici	5.5 - 8.0	7.0 - 10.0
Oligosaccaridi	6.8 - 8.0	5.0 - 7.0
Polisaccaridi	50.0 - 55.0	37.0 - 47.0

Fig. 4.

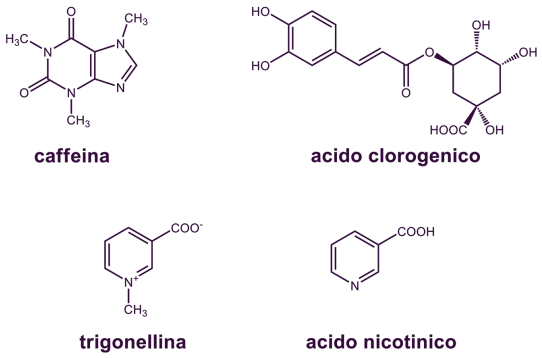


Fig. 5.

La storia dell'analisi dei componenti volatili dell'infuso del caffè tostato (=caffè) inizia più di un secolo fa (*Fig. 6*). Dei dieci componenti volatili isolati puri all'inizio del secolo scorso nessuno possedeva il gusto del caffè. Ma il primo studio sistematico sull'aroma del caffè iniziò dopo la fine della prima guerra mondiale in Svizzera, paese che ebbe molto a soffrire durante la guerra per la mancanza di caffè. In seguito agli studi di Reichstein (premio Nobel per la chimica nel 1939) e di Staudinger furono isolati attraverso distillazione e condensazione a -180°C settanta composti odorosi, ventinove dei quali erano liquidi.

Nessuno dei composti isolati ricordava nell'odore il caffè. Solo uno, il 2-furfurilmetantiolo, conferiva alla soluzione acquosa molto diluita un vago profumo di caffè molto tostato (*Fig. 7*).

Fig. 6.

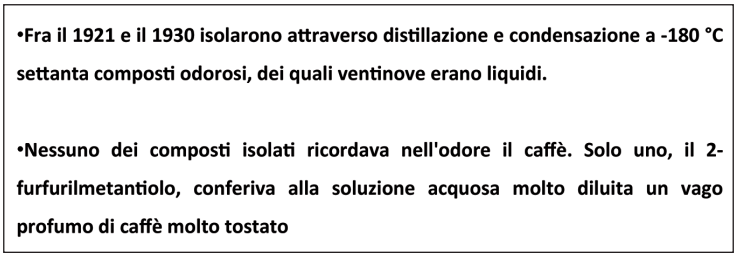


Fig. 7.

La parte più importante ai fini dello sviluppo dell'aroma del caffè é la tostatura. Questo processo che può raggiungere anche i 220°C determina una drastica alterazione della composizione chimica del caffè, mantenendo inalterato il contenuto di caffeina e di lipidi (*Fig. 8*). Diminuiscono soprattutto zuccheri e amminoacidi che vengono consumati in reazioni di Maillard. Si tratta di reazioni fra l'ammino gruppo degli amminoacidi e il carbonile degli zuccheri. La decomposizione successiva durante l'infusione di questi intermedi labili genera compo-

sti a basso peso molecolare, che in parte polimerizzano dando luogo alle melanoidine con peso molecolare di *ca.* 10.000 Da (*Fig. 9*).

La tostatura

Durante la tostatura si realizzano reazioni pirolitiche all'interno della cellule del caffè che possono essere assimilate a delle piccole autoclavi in cui si raggiunge anche la pressione di 25 bar, grazie all'inusuale robustezza delle pareti cellulari. La temperatura interna più elevata raggiunta per la preparazione dell'espresso è di 220 °C.

La tostatura modifica drasticamente la composizione chimica, mantenendo inalterato il contenuto di caffeina e di lipidi.

Fig. 8.

Le quantità di zuccheri e di ammino acidi liberi o legati alle proteine diminuisce di molto, venendo consumati nelle reazioni di *Maillard*. Queste sono reazioni di condensazione fra il carbonile di uno zucchero riducente e l'ammino gruppo di un ammino acido. La decomposizione di tali intermedi labili genera una moltitudine di composti a basso peso molecolare, i più reattivi dei quali polimerizzano in strutture poco definite.

Accanto ai materiali polimerici (massa ≥ 10.000 Da), denominati *melanoidine*, colorate di marrone e contenenti azoto, si sviluppano anidride carbonica e centinaia di composti volatili che impartiscono al caffè il tipico aroma. Le *melanoidine*, solubili in acqua, agiscono come pigmenti e sono rilevanti ai fini del binding degli aroma e della stabilizzazione della schiuma.

Fig. 9.

In *Fig. 10* è riportato il numero di componenti aromatici di caffè, cacao e tè per le varie classi strutturali. Colpisce l'elevato numero di fenoli, furani e tiofeni del caffè in paragone con cacao e tè.

Nella *Fig. 11* sono riportate le valutazioni aromatiche di soluzioni acquose dei singoli componenti volatili del caffè. In *Fig. 12* ci sono le strutture e le soglie di percezione in acqua dei più importanti componenti aromatici del caffè. In *Fig. 13* sono riportate le strutture di un aldo-esoso e di tre furani solforati molto importanti ai fini dell'aroma complessivo

del caffè. La *Fig. 14* riporta la soglia di percezione dei dieci componenti aromatici ritenuti più importanti per l'espressione di un buon aroma caffè, mentre nella *Fig. 15* sono riportate le concentrazioni delle varie sostanze nelle varietà arabica e robusta. Da notare le differenze di quantità di damascenone, guaiacolo e 4-vinilguaiacolo fra le due.

Costituenti aromatici di: caffè, cacao e tè

	caffè	cacao	tè
1. idrocarburi	51	45	28
2. alcooli	19	27	45
3. aldeidi	28	24	54
4. chetoni	70	28	55
5. acidi	20	37	27
6. esteri	30	55	62
7. lattoni	8	6	14
8. ammine	21	9	6
9. tioli, solfuri	13	15	5
10. fenoli	44	8	15
11. furani	108	16	12
12. tiofeni	26	-	1
13. pirroli	74	10	10
14. pirazine	79	77	23
15. tiazoli	27	3	7
16. ossazoli	28	4	1

Fig. 10.

Aroma di soluzioni acquose di singoli componenti volatili del caffè

Composto	Aroma
2-furfuriltiolo	caffè troppo tostato
4-vinilguaiacolo	aromatico-fenolico
4-idrossi-2,5-dimetil-3[2H]-furanone	caramello
2-isobutil-3-metossipirazina	paprica
metionale	patata fritta
3-mercapto-3-metilbutilformiato	urina di gatto
2-etil-3,5-dimetilpirazina	terroso-tostato
2,3-dietil-5-metilpirazina	" "
3-idrossi-4,5-dimetil-2[5H]-furanone	minestra di verdura
5-etil-3-idrossi-4-metil-2[5H]-furanone	" "
(E)-damascenone	miele-fruttato
vanillina	Vaniglia
2-metil-3-furantiolo	carne arrostita
acidi 2/3-metilbutirrici	sudore
3-mercapto-3-metilbutan-1-olo	brodo di carne
2,6-dimetil-3-isopropilpirazina	patata lessata con la buccia
2-metilditiometilfurano	pane bianco fresco di forno

Fig. 11.

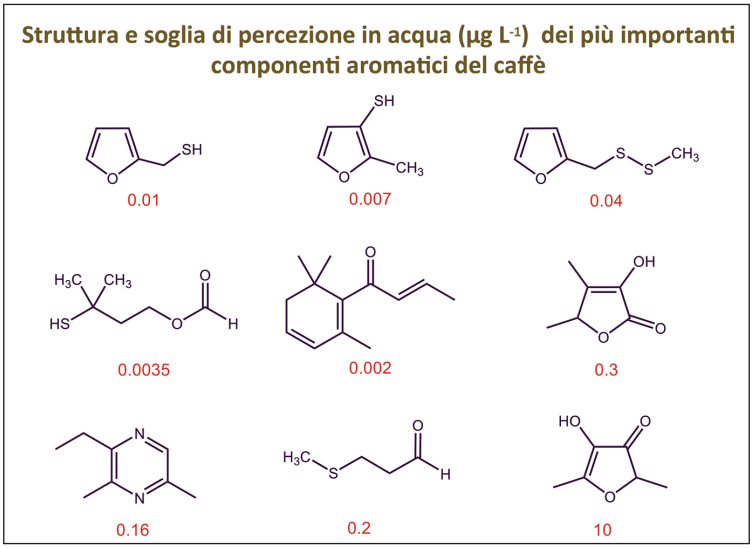


Fig. 12.

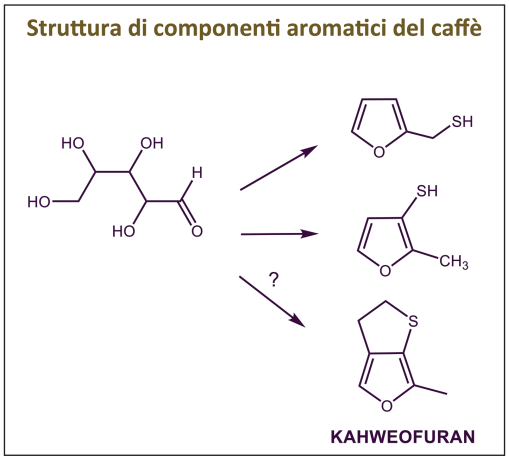


Fig. 13.

In *Fig. 16* sono riportate le concentrazioni e il valore aromatico dei principali componenti odorosi dei caffè arabica e robusta tostati, da cui si evince che sono i primi quattro composti della lista quelli che caratterizzano la bontà del caffè.

Soglia di percezione dei principali componenti aromatici del caffè sciolti in acqua in µg L^{-1}

2-furfurilmetantiolo	0,01
2,3-butandione	15
4-idrossi-2,5-dimetil-3[2*H*]-furanone	10
metionale	0,2
3-mercapto-3-metilbutilformiato	0,0035
2-etil-3,5-dimetilpirazina	0,16
3-idrossi-4,5-dimetil-2[5*H*]-furanone	0,3
β-damascenone	0,002
2-metil-3-furantiolo	0,007
2-metilditiometilfurano	0,04

Fig. 14.

Concentrazione dei principali componenti odorosi dei caffè arabica e robusta tostati

composto	arabica [mg kg^{-1}]	robusta [mg kg^{-1}]
2-furfurilmetantiolo	1,08	1,73
β-damascenone	1,195	0,205
3-mercapto-3-metilbutilformiato	0,130	0,115
3-metil-2-butentiolo	0,0082	0,0083
2-isobutil-3-metossipirazina	0,083	0,012
4-idrossi-2,5-dimetil-3[2*H*]-furanone	109,0	57,0
guaiacolo	4,2	28,2
4-vinilguaiacolo	64,8	177,7
2,3-butandione	50,8	47,8
2-etil-3,5-dimetilpirazina	0,33	0,94

Fig. 15.

Quanto esposto finora giustifica la grande variabilità di gusto che il consumatore ritrova nei vari 'preparati' a base di caffè tostato. La composizione tipica del caffè grezzo (*Fig. 4*) influisce relativamente poco. Determinanti ai fini del gusto sono la tostatura e il modo di preparazione della bevanda. Nel primo processo si forma un numero enorme di composti aromatici volatili, nessuno dei quali possiede le caratteristiche di aroma caratterizzante ('impact flavour'). Per di più, la composizione degli aromi volatili di nun determinato macinato dipende molto dal modo e dalla temperatura di estrazione con acqua che avviene nella preparazione della bevanda. Tuttavia, appare da studi recenti che 'un buon' caffè debba avere concentrazioni elevate di tutti i primi quattro composti della

Fig. 16. Infine, è stato dimostrato che la nota caratteristica del caffè 'all'italiana' sia dovuta alla presenza di concentrazioni elevate di composti fenolici (*Fig. 15*) tipica della varietà *robusta*.

Concentrazione e *valore aromatico* dei principali componenti odorosi dei caffè arabica e robusta tostati

composto	arabica [mg.kg^{-1}]		robusta [mg.kg^{-1}]	
2-furfurilmetantiolo	1,08	$1{,}1{\cdot}10^{5}$	1,73	$1{,}7{\cdot}10^{5}$
β-damascenone	1,195	$9{,}8{\cdot}10^{4}$	0,20	$1{\cdot}10^{5}$
3-mercapto-3-metilbutilformiato	0,130	$3{,}7{\cdot}10^{4}$	0,115	$3{,}3{\cdot}10^{4}$
3-metil-2-butentiolo	0,0082	$2{,}7{\cdot}10^{4}$	0,0083	$2{,}8{\cdot}10^{4}$
2-isobutil-3-metossipirazina	0,08	$1{,}7{\cdot}10^{4}$	0,012	$2{,}4{\cdot}10^{3}$
4-idrossi-2,5-dimetil-3[2H]-furanone	109,0	$1{,}1{\cdot}10^{4}$	57,0	$5{,}7{\cdot}10^{3}$
guaiacolo	4,2	$1{,}7{\cdot}10^{3}$	28,2	$1{,}1{\cdot}10^{4}$
4-vinilguaiacolo	64,8	$3{,}2{\cdot}10^{3}$	177,7	$8{,}9{\cdot}10^{3}$
2,3-butandione	50,8	$3{,}4{\cdot}10^{3}$	47,8	$3{,}2{\cdot}10^{3}$
2-etil-3,5-dimetilpirazina	0,33	$2{,}1{\cdot}10^{3}$	0,94	$5{,}9{\cdot}10^{3}$

Fig. 16.

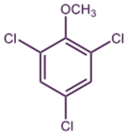
Infine, uno sguardo ai composti che 'spogliano' l'aroma del caffè. Questi sono il tricloroanisolo (odore di terra bagnata), derivato dalla degradazione microbica di fungicidi a base di pentaclorofenolo, e il metilisoborneolo (odore di tappo), metabolita di muffe che infettano il caffè (*Fig. 17*).

'AROMA DEFECTS' del caffè

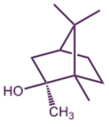
E' abbastanza frequente che intere partite di chicchi siano 'ammalorate' a causa di *'off flavors'* definiti come:

'sacco di juta'
'pisello'
'Rio-taste'

Fra gli *'agenti chimici'* responsabili ci sono:
- il 2,4,6-tricloroanisolo (ODORE DI TERRA BAGNATA) prodotto di degradazione microbica e metilazione di fungicidi a base di pentaclorofenolo.
- il metilisoborneolo, metabolita di muffe (ODORE DI TAPPO)



soglia di percezione in acqua: 3,0.10^{-2} µg/L



soglia di percezione in acqua: 3,0.10^{-2} µg/L

Fig. 17.

Altri componenti maleodoranti del caffè, dotati anche loro di soglie di percezione piuttosto basse sono lo scatolo e la geosmina (*Fig. 18*). Nelle *Fig. 19* e *Fig. 20* sono riportati dati relativi agli effetti fisiologici del caffè, assieme alla bibliografia rilevante sull'argomento.

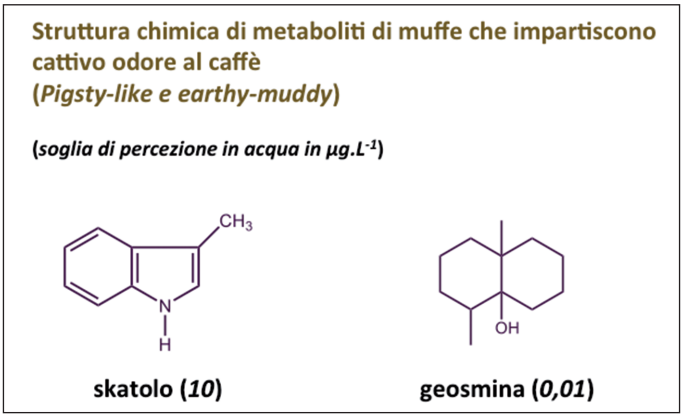


Fig. 18.

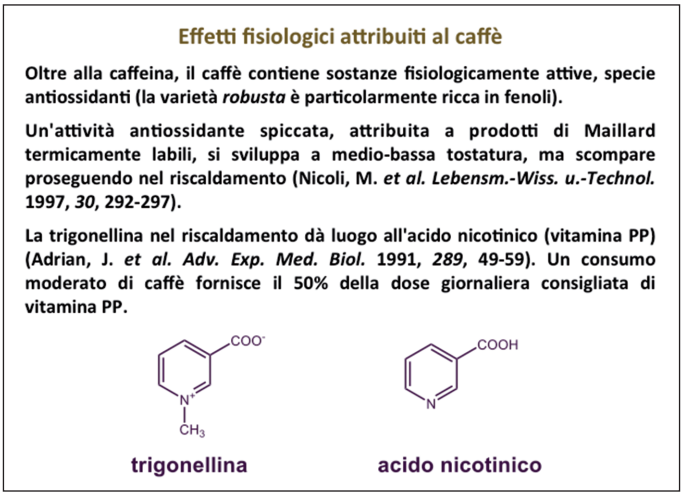


Fig. 19.

Notevole attività antiossidante è stata osservata nel siero umano dopo somministrazione orale di 200 mL di caffè preparato da 12 g di caffè tostato (Natella, F. *et al. J. Agric. Food. Chem*. 2002, *50*, 6211-6216). Questa azione è stata attribuita alla presenza di un prodotto di Maillard ionico (metilpiridinio derivante dalla trigonellina) (Somoza, V. *et al. J. Agric. Food Chem*. 2003, *51*, 6861-6869)

BIBLIOGRAFIA SULL'AROMA DEL CAFFE'

Flament, I. et al., *Chemistry and Industry* **1988**, 592-596
Grosch, W. *Chem. unserer Zeit* **1996**, *30*, 126-133
Petracco, M. *J. Chem. Educ.* **2005**, *82*, 1161-1167
Brenna, E. et al., *J. Chem. Res.* **1998**, 74-75

Fig. 20.

Istituto Lombardo (Rend. Scienze) 146, 311-335 (2012)

AVALANCHE DYNAMICS IN DISORDERED SYSTEMS

ALBERTO ROSSO (*)

Nota presentata dal m.e. Attilio Rigamonti

(Adunanza del 20 dicembre 2012)

SUNTO. Descriviamo la dinamica complessa di un'interfaccia elastica che si muove in un mezzo disordinato e in presenza di una forza esterna. Quando il valore della forza esterna è abbastanza piccolo, il moto dell'interfaccia è lento e intermittente. Due diversi regimi sono osservati. Un primo regime detto di *creep* si presenta per valori della forza esterna molto piccoli. Qui il moto è possibile solo per attivazione termica e l'interfaccia appare per lo più bloccata in profondi stati metastabili separati da grandi barriere energetiche. Un secondo regime detto di *depinning* è caratterizzato dalla presenza di moto anche a temperatura nulla grazie ad una dinamica che procede attraverso riorganizzazioni di grandi porzioni dell'interfaccia dette valanghe. I nostri risultati gettano nuova luce sulla transizione tra i due regimi e sulla statistica delle valanghe.

***

ABSTRACT. We describe the complex dynamics of an interface (like a magnetic domain wall) on a disordered landscape. In presence of a small external drive (*i.e.* an external magnetic filed) the motion of the interface is slow and jerky. Two different regimes can be distinguished: the thermally activated regime where the interface spend most of the time in deep metastable states separated by large barriers (creep regime) and the avalanche regime where motion persists also at zero temperature, but it involves collective re-organizations (depinning). Our results shed light on the interplay between these two regimes and on the avalanche statistics.

(*) Dipartimento di Fisica, Laboratoire Physique Théorique et Modèles Statistiques (UMR CNRS 8626), Université Paris-Sud, Orsay Cedex, France.
E-mail: alberto.rosso@u-psud.fr

1. INTRODUCTION

The growth of interfaces in a disordered medium is a longstanding problem with important applications for a host of experimental systems. Such problems can be split into two broad categories: (i) propagating interfaces such as magnetic domain walls [1, 2], contact line in wetting [3], and crack propagation [4, 5, 40]; (ii) periodic systems such as vortex lattices [7], and charge density waves [8]. In all these systems the basic physical ingredients are identical: the elastic forces tend to keep the structure ordered (flat for an interface and periodic for lattices), whereas the impurities locally promote the wandering (see *Fig. 1*). From the competition between disorder and elasticity emerges a complex energy landscape with many metastable states. This results in glassy properties such as hysteresis and history dependence of the static configuration.

If we consider an interface at equilibrium with the random environment, scaling arguments due to Larkin [9] show that below four (internal) dimensions, displacements grow unboundedly with the distance, resulting in rough interfaces even at zero temperature. The roughness exponent, ζ_{eq}, characterise the spatial wandering of the interface and depends only on the correlation of the disorder, the dimension of the interface, and the nature of the elastic forces.

The competition between disorder and elasticity manifests also in the dynamics of such systems. Among the dynamical properties, the response of the system to an external force f is specially crucial, both from a theoretical point of view and respect to measurements. Indeed, in most systems the velocity v versus force f characteristics is directly measurable and is simply linked to the transport properties (velocity-applied magnetic field for magnetic domain walls, crack velocity-elastic energy release rate G,...). Different dynamical regimes can be observed as shown in *Fig. 2*.

The Creep regime. It is natural to expect that under the action of a small applied force, at zero temperature, the system remains pinned and only moves until it locks on a local minimum of the tilted energy landscape. In this regime, the system is expected to move through thermal activation. What is the nature of this motion and what is the velocity? The first model of such a motion, called TAFF (Thermally Assisted Flux Flow) model, found a linear response [10]. The idea is to consider that a

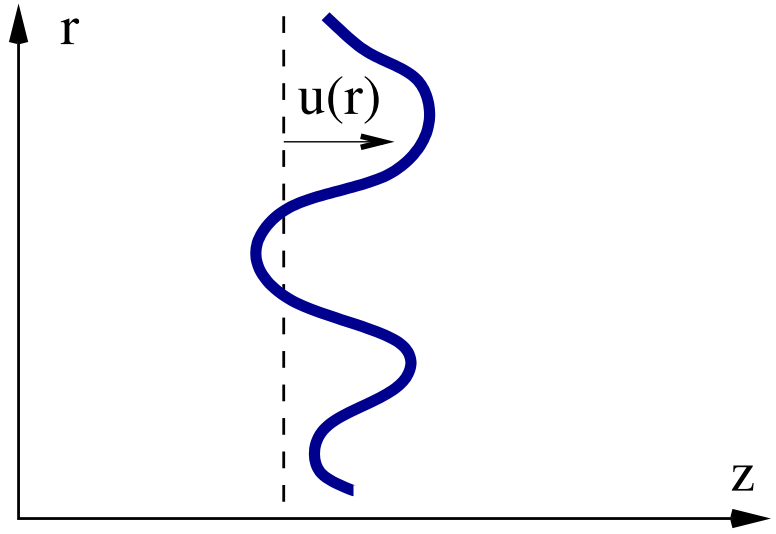


Fig. 1 – A one dimensional interface (domain wall) living in a two dimensional space (film). The position of the interface is determined (provided there are no overhangs or bubbles) by the displacement u from a flat configuration. The roughness of the interface is characterised by the exponent ζ, $(u(r) - u(0)) \sim r^{\zeta}$.

blob of pinned material has to move in an energy landscape with characteristic barriers U_b as shown in *Fig. 3*. The external force f tilts the energy landscape making forward motion possible. The barriers are overcome by thermal activation (hence the name: Thermally Assisted Flux Flow) with an Arrhenius law. If the minima are separated by a distance a the velocity is

$$v \propto e^{-(U_b - fa/2)/T} - e^{-(U_b + fa/2)/T} \simeq e^{-U_b/T} f \qquad (1)$$

The response is thus linear, but exponentially small. However, this argument is grossly inadequate for a glassy system. The reason is easy to understand if one remembers that the static system is in a vitreous state. In such states a characteristic barrier U_b does not exist, since barriers are expected to diverge as one gets closer to the ground state of the system. The TAFF formula is then valid in systems where the glassy aspect is somehow killed and the barriers do saturate. This could be the case for

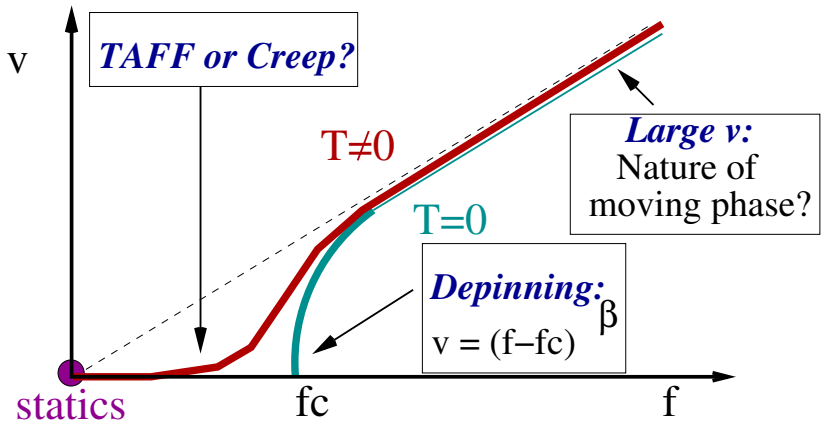


Fig. 2 – The velocity v-f characteristic of a disordered elastic system. At zero temperature the system stays pinned until a critical force f_c is reached. At finite temperature a motion can occur even for forces below threshold $f < f_c$ since the barriers to motion can always be passed by thermal activation.

example for a finite size interface. When the glassy nature of the system persists up to arbitrarily large length scales the motion is actually dominated by barriers which diverge as the drive f goes to zero. To evaluate the size of these barriers, scaling arguments rely on strong assumptions such the use of statics properties to describe an out of equilibrium system [11]. This phenomenological analysis, confirmed by a functional renormalization group calculation [12], leads to the so called creep formula

$$v \sim \exp(-C\, f^{-\mu_{\rm eq}}/T) \tag{2}$$

where $\mu_{\rm eq}$ is an equilibrium critical exponent. This formula has recently been verified by experiments on magnetic domain walls [1, 2].

The depinning regime At larger drive, the system follows the force f, even at zero temperature, and acquires a non-zero asymptotic velocity v. What is the depinning threshold f_c below which the interface is immobile and above which a steady-state motion sets in? An estimate of f_c can be obtained via scaling arguments [9] or computed numerically

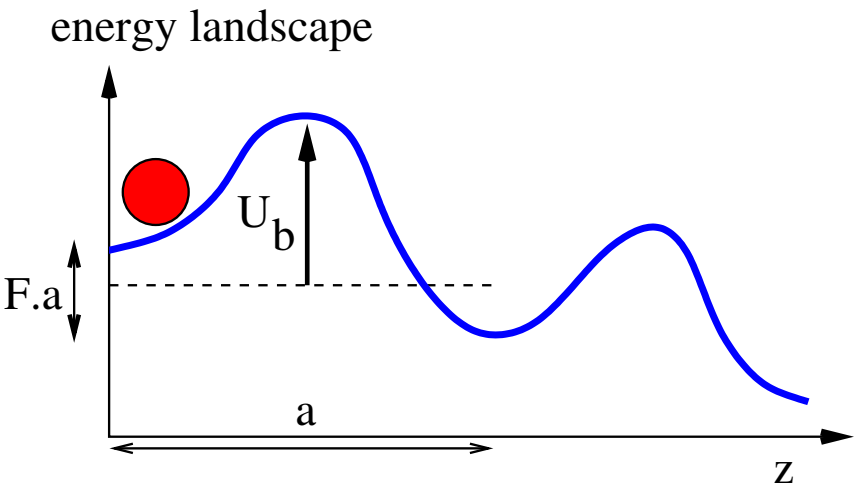


Fig. 3 – Left: In the Thermally Assisted Flux Flow (TAFF) [10] a region of pinned material is considered as a particle moving in an energy landscape characterized by characteristic barriers U_b. This leads to a small but linear response.

by an exact algorithm [13]. Close to f_c the driven dynamics proceeds by discontinuous collective rearrangements called avalanches. The size of these rearrangements grows as we approach to the threshold and, at the transition $f = f_c$, the avalanche statistics is expected to be scale free [14]. In analogy with equilibrium critical phenomena, the scaling behaviour is characterised by universal critical exponents, such as the correlation length exponent, $\nu_{\rm dep}$, related to the linear extension of the avalanches ($\xi(f) \sim |f - f_c|^{-\nu_{\rm dep}}$) or the velocity exponent, $\beta_{\rm dep}$, defined as $v \sim (f - f_c)^{\beta_{\rm dep}}$. The value of these exponents has been determined by analytical techniques such as the functional renormalization group [15, 16] and by numerical approaches [17].

In this paper we discuss the recent advanced in our comprehension of the dynamical phase diagram shown in *Fig. 2*. In particular, in Section 2 we account for a unified description of the dynamics below f_c and in particular we explain how to match the creep regime with the depinning transition [18, 19]. In Section 3 we focus on the depinning transition and characterize the critical force and the avalanche statistics.

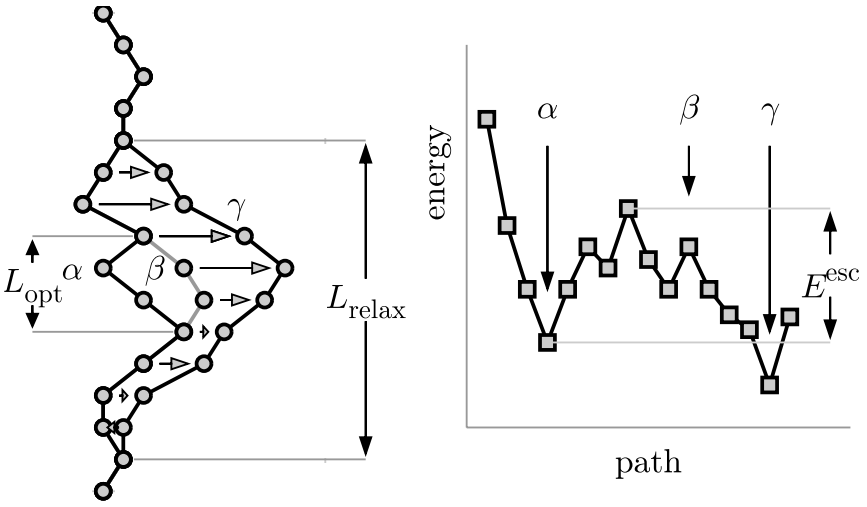


Fig. 4 – Low temperature dynamics of the driven elastic string below the depinning threshold: the optimal path to escape from a given metastable configuration α pass through a saddle configuration β that can relax deterministically to the next metastable configuration γ with $E_\gamma < E_\alpha$.

2. Depinning *vs* creep: the dynamical phase diagram below threshold

2.1 *Creep formula: a divergent length scale when $f \to 0$*

The slow dynamics of the interface for $f \to 0$, so-called creep, is controlled by thermally activated jumps of correlated regions over the pinning energy barriers separating different metastable states. Scaling arguments are based on the assumption that the dynamical barriers of this out-of-equilibrium system can be determined purely from the statics.

In general three different dynamical steps can be isolated (see *Fig. 4*): (i) Starting from a deep metastable state, α, the interface explores the neighborhoods jumping up and down over energy barriers. (ii) This continues until a saddle configuration β is found. (iii) Finally the configuration β can relax deterministically to a new and deeper metastable configuration γ. Configurations α and β differ on a length $L_{\rm opt}$ while α and γ differ on a length $L_{\rm relax}$.

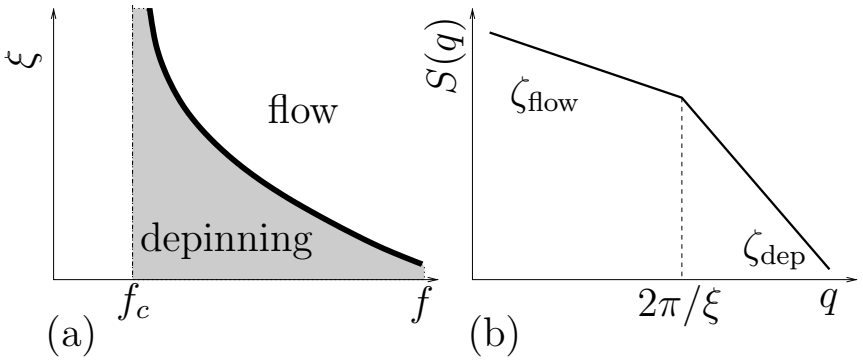


Fig. 5 – (a) Zero-temperature dynamical phase diagram. (b) Schematic structure factor for $f > f_c$. The correlation length ξ fixes the crossover between depinning and fast flow. As an example, for a one dimensional interface with short range disorder we have $\zeta_{dep} \sim 1.25$ and $\zeta_{flow} = 0.5$.

The size of the optimal excitation L_{opt} is obtained by the balance of the gain in energy of being pinned in a deep metastable state, $E_{\text{pinned}}(L)$, with the gain in energy of moving the interface forward, $E_{\text{depinned}} \sim f \cdot (u_\beta(L) - u_\alpha(L))$. At equilibrium we have that $u_\beta(L) - u_\alpha(L) \sim L^{\zeta_{eq}}$ and that $E_{\text{pinned}}(L) \sim L^\vartheta$, with $\vartheta = 2\zeta_{eq} - 1$ (this exact relation is due to the statistical tilt symmetry [20]). The balance gives that

$$L_{opt} \sim f^{-\nu_{eq}} \quad \nu_{eq} = \frac{1}{2 - \zeta_{eq}}, \tag{3}$$

which means that L_{opt} diverges when $f \to 0$.

Moreover, numerical simulations give a clear evidence that the energy landscape is characterized by a unique energy scale, and that the energy difference between neighbor metastable states is equal to the energy barrier separating them [21]. Thus, we expect that the barrier that the interface has to overcome to escape from the state α grows as $L_{opt}^\vartheta \sim f^{-\frac{2\zeta_{eq}-1}{2-\zeta_{eq}}}$. Using the Arrhenius activation, we recover the creep formula Eq. (2) with $\mu = (2\zeta_{eq} - 1)/(2 - \zeta_{eq})$.

2.2 The depinning transition: a divergent length scale when $f \to f_c$

Above the critical force and at zero temperature the analogy of the depinning transition with critical phenomena is well understood [14]: As in a thermodynamic second-order phase transition, the connected two-point correlation function of the order parameter is characterized by a correlation length which diverges at the critical point. The steady-state velocity is the order parameter of the depinning transition and its two-point correlation function,

$$\overline{\langle (v(x,t)-v)(v(0,t)-v)\rangle} \sim e^{-|x|/\xi}, \tag{4}$$

indeed diverges for $f \to f_c^+$. In practice, this correlation function is not easily accessible because the steady-state velocities can be obtained only after a long-time integration of the equation of motion [22].

The correlation length ξ separates two length scales in the manifold (see *Fig. 5a*): on scales smaller than ξ, the geometry, i. e. the *roughness*, of the interface is characterized by depinning exponents (the exponents of the critical phase in the language of magnetic transitions). In contrast, on length scales larger than ξ, the interface is governed by the fast flow exponents (analogous to the exponents of the ferromagnetic ordered phase). The length ξ can be measured through the power spectrum or structure factor [23],

$$\begin{aligned} S(q) &= \overline{\left\langle \left| \frac{1}{L^{1/2}} \int dx\; u(x,t)\; e^{-iqx} \right|^2 \right\rangle} \\ &= \int dx\; e^{-iqx}\; \overline{\langle u(x,t)u(0,t)\rangle}, \end{aligned} \tag{5}$$

where the second equality makes use of spatial translation invariance. For inverse lengths q belonging to a self-affine regime with a single roughness exponent ζ, the structure factor takes the form

$$S(q) \sim q^{-(1+2\zeta)}.$$

The crossover between the depinning and the fast-flow regimes can be conveniently extracted from the change of slope of $S(q)$ (see *Fig. 5b*).

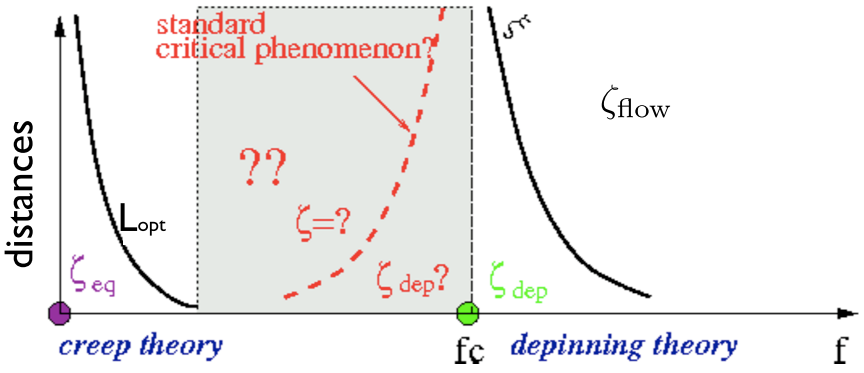


Fig. 6 – Dynamical phase diagram, the geometry of the interface in the steady state regime. Below threshold we expect two divergent correlation length: the activation size $L_{opt}(f)$ which diverges as $f \to 0$ and the distance from consecutive metastable states $\xi(f)$ which diverges as $f \to f_c$. Below the scale L_{opt} we expect that the system is at equilibrium, below the scale $\xi(f)$ the system is at the depinning critical point. How behaves the system at larger length scales?

Below the depinning threshold f_c, at zero temperature, the manifold is permanently pinned. The interpretation of depinning as a critical phenomenon suggests that a similar diverging length scale should also exists below f_c. Indeed, in standard critical phenomena, for length scales smaller than the correlation length the system is critical, and crosses over beyond this correlation length to the broken symmetry phase on one side of the transition and to the symmetric phase on the other side. An important question is thus whether an equivalent correlation length is observable in the limit $f \to f_c^-$.

A classical protocol to study the depinning transition below threshold is defined as follows: we prepare the system with an initial non-steady state configuration, for example the flat configuration, and then we apply an external force $f < f_c$. The interface moves forward and re-organize on a length scale $\xi(f)$ which diverges when $f \to f_c$. We will see in Section 3 that this collective re-organization is a manifestation of the avalanche dynamics in these systems. The geometry of the line is governed by the scale ξ: on scales smaller than ξ, the *roughness* of the interface is characterized by depinning exponents. In contrast, on length scales larger than ξ, the interface is reminiscent of the non-steady state

initial condition (in our example it will be flat). This means that, below threshold, a correlation length $\xi(f)$ exists in the transient regime. It is natural to ask if it is possible to study the steady state regime below threshold. If yes, how behaves the system at large length scales? In analogy with what happens above threshold one should expect that, below threshold, the large scale behavior is described by equilibrium exponents of the $f = 0$ fixed point. However, as shown in *Fig. 6* this does not match with the creep scaling arguments, for which $f = 0$ is unstable fixed point. In others words, for the creep scenario, when $f > 0$, the interface is at equilibrium only at length scales shorter than $L_{\rm opt}(f)$.

2.3 Dynamical phase diagram for $T \to 0^+$

In [18, 19] we studied the steady state regime of an interface in the limit of vanishing temperature. The temperature allows to forget the initial condition and reach the stationary state. A direct numerical study of the motion of such systems at low temperature is very difficult for forces below the depinning force. Indeed, in that case, the motion takes place by thermal activation over barriers leading to extremely long activation times. In this regime numerical techniques such as the molecular dynamics are inefficient. For this reason we introduced a novel numerical method which allows to follow the motion of an interface at finite temperature, without running into the above-mentioned difficulties. This method directly implement the dynamics of *Fig. 4*. We have proved theorems that assure that this dynamics is the correct low temperature dynamics for a finite interface.

Our results are summarized in *Fig. 7*. The crossover length, L_{opt}, separates two roughness regimes: On length scales smaller than L_{opt} the roughness of the interface is described by the equilibrium exponent ζ_{eq}, corresponding to the paramagnetic phase in the language of magnetic transitions. For distances bigger than L_{opt}, the roughness is described by depinning exponents. This is at variance with standard critical phenomenon, where the critical phase at large length scales appears only at the critical point. Increasing the external force, *Lopt* decreases and when the depinning threshold is reached ($f \to f_c^-$), L_{opt} becomes of the order of the lattice space. The depinning divergent length scale $\xi(f)$ can be identified with the characteristic length L_{relax} of the creep regime.

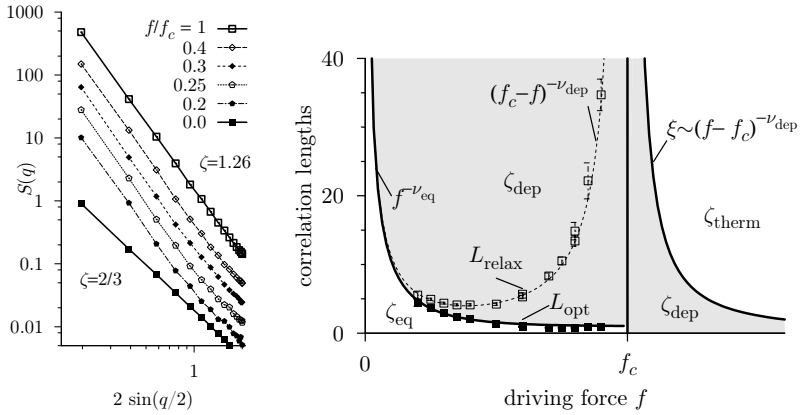


Fig. 7 – Results: (a) Steady state structure factor of the line in the $T \to 0$ limit, averaged on 1000 *samples. (a) $S(q)$ for $L = 32$, $M = 92$ and different forces (curves are shifted for clarity). (b) Summary: The steady state properties of the elastic string at $T \to 0$ are determined by L_{opt} (filled symbols) and by ξ. They separate regions characterized by the equilibrium exponent, the depinning exponent (gray region), and the thermal or flow one. The divergent length L_{relax} (open symbols) is associated only with transient dynamics. Lines are guides to the eye.*

In the creep $L_{relax} \sim L_{opt}$, at the depinning transition L_{relax} diverges, while L_{opt} remains finite.

Our algorithm cannot access very small forces, but our results are compatible with predictions in the creep regime. In particular, the functional renormalization group [12] predicts that on scales below L_{opt} the system is in equilibrium and that scales larger than L_{opt} are characterized by deterministic forward motion.

3. AVALANCHE DYNAMICS

The driven dynamics of an elastic interface proceeds by discontinuous collective rearrangements. We have introduced two different type of rearrangements: the activated rearrangements of scale L_{opt} and responsible of the creep regime, and the deterministic re-arrangements of scale L_{relax} and called avalanches. Here we focus on avalanches.

Avalanches are observed in a number of systems. Some examples are the Barkhausen noise in magnets [24, 25], dislocations and crack

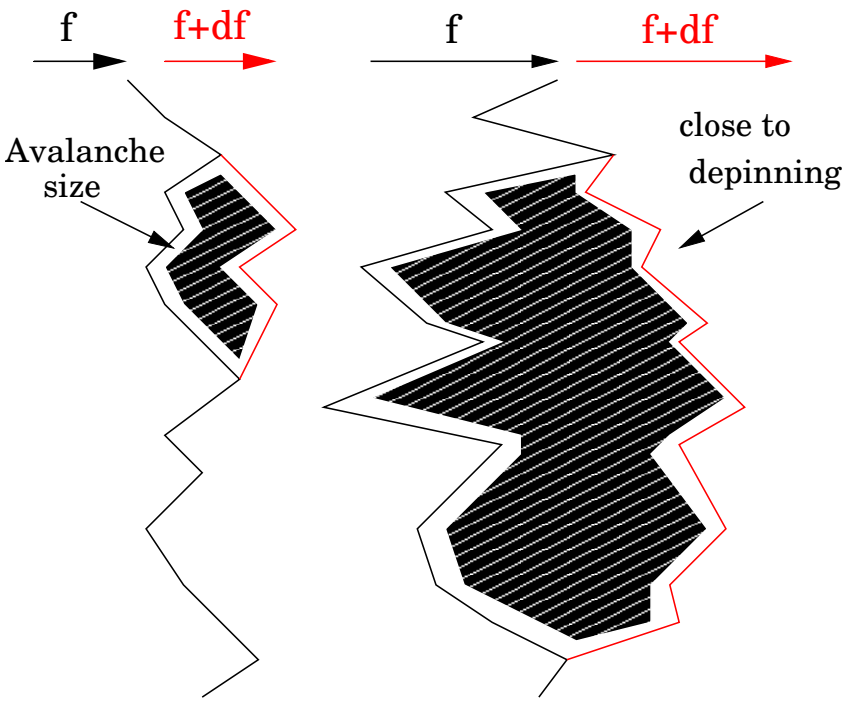


Fig. 8 – Consider an interface pinned in a metastable state in the presence of a force f. If we apply an additional infinitesimal force $f \to f + \delta f$ the metastable state can become unstable and re-organize on very large length scale. The characteristic (maximal) size of the avalanche is called S_m and diverges when we approach the depinning transition.

propagation [26, 27], and stick-slip motion of tectonic plates (responsible for earthquakes) [28, 29]. Avalanches have also been studied in models without quenched substrate disorder, such as sandpile models and granular matter [30, 31].

Important characteristics of avalanche motion are its scale invariance, self-organized criticality, and a broad distribution $P(S) \sim S^{-\tau}$ of the avalanche size S, for sizes S between small- and large-scale cutoffs $S_{\min} \ll S \ll S_m$. Pinned elastic manifolds are an important prototype for a much wider class of phenomena, exceeding physics and reaching for example economy and finance, where extreme (and sometimes catastrophic) events are sufficiently frequent and large to dominate most

observables. In this context, it is clearly of importance to understand how the avalanche-size probability is cutted off at large scales, $S > S_m$.

Although avalanche motion of pinned manifolds has been studied for a while [32, 33, 34], most work was focused on measuring the avalanche-size exponent τ. A scaling argument proposed by Narayan and Fisher [35] on the assumption that the avalanche density remains finite at the depinning threshold, relates the exponent τ with the roughness exponent ζ via:

$$\tau = 2 - \frac{2}{d+\zeta} \,. \tag{6}$$

Here ζ is the roughness exponent at the depinning transition. On the other hand the cut-off of large avalanches depend either on finite size effects either on the value of the external force. In the latter case scaling arguments give:

$$S_m \sim L_{relax}^{d+\zeta} \sim |f - f_c|^{-\frac{d+\zeta}{2-\zeta}}. \tag{7}$$

To describe the full statistics of avalanches we have to specify how the dynamics is implemented for a system of finite size L. An intuitive protocol, used in experiments of Barkausen noise, is shown in *Fig. 8* where the interface is prepared in a metastable state at force $f = 0$ and force is slowly increased up to the critical value f_c. Avalanches obtained using this method are not identically distributed because the cut-off S_m depends on the distance to the critical point which is not constant during the evolution.

In our works we have replaced the external force f with a an external parabolic potential to provide a proper definition of the steady state avalanche dynamics. To be more precise, the zero-temperature Langevin dynamics is described by the equation of motion:

$$\begin{aligned} \partial_t u(x,t) &= \mathcal{F}_{w(t)}(x, u(x,t)) \\ \mathcal{F}_w(x, u(x)) &= m^2(w - u(x)) + c\nabla^2 u(x) + F(x, u(x)), \end{aligned} \tag{8}$$

where $\mathcal{F}_w(x,u)$ is the total force acting on the manifold, c is the elastic constant and $F(x,u)$ is the random pinning force. The parabolic potential is centred in w and its curvature, m^2, acts as a mass for the field

u. For random field (RF) disorder, $F(x,u)$ is short-ranged with correlations $\overline{F(x,0)F(x',u)} = \Delta_0(u)\delta^d(x-x')$. For random bond (RB) disorder this force is derived from a short-ranged random potential $V(x,u)$, $F(x,u) = -\partial_u V(x,u)$.

Starting from an arbitrary initial condition $u_{\text{init}}(x)$, and giving $w = w_0$, the manifold moves to a locally stable state $u_{w_0}(x)$, i.e., a zero-force state $\mathcal{F}_{w_0}(x, u_{w_0}(x)) = 0$ which is stable to small deformations. Increasing w, $u(x)$ increases slightly, while the configuration remains stable. At some $w = w_1$, the state becomes unstable and the manifold moves until it is blocked again in a new locally stable state $u_{w_1}(x)$. We are interested in the centre of mass displacement

$$u(w) = L^{-d} \int dx\, u_w(x). \tag{9}$$

The function $u(w)$ exhibits jumps at discrete values of w and is in general dependent on the initial condition. However, due to the no-passing rule [36], we can prove that there exists a $w^* > w_0$ such that the orbit $u_{w>w^*}(x)$ becomes independent of the initial condition $u_{\text{init}}(x)$, and w_0. A stationary state is thus reached after a finite $w - w_0$, on which we focus. In *Fig. 9* we show how the algorithm works, for dw very small we are able to isolate a single avalanche event.

3.1 Critical force and autocorrelation function

It is useful to introduce the process [37]:

$$x(w) = m^2\,[w - u(w)]. \tag{10}$$

For a given location of the parabolic well, $x(w)$ represents the average force on the elastic interface in presence of a parabolic well. This force should is compensated by the presence of the impurity which pinned the system. The critical force for this system is thus given by:

$$f_c(m) = \overline{x(w)} = m^2\,\overline{[w - u(w)]} \tag{11}$$

when $m \to 0$ we have $f_c(m) \to f_c$ ($f_c(m)$ is independent of L if $Lm \gg 1$). In *Fig. 10* we show a typical realization of the process $x(w)$.

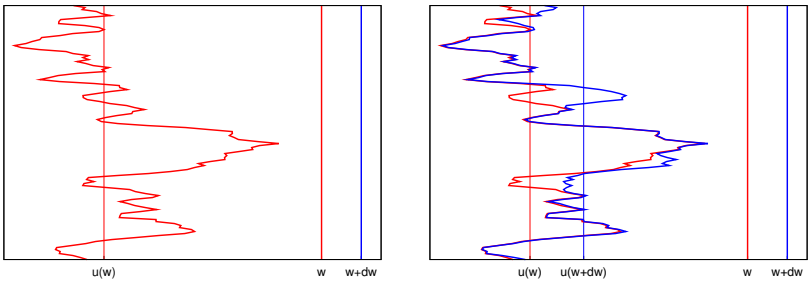


Fig. 9 – For a given w the manifold moves to a metastable state $u_w(x)$, i.e. a state dynamically stable to infinitesimally small deformations. Then one increases w, and a smooth forward deformation of $u_w(x)$ results (for smooth short-scale disorder) while the state remains stable. At some $w = w_1$ the state becomes unstable and the manifold, for $w = w_1^+$ moves until it is blocked again in a new metastable state $u_{w_1}(x)$ (also locally stable). This process is called an avalanche and the centre of mass $u(w)$ has a jump located at $w = w_1$.

We see that, due to the presence of avalanches, this process display vertical jumps and fluctuation around the mean value. The statistical tilt symmetry assures that the quadratic part of the hamiltonian (and thus the mass m) is not renormalized. This means that the length associated to the bare mass $L_m = 1/m$ fixes the distance from the critical point (located at f_c) as $m^{1/\nu} = m^{2-\zeta}$. The finite size effects on the force takes then the form:

$$f_c(m) = f_c + c_1 m^{2-\zeta} + \dots . \tag{12}$$

The fluctuations around this value depend on L and m. In the limt $L \gg L_m = 1/m$, the interface can be modelled as a collection of independent interfaces of size L_m and the central limit theorem assures that fluctuations should decrease as $\sqrt{L_m/L}$. On the other hand using that the distance from the critical point is $m^{1/\nu} = m^{2-\zeta}$ and that w scales as L_m^ζ we can write

$$x(w) = f_c(m) + \frac{m^{1/\nu}}{\sqrt{L/L_m}} y(w/L_m^\zeta) = f_c(m) + \frac{m^{3/2-\zeta}}{\sqrt{L}} y(w m^\zeta), \tag{13}$$

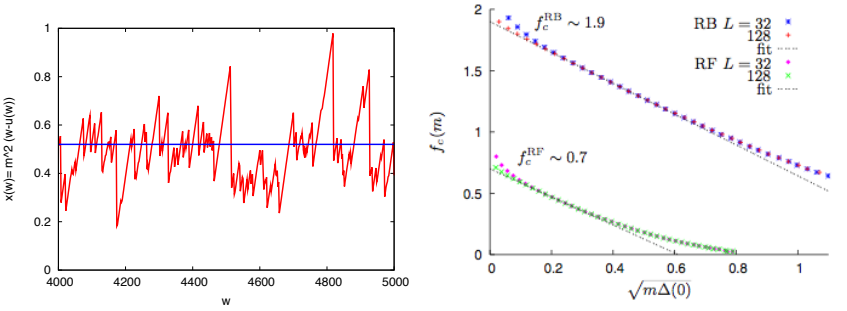


Fig. 10 – Left: Typical realization of the process $x(w)$. Right: Critical force, finite size effect study.

where the rescaled process $y(z)$ is independent of m and L as $m \to 0$ and $Lm \gg 1$. Approaching the critical point $y(z)$ displays universal behavior and becomes independent of most microscopic details. Using the set up with the parabolic well we can measure the autocorrelation function of $x(w)$:

$$m^4 \overline{[w - u(w)][w' - u(w')]}^c = L^{-d} \Delta_m(w - w').$$
$$\Delta(w) = m^{3-2\zeta} \tilde{\Delta}(w m^{\zeta}) \tag{14}$$

where $\Delta_m(w) \to \Delta(w)$ as $m \to 0$ and $\tilde{\Delta}(z)$ is the autocorrelation function of the process $y(z)$.

We have studied the behavior of the critical force $f_c(m)$ for two classes of disorder the RF and the RB. From (14) one has $\sqrt{\Delta(0)m} \sim m^{2-\zeta}$, yielding a parameter-free linear scaling shown in *Fig. 10* [37]. For large m the scaling is non-linear, while for smaller m it is linear up to the scale where the correlation length becomes of the order of L ($mL \approx 5$). The critical force of the infinite system is defined here in an unambiguous way, as $f_c = f_c(m = 0)$. One can see on *Fig. 10* that the slope, c_1, of the two curves coincides. Indeed, c_1 is a universal amplitude, depending on microscopic details only through the renormalized elastic coefficient c; here $c \approx 1$ for both models of *Fig. 10*.

3.2 *Avalanche statistics*

We can now study in detail the avalanche statistics [38]. The size S of an avalanche is defined as the area swept by the line as it jumps between the two consecutive metastable states. The distribution of avalanche sizes is expected to exhibit universality, i.e. independence of short scale details, for sizes $S > S_{\min}$. The short-scale cutoff $S_{\min}$ corresponds to the area spanned by a single monomer on the scale of the discretization of the disorder (in our units $S_{\min} \simeq 1$). We define the (normalized) distribution of avalanche sizes $P(S)$, as well as its moments

$$\langle S^n \rangle = \frac{1}{N} \sum_{i=1}^{N} S_i^n = \int_0^\infty dS\, S^n P(S) \tag{15}$$

from the sequence of measured avalanches S_i, $i = 1, \ldots, N$.

In the limit $m = 0$ a critical point is reached, resulting in a power-law distribution of avalanche sizes. To properly define the problem, including the stationary measure, it is essential to consider a small $m > 0$. Then, the correlation length $L_m = 1/m$ is large in the small-m regime considered here. As a result, the distribution of avalanche sizes is cut-off by the large scale $S_m \gg S_{\min}$. Using the setting with a parabolic well the large scale cut off is expected to scale as $S_m \sim L_m^{d+\zeta} \sim m^{-d-\zeta}$ at small m. It is useful to express S_m in terms of the moments $\langle S^n \rangle$. The definition depends on the value of τ , the critical exponent of avalanche size statistics.

- For $\tau < 1$ $S_{\min} \to 0$, $\langle S \rangle \sim S_m$
- For $\tau > 1$ $S_{\min}$ must be finite in order to have finite moments. In particular the moments write $\langle S^n \rangle \sim S_m^{n+1-\tau} S_{\min}^{\tau-1}$, so that we can define

$$S_m := \frac{\langle S^2 \rangle}{2\langle S \rangle} \,. \tag{16}$$

The scale S_m is important as it allows to define universal functions. In the variable $s := S/S_m$ avalanche statistics becomes universal. Nevertheless, if the exponent τ satisfies $2 > \tau > 1$, which is the case here,

then the distribution of avalanche should always depend on $S_{\min}$ which assures a proper normalization.

A fully universal quantity is the function $p(s)$, defined via the relation:

$$P(S)dS = \frac{\langle S\rangle}{S_m} p\left(\frac{S}{S_m}\right)\frac{dS}{S_m}\,. \tag{17}$$

The function $p(s)$ is universal and depends only on the space dimension d. Note that the normalized probability $P(S)$ depends on the cut-off $S_{\min}$ via the first moment $\langle S\rangle$ which cannot be predicted by the theory, hence is an input from the numerics. It is important to stress that while the function $p(s)$ is universal and convenient for data analysis, *it is not a probability distribution* and it is not normalized to unity. Rather, it satisfies from its definition (17) and using (16), the two normalization conditions

$$\langle s\rangle_p = \int ds\, sp(s) = 1 \tag{18}$$

$$\langle s^2\rangle_p = \int ds\, s^2 p(s) = 2\,. \tag{19}$$

A practical algorithm to compute $p(s)$ from a sequence $S_1, S_2, \ldots$ of avalanches is defined as:

- compute $\langle S\rangle$ and $\langle S^2\rangle$
- compute $S_m = \frac{\langle S^2\rangle}{2\langle S\rangle}$
- compute $s_1 = S_1/S_m, S_2/S_m, \ldots s_n = S_n/S_m$
- make the histogram of the s_i
- multiply the heights of the histograms by $\frac{\langle S^2\rangle}{2\langle S\rangle^2} = \frac{S_m}{\langle S\rangle}$

The rescaled avalanche-size distribution can be written as

$$p(s) = s^{-\tau} f(s) \tag{20}$$

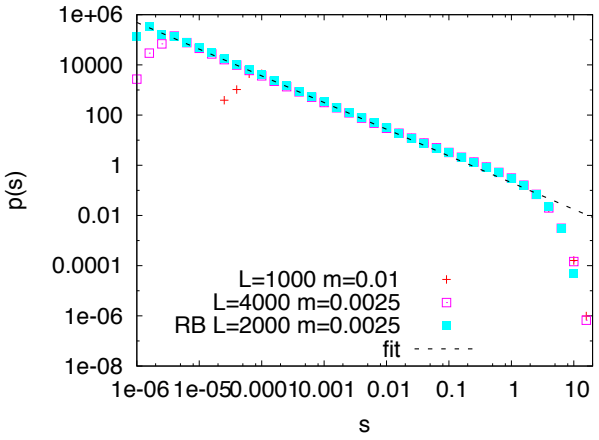


Fig. 11 – The function $p(s)$ for RF and RB disorder ($d = 1$). A fit with a power law gives the exponent $\tau = 1.08 \pm 0.02$. The agreement with Eq. (6) is discussed in the text.

where τ is the avalanche-size exponent, and $f(s)$ the universal cut-off function[1], which tends to a constant for $s \to 0$.

For the present model, the only analytical prediction concerns the exponent τ, via the above mentioned conjecture [35] of Eq.(6). For $d = 1$, a direct power-law fit of our numerical data gives

$$\tau_{\text{num}}^{d=1} = 1.08 \pm 0.02 \,. \tag{21}$$

This value has to be compared with the conjecture of Eq. (6). The roughness exponent is known numerically with a good accuracy from system sizes ($L \sim 10^3$) and $m = 0$, as $\zeta = 1.26 \pm 0.01$ [17]. This value for ζ gives $\tau_{\text{conj}} = 1.115 \pm 0.005$. Hence the estimate (21) is slightly smaller, but still consistent with the value of τ obtained from the conjecture.

Exact solution of a mean field toy model of avalanches allows to compute the full universal scaling function:

$$f_{\text{MF}}(s) = \frac{1}{2\sqrt{\pi}} e^{-s/4} \,. \tag{22}$$

[1] By universal we mean w.r.t. short scale details. Of course this function is characteristic of a large scale cut-off provided by a parabolic well.

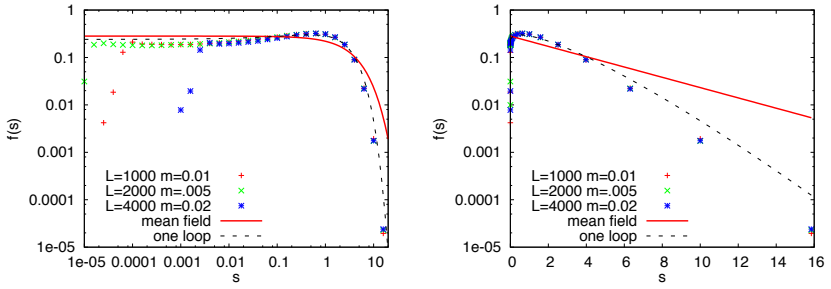


Fig. 12 – Function $f(s)$ for RF disorder ($d = 1$). The red solid curve is given by Eq.(22), the black dashed line by Eq.(23), with $A = 0.852$, $B = 1.56$ and $C = 0.56$. Left: Blow up of the power-law region. Right: Blow up of the tail region.

which is expected to be exact above the upper critical dimension $d = d_{\rm uc} = 4$. The Functional Renormalization Group calculation gives a finite dimension correction to the mean field result, at the first order in the $\varepsilon = 4 - d$ expansion we have the following prediction [39]

$$f(s) = \frac{A}{2\sqrt{\pi}} \exp\left(C\sqrt{s} - \frac{B}{4} s^{\delta} \right) , \tag{23}$$

with exponents

$$\tau = \frac{3}{2} + \frac{3}{8}\alpha = \frac{3}{2} - \frac{1}{8}(1 - \zeta_1)\varepsilon \tag{24}$$

$$\delta = 1 - \frac{\alpha}{4} = 1 + \frac{1}{12}(1 - \zeta_1)\varepsilon \tag{25}$$

where $\alpha = -\frac{1}{3}(1 - \zeta_1)\varepsilon$ and $\zeta_1 = 1/3$ for the RF class, relevant to the present study. The constants A, B and C depend on ε, and must satisfy the normalization conditions (18), (19). At first order in ε they are $C = -\frac{1}{2}\sqrt{\pi}\alpha$, $B = 1 - \alpha(1 + \frac{\gamma_{\rm E}}{4})$, $A = 1 + \frac{1}{8}(2 - 3\gamma_{\rm E})\alpha$, $\gamma_{\rm E} = 0.577216$. As usual, the one-loop results for the exponents τ, δ and for the parameters A, B, and C are exact up to $O(\varepsilon^2)$.

Our numerical data for the avalanche-size distribution $p(s)$ are plotted in *Fig. 12*, with emphasis either on the power-law region or on the tail. Note that for the different values of m and L used here, the

data have converged, with the exception of the last point for very large avalanches (over-suppressed by the finite size of the interface in the smallest samples), and the region of very small avalanches (which are cut off at $s \approx 1/S_m$).

4. CONCLUSIONS

In Conclusion we have discussed the results on two aspects of the dynamics of an elastic interface in random medium: the activated dynamics or creep, and the avalanche dynamics or depinning.

Despite its relevance for experiments and the theoretical efforts made, the description of the creep regime remains a very challenging problem. Most of the few predictions proposed in litterature are based on strong assumptions that should be verified, at least numerically. In particular it should be shown clearly that a small drive does not change the short scale properties of the interface that remains locally in equilibrium even when it is driven in an out of equilibrium steady state. Verifications are needed not only for the assumptions of creep theory, but for the predictions too. It is true that the main prediction, i.e. the creep formula for the velocity, has received an important confirmation with the experiment of magnetic domain wall [1], but on one side this is the unique experiment that clearly shows a creep regime and on the other side numerical simulations are still far to be conclusive. New algorithms have been developed to capture the secrets hidden in the deep metastable states of the disorder landscape. We are actively working in this direction.

The depinning transition has been much more understood, but different directions are open for future researches. Let me mention some directions of my present research activity:

- The description of the universal behavior is of a little help for the estimation of the threshold above which the interface moves even at zero temperature. In fracture the estimation of this threshold should give information about the toughness of the material [40]. This should be true at least for fragile materials for which plasticity plays a minor role. If this is true, to understand how the threshold behaves as a function of the impurity concentration, the impurity

size and the elastic moduli of the material, will allow to control the toughness of a family of materials. This is a major task in fracture theory.

- In a real system, the sequence of avalanches as a function of time display strong correlation in time. In particular the phenomenological Omori law, verified for the seismic activity, predicts power law correlations. Unfortunately an elastic interface in the steady state regime shows a quasi-Poissonian statistics with absence of correlations, both in time and in intensity. Which is the missing ingredient in the model? Recently, a model reproducing aftershock statistics has been proposed [29]. This model incorporates an internal degree of freedom which relax the contact between the disordered system and the elastic interface. We studied a simple model of viscoelastic interface that reproduce a large part of the observations of the earthquakes statistics [41]

- Collective re-organizations are also observed in the dynamics of amorphous solids under shear deformation [42, 43]. Here, when a site slips, it triggers a shock wave which in turn can either stabilize or destabilize other sites. This is at variance with what is observed in the avalanches discussed so far, where an unstable site can only destabilize other sites. We have recently introduced a scaling description of the yielding transition in term of universal exponents [44].

AKNOWLEDGEMENTS

We acknowledge very fruitful collaborations with T. Giamarchi, A. Kolton, P. Le Doussal, W. Krauth and K. Wiese.

REFERENCES

[1] S. Lemerle *et al.*, Phys. Rev. Lett. 80, 849 (1998).

[2] V. Repain *et al.*, Europhys. Lett. 68, 460 (2004).

[3] S. Moulinet, C. Guthmann, and E. Rolley, Eur. Phys. J. E 8, 437 (2002).

[4] E. Bouchaud, J.-P. Bouchaud, D.S. Fisher, S. Ramanathan, J.R. Rice, Journal of the Mechanics and Physics of Solids 50, 1703 (2002).

[5] D. Bonamy, E. Bouchaud, Physics Reports 498, 1-44 (2011).

[6] L. Ponson, Phys. Rev. Lett. 103, 055501 (2009).

[7] G. Blatter *et al.*, Rev. Mod. Phys. 66, 1125 (1994).

[8] G. Grüner, Rev. Mod. Phys. 60, 1129 (1988).

[9] A. I. Larkin and Y. N. Ovchinnikov, J. Low Temp. Phys 34, 409 (1979).

[10] P. W. Anderson and Y. B. Kim, Rev. Mod. Phys. 36, 39 (1964).

[11] L. B. Ioffe and V. M. Vinokur, J. Phys. C 20, 6149 (1987).

[12] P. Chauve, T. Giamarchi, and P. Le Doussal, Phys. Rev. B 62, 6241 (2000).

[13] A. Rosso, W. Krauth, Phys. Rev. E 65, 025101(R) (2002).

[14] D. S. Fisher, Phys. Rev. B 31, 1396 (1985).

[15] T. Nattermann, S. Stepanow, L. H. Tang, and H. Leschhorn, J. Phys. (Paris) 2, 1483 (1992).

[16] O. Narayan and D. Fisher, Phys. Rev. B 48, 7030 (1993).

[17] A. Rosso, A. K. Hartmann, and W. Krauth, Phys. Rev. E 67, 021602 (2003).

[18] A. B. Kolton, A. Rosso, T. Giamarchi, W. Krauth, Physical Review Letters 97, 057001 (2006).

[19] A. B. Kolton, A. Rosso, T. Giamarchi, W. Krauth, Phys. Rev. B 79, 184207 (2009).

[20] U. Schulz, J. Villain, E. Brézin, and H. Orland, J. Stat. Phys. 51, 1 (1988).

[21] B. Drossel, M. Kardar, Phys. Rev. E 52, 4841 (1995).

[22] O. Duemmer, W. Krauth, Phys. Rev. E 71, 61601 (2005).

[23] A. B. Kolton, A. Rosso, T. Giamarchi, Phys. Rev. Lett. 94, 047002 (2005).

[24] J. S. Urbach, R. C. Madison and J. T. Markert, Phys. Rev. Lett. 75 (1995) 276–279.

[25] JP. Sethna, KA. Dahmen and CR. Myers, Nature 410 (2001) 242–250.

[26] P. Moretti, M.C. Miguel, M. Zaiser and S. Zapperi, Phys. Rev. B 69 (2004) 214103.

[27] D. Bonamy, S. Santucci and L. Ponson, Phys. Rev. Lett. 101 (2008) 045501.

[28] D.S. Fisher, K. Dahmen, S. Ramanathan and Y. Ben-Zion, Phys. Rev. Lett. 78 (1997) 4885–4888.

[29] E. A. Jagla and A. B. Kolton, arXiv: 0901.1907 (2009).

[30] C. Tang and P. Bak, Journal Of Statistical Physics 51 (1988) 797–802.

[31] D. Dhar, Physica A 369 (2006) 29–70.

[32] O. Narayan and AA. Middleton, Phys. Rev. B 49 (1994) 244–256.

[33] F. Lacombe, S. Zapperi and H.J. Herrmann, Phys. Rev. B 63 (2001) 104104 (7 pages).

[34] A. Tanguy, M. Gounelle and S. Roux, Phys. Rev. E 58 (1998) 1577–90.

[35] O. Narayan and D.S. Fisher, Phys. Rev. B 48 (1993) 7030–42.

[36] A. A. Middleton, Phys. Rev. Lett. 68, 671 (1992).

[37] A. Rosso, P. Le Doussal and K.J. Wiese, Phys. Rev. B 75 (2007) 220201, cond-mat/ 0610821.

[38] A. Rosso, P.Le Doussal, K.J. Wiese, Phys. Rev. B 80, 144204 (2009).

[39] P. Le Doussal and K.J. Wiese, arXiv: 0812.1893 (2008).

[40] V. Demery, L. Ponson, A. Rosso, EPL, 105, 34003, (2014).

[41] E. A. Jagla, F. Landes, A. Rosso, Phys Rev Lett. 174301 (2014).

[42] P. Moretti, B. Cerruti, M.C. Miguel, PLoS One 6 (6), e20418 (2011).

[43] M. Talamali, V. PetŠjŠ, D. Vandembroucq, S. Roux, PRE 84 016115 (2011).

[44] J. Lin, E. Lerner, A. Rosso and M. Wyart, PNAS (2014)

GABA RECEPTORS AND PAIN REGULATION

VALERIO MAGNAGHI, MARCELLA MOTTA (*)

Nota presentata dal m.e. Marcella Motta
(Adunanza del 20 dicembre 2012)

SUNTO. – Il dolore è una risposta dell'organismo ad un insulto, esterno o interno. La sua funzione è quella di avvertire e proteggere l'organismo da ulteriori danni. L'Associazione Internazionale per lo Studio del Dolore ha definito il dolore come "esperienza sensoriale o emozionale spiacevole, associata ad un potenziale o reale danno tissutale o descritto come tale". Il dolore neuropatico è un tipo di dolore cronico che origina direttamente da un danno al sistema nervoso stesso, centrale o periferico. E' un tipo di dolore a lungo termine, con alta incidenza e di difficile trattamento farmacologico. I meccanismi fisiopatologici alla base del dolore cronico hanno identificato i neuroni sensoriali afferenti come il bersaglio primario per la terapia. Tuttavia, anche le cellule gliali centrali o periferiche (ad esempio le cellule di Schwann) sembrano essere fondamentali per il controllo del dolore. I circuiti nervosi, che coinvolgono l'acido γ-aminobutirrico (GABA) e i suoi recettori GABA-B, sono stati a lungo considerati in studi preclinici di dolore acuto e cronico. I recettori GABA-B coinvolti nelle vie del dolore sono recettori sopraspinali e spinali, presinaptici e postsinaptici. I recettori GABA-B sono presenti anche nei gangli dorsali spinali e negli assoni periferici. Di recente si è dimostrato che i recettori GABA-A e GABA-B sono funzionalmente attivi anche sulle cellule di Schwann. L'attivazione dei recettori GABA-A controlla la sintesi di GABA, attivando un meccanismo autocrino che regola la mielinizzazione periferica operata dalle cellule di Schwann. L'attivazione dei recettori GABA-B, invece, induce meccanismi autonomi e non-autonomi delle cellule di Schwann. I recettori GABA-B controllano la struttura e l'espressione delle proteine mieliniche. Inoltre, i recettori GABA-B inducono modificazioni indirette del compartimento neuronale periferico, controllando il numero di fibre

(*) Dipartimento di Scienze Farmacologiche e Biomolecolari, Università degli Studi di Milano, Via G. Balzaretti 9, 20133 Milano, Italy.
E-mail: valerio.magnaghi@unimi.it

amieliniche, deputate alla trasmissione nocicettiva, quindi partecipando al dolore periferico. Queste osservazioni sono importanti per lo sviluppo della futura farmacoterapia del dolore, che dovrebbe tenere in considerazione il coinvolgimento delle cellule gliali e i meccanismi neuropatologici alla base del dolore cronico.

***

ABSTRACT. – Pain is an adaptive sensation that normally appears as a warning, activated in response to a damage of the organism. Pain serves to protect the organism to further tissue injuries. The International Association for the Study of Pain (IASP) defined pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (see definition on the IASP homepage at www.iasp-pain.org). Pain can be generally divided into two categories, acute and chronic pain: acute pain is properly a sudden warning pain which signals that something is wrong in the body. If the cause is not removed acute pain may develop in chronic pain, which is persistent and debilitating.

When a damage or a dysfunction affects directly the central or the peripheral nervous system, so that the pain becomes the pathology, it is termed neuropathic pain. It can be described as burning, tingling, shooting, stabbing, or shocking. Injury to the brain, brain tumors, diabetic neuropathy as well as herpes zoster are among the pathologies that may cause this type of pain [1]. Neuropathic pain is often underdiagnosed and undertreated, and it is associated with suffering, disability, impaired quality of life, and increased cost. Neuropathic pain is a long-term process that results difficult to treat.

The increasing knowledge of the mechanisms of pain identified the primary afferent neurons as an important target for the development of novel analgesics. However, neuropathic pain is an heterogeneous process, in terms of etiological factors, molecular mechanisms, but also cells involvement [2]. The central (*i.e.*, astrocytes, oligodendrocytes) and peripheral glial cells (*i.e.*, Schwann cells) are fundamental for the regulation of neuronal activity and also for the control of the nociceptive transmission.

γ-aminobutiric acid (GABA) is generally considered the main inhibitory transmitter in the nervous system, being involved in the processing of nociceptive information. These effects are mediated through activation of ionotropic (*i.e.* GABA-A) and metabotropic (*i.e.* GABA-B) receptors. In particular, the importance of GABA-B receptors in nociceptive processing was well documented in the early 90's, in a series of preclinical studies in which the GABA-B specific agonist baclofen

proved to be antinociceptive in models of acute and chronic pain [3,4]. These effects are likely mediated by spinal and supraspinal GABA-B receptors. The supraspinal effects appear to reflect depression of ascending adrenergic and dopaminergic inputs to the brainstem, and facilitation of descending nor-adrenergic input to the spinal cord dorsal horn [5]. At spinal cord level, the baclofen-induced analgesic effect is attributed to the activation of presynaptic GABA-B receptors, localized on the terminals of peptidergic primary afferents fibers, which determine the suppression of transmitter release [6]. GABA-B receptors are also involved in the inhibitory effects of other co-neurotransmitters on glutamatergic synaptic transmission and pain [7]. However, postsynaptic mechanisms are involved. Indeed, the decrease of dorsal horn neuron excitability and the regulation of intrinsic neuronal properties [8,9] suggest an additional postsynaptic GABA-B mechanism on pain. GABA-B receptors exert a basal inhibition that is induced by a tonic excitatory control through the group I metabotropic glutamate receptors (mGluR), thus resulting in a dynamic balance of inhibitory and excitatory inputs on the dorsal horn neurons [9]. Both GABA-B receptors and group I mGluRs modulate inwardly rectifying potassium (Kir) channels, L-type and N-type voltage-gated calcium channels [10,11]. The contribution of GABA-B receptors to the nociceptive processing was further explored with GABA-B1 knock-out mice [12]. Acute mechanical and thermal thresholds to painful stimuli were reduced in knock-out animals. Thus, tonic GABA-B receptor activation appears to contribute to the establishment of a nociceptive threshold. However, the use of baclofen in clinical practice has been limited due to a rapid tolerance and the possible desensitization of GABA-B receptors.

In the mid-70's the studies on the presence of GABA and its receptors outside the central nervous system were performed. Indeed, the accumulation of ^{3}H-GABA was shown in Schwann cells of the taste buds of the amphibian *necturus maculosus* [13]. At the peripheral nervous system (PNS) level, the presence of GABA-B receptors has been then demonstrated in the rat dorsal root ganglion (DRG), in peripheral axons, in autonomic nerve terminals and in pig nodose ganglion cells [14-19]. Recently, we showed that Schwann cells express different active isoforms of GABA-B receptors, such as -1a, -1b, -1c, and -2 [20,21]. Exposure of Schwann cells to the specific GABA-B agonist baclofen, instead, decreases the cell proliferation and the percentage of Schwann-BrdU immunopositive cells. Baclofen

also decreases the levels of the specific P0 and PMP22 peripheral myelin proteins, suggesting a GABA-B mediated control of peripheral myelination [21]. Interestingly, the presence of possible autocrine GABA-B mediated mechanisms in the PNS has been suggested by the capability of Schwann cells to synthesize GABA. It may activate autocrinally either the GABA-B or the GABA-A receptors, regulating the Schwann cell development [22]. Therefore, the PNS is a potential target for GABA's action, suggesting that GABA-A and GABA-B receptors cross-interact in a complex mechanism influencing the Schwann cell biology [23].

Given the thermal hyperalgesia and the lack of allodynic behavior observed in GABA-B1 knock-out mice, this type of animals has been used to assess the peripheral contribution of GABA-B receptors to the nociceptive processing [24]. Indeed, the morphologic and biomolecular analysis of GABA-B1 knock-out mice further evidences the role of this receptor in the PNS. GABA-B1-deficient mice exhibited morphological and molecular changes in peripheral myelin, including an increase in the number of irregular fibers and in the expression levels of PMP22 and P0 [24]. In particular, these mice presented a high number of small myelinated fibers and small neurons of the lumbar DRG [24], suggesting a role for GABA-B in the control of the peripheral nociceptive fibers. Although, GABA-B receptors may have activational effect in PNS, the possibility that the sensory nerve fiber dysmyelination contributes to the nociceptive phenotype is plausible. A peripheral co-regulation between GABA-A and GABA-B receptors in the control of Schwann cell physiology may contribute to add more complexity to these mechanisms [25]. GABA-A receptors, in fact, are present in Schwann cells, they participate in the control of myelin protein expression and they show opposite effects to GABA-B activation [25]. A similar cross-interaction between GABA-A and GABA-B receptors may explain the decrease in antinociceptive potency of baclofen in those mice lacking the GABA-A 3 subunit [26]. In a model of diabetic neuropathy, bearing peripheral demyelination and neuropathic pain, the increased glutamatergic input from primary afferents to dorsal horn neurons is the result of a reduction in presynaptic GABA-B receptors [27]. However, in the PNS of GABA-B1 knock-out mice not only the peripheral myelin but also the neuronal compartment was modified [24]. Electron microscopy analysis of the lumbar DRG, where the soma of the sensory neurons that synapse the lamina I-IV of the spinal dorsal

horn are located, showed an increase in small neurons [24]. The small sensory neurons of the DRG with thin myelinated axons usually origin the Aδ fibers [28]. The altered morphological findings in the PNS of GABA-B1 knock-out mice, such as the high number of myelinated fibers of small caliber might be ascribed to an increase in Aδ and C fibers. Moreover, the neuropathologic analysis of the skin, revealed that the number of intraepidermal unmyelinated sensory fibers, mostly representing C fibers, is normal in GABA B1 knock-out mice. In agreement, the immunolabeling of the peripheral nerves for specific neuropeptides (*e.g.* CGRP) or neurofilaments (*i.e.* NF200) confirmed that likely Aδ-sensory fibers are increased in the GABA-B1 knock-out mice. However, some extrasynaptic roles for GABA-B receptors in nociception have been also demonstrated at the spinal level. Post synaptic effects of GABA-B1b-containing receptors have been demonstrated in somatosensory neurons, whereas GABA-B2/GABA-B1b dimers specifically induce postsynaptic, long-lasting inhibition of dendritic Ca^{2+} spikes [29]. The involvement of calcium channels in GABA-B-mediated control of neuronal activity has been documented in the spinal cord.

In conclusion, chronic and neuropathic pains are still far from the identification of clinic reliable therapies. The involvement of glial cells in the neuropathologic and molecular mechanisms of pain should be taken in consideration to improve the future pharmacotherapy of neuropathic pain. We showed that the GABAergic system (GABA and its GABA-A and GABA.B receptors) is functionally present in Schwann cells. GABA-A receptors, via activation of the PK-A and PK-C signaling pathways, control the GABA synthesis in Schwann cells and supply an autocrine loop that, in turn, regulates peripheral myelination. GABA-B receptors induce either Schwann cell autonomous or non-autonomous mechanisms. GABA-B receptors control myelin structure, myelin protein expression and Schwann cell structure. Moreover, GABA-B receptors induce Schwann cell mediated changes in the neuronal compartment, controlling the number of unmyelinated fibers, thus participating actively to the peripheral nociception. Future pharmacotherapy of neuropathic pain should necessarily take in account these issues.

BIBLIOGRAPHY

[1] TREEDE R.D. *et al.* (2008), *Neuropathic pain: redefinition and a grading system for clinical and research purposes,* Neurology **70**, 1630-5.

[2] YAMANAKA H. AND NOGUCHI K. (2012), *Pathophysiology of neuropathic pain: molecular mechanisms underlying central sensitization in the dorsal horn in neuropathic pain,* Brain Nerve **64,** 1255-65.

[3] MALCANGIO M. AND BOWERY N.G. (1995), *Possible therapeutic application of GABAB receptor agonists and antagonists,* Clin Neuropharmacol **18**, 285-305.

[4] PATEL S. *et al.* (2001), *The effects of GABA(B) agonists and gabapentin on mechanical hyperalgesia in models of neuropathic and inflammatory pain in the rat,* Pain **90**, 217-26.

[5] SAWYNOK J. (1984), *GABAergic mechanisms in antinociception,* Prog Neuropsychopharmacol Biol Psychiatry **8**, 581-6.

[6] PRICE G.W. *et al.* (1984), *Are baclofen-sensitive GABAB receptors present on primary afferent terminals of the spinal cord?*, Nature **307**, 71-4.

[7] SUZUKI T. *et al.* (2005), *Effect of a selective GABA(B) receptor agonist baclofen on the mu-opioid receptor agonist-induced antinociceptive, emetic and rewarding effects,* Neuropharmacology **49**, 1121-31.

[8] KANGRGA I., JIANG M.C. AND RANDIC M. (1991), *Actions of (-)-baclofen on rat dorsal horn neurons,* Brain Res **562**, 265-75.

[9] DERJEAN D. *et al.* (2003), *Dynamic balance of metabotropic inputs causes dorsal horn neurons to switch functional states,* Nat Neurosci **6**, 274-81.

[10] VOISIN D.L. AND NAGY F. (2001), *Sustained L-type calcium currents in dissociated deep dorsal horn neurons of the rat: characteristics and modulation,* Neuroscience **102**, 461-72.

[11] TOMBLER E. *et al.* (2006), *G protein-induced trafficking of voltage-dependent calcium channels,* J Biol Chem **281**, 1827-39.

[12] SCHULER V. *et al.* (2001), *Epilepsy, hyperalgesia, impaired memory, and loss of pre- and postsynaptic GABA(B) responses in mice lacking GABA(B(1))*, Neuron **31**, 47-58.

[13] NAGAI T. *et al.* (1998), *Uptake and release of neurotransmitter candidates, [3H]serotonin, [3H]glutamate, and [3H]gamma-aminobutyric acid, in taste buds of the mudpuppy, Necturus maculosus,* J Comp Neurol **392**, 199-208.

[14] BOWERY N.G. *et al.* (1981), *Bicuculline-insensitive GABA receptors on peripheral autonomic nerve terminals,* Eur J Pharmacol **71**, 53-70.

[15] DESARMENIEN M. *et al.* (1984), *Coexistence of GABAA and GABAB receptors on A delta and C primary afferent,* Br J Pharmacol **81**, 327-33.

[16] LISKE S. AND MORRIS M.E. (1994), *Extrasynaptic effects of GABA (gamma-aminobutyric acid) agonists on myelinated axons of peripheral nerve.* Can J Physiol Pharmacol **72**, 368-74.

[17] SUN B.B. AND CHIU S.Y. (1999), *N-type calcium channels and their regulation by GABAB receptors in axons of neonatal rat optic nerve,* J Neurosci **19**, 5185-94.
[18] TOWERS S. *et al.*(2000), *GABAB receptor protein and mRNA distribution in rat spinal cord and dorsal root gangli,* Eur J Neurosci **12**, 3201-10.
[19] ZAGORODNYUK V.P. *et al.* (2002), *Functional GABAB receptors are present in guinea pig nodose ganglion cell bodies but not in peripheral mechanosensitive endings,* Auton Neurosci **102**, 20-9.
[20] MAGNAGHI V. *et al.* (2006), *GABA receptor-mediated effects in the peripheral nervous system: A cross-interaction with neuroactive steroids,* J Mol Neurosci **28**, 89-102.
[21] MAGNAGHI V. *et al.* (2004), *GABAB receptors in Schwann cells influence proliferation and myelin protein expression,* Eur J Neurosci **19**, 2641-9.
[22] MAGNAGHI V. *et al.* (2010), *GABA synthesis in Schwann cells is induced by the neuroactive steroid allopregnanolone,* J Neurochem **112**, 980-90.
[23] MAGNAGHI V., PROCACCI P. AND TATA A.M. (2009), *Novel pharmacological approaches to schwann cells as neuroprotective agents for peripheral nerve regeneratio,* Int Rev Neurobiol **87**, 295-315.
[24] MAGNAGHI V. *et al.* (2008), *Altered peripheral myelination in mice lacking GABAB receptor,* Mol. Cell. Neurosci. **37**, 599-609.
[25] MAGNAGHI V. (2007), *GABA and neuroactive steroid interactions in glia: new roles for old players?,* Curr Neuropharmacol **5**, 47-64.
[26] UGARTE S.D. *et al.* (2000), *Sensory thresholds and the antinociceptive effects of GABA receptor agonists in mice lacking the beta3 subunit of the GABA(A) receptor,* Neuroscience **95**, 795-806.
[27] WANG X.L. *et al.* (2007), *Altered synaptic input and GABAB receptor function in spinal superficial dorsal horn neurons in rats with diabetic neuropath,* J Physiol **579**, 849-61.
[28] RUSCHEWEYH R. *et al.* (2007), *Modification of classical neurochemical markers in identified primary afferent neurons with Abeta-, Adelta-, and C-fibers after chronic constriction injury in mice,* J Comp Neurol **502**, 325-36.
[29] PEREZ-GARCI E. *et al.* (2006), *The GABAB1b isoform mediates long-lasting inhibition of dendritic Ca2+ spikes in layer 5 somatosensory pyramidal neurons,* Neuron **50**, 603-16.

PLANT BIOLOGY
A RESOURCE FOR PROSPERITY

MARTIN M. KATER (*)

Nota presentata dal s.c. Martino Bolognesi
(Adunanza del 20 dicembre 2012)

SUNTO. – L'agricoltura in tutto il mondo si trova ad affrontare molte sfide. La popolazione mondiale in rapida crescita, i cambiamenti climatici, il bisogno crescente di un'agricoltura sostenibile richiedono soluzioni immediate per garantire una migliore stabilità economica, sociale ed ambientale. Un contributo importante è richiesto alla ricerca sulle piante attraverso lo sviluppo di tecniche sempre più innovative. Tuttavia c'è ancora poco investimento in questo campo di ricerca nonostante la crescente attenzione sia scientifica che politica verso i problemi legati alla agricoltura. E' quindi importante che questo cambi e che si comprenda come investire nella ricerca sulle piante sia un investimento per il futuro della nostra società.

***

ABSTRACT. – Worldwide agriculture is facing many challenges. A fast growing world population, a changing climate and an increasing need for sustainability call for immediate solutions to guarantee environmental, social and economic stability. An important contribution will have to come from innovative plant research. However, despite that there is more and more scientific and political attention for problems related to agriculture, there is still little investment in this field of research. It will be important that this will change and that we start to realize that investing in plant research is an investment in the future of our society.

(*) Department of Biosciences, Università degli Studi di Milano, Milan, Italy.
E-mail: martin.kater@unimi.it

Plants are essential for our existence and form the fundaments for our modern society. Apart from oxygen many products come from plants and our daily life without them will be absolutely impossible. For instance they are the main source of food (cereals, fruits, beans etc.), medicines, clothes, paper, wood, etc. They are also becoming an important resource for bioenergy. It is clear that plants are important for our society but also provide many opportunities for the future.

Unfortunately, the importance of plant research is often not enough recognized, despite the fact that difficulties related to agriculture are more and more becoming evident. Therefore, plant scientists are asking for special attention for these problems since many difficulties are foreseen within the next 50 years. Examples are, the rapid increase of the world population, water shortage, climate change, limitations in natural resources, costs of energy and this all goes together with a deep global economical crisis.

For many years in Europe it was politically wrong to discuss about yield improvement in agriculture, since we throw in our rich Western society something like 30% of our food in the garbage bin. However, in the mean time the world population has grown with unprecedented speed resulting today in almost 1 billion people on Earth that have not enough calorie intake. Remain the contrasting data that show that in our Western society more than one billion people are obese! Nonetheless, the prognoses for the growth of the world population are showing a further rapid increase towards 9 billion people that will be reached between 2040 and 2050, which means feeding about 2 times the population of China extra within in the next 40 years. In principle, based on the agricultural practice that we apply to day, this will be feasible. We only have to use more land, more water, more fertilizers, more pesticides etc. It is clear that this is not a sustainable option and will not bring solutions for the very poor people. Furthermore, not solving this problem is rather risky since history has learnt that hunger has always resulted in social instability.

What is needed is a second green revolution but this time it should be a sustainable agricultural revolution. Is this possible? Plant scientists think it is. Many examples from basic and applied research show that in nature characters can be identified that would be very helpful for agriculture, for example, drought tolerance of plant species that grow in the desert, disease resistance available in often untapped wild varieties. (Molecular) Geneticists and plant breeders are nowadays closely collaborating and using the available biodiversity for introgres-

sion of important characters into local crops for yield improvement and to grow them cheaper and more environmentally sustainable. Examples are the identification of the *SUB1A* [1] and *PSTOL1* [2] genes. Both genes have been lost during the domestication of rice and are only found in wild varieties. SUB1A provides resistance to flooding of rice paddy fields, which often occurs in countries like Bangladesh and which leads to important harvest losses. The *SUB1A* gene prevents that rice plants start to elongate when they are under water and secures that when water levels return to normality the plants produce an almost normal amount of seeds. The *PSTOL1* gene is important for root development and increases drastically the number of roots. Plant breeders for a long time neglected looking at roots and they focused mainly on the areal parts of the plant. The presence of the *PSTOL1* gene, again identified in a non-commercial rice variety, increases the uptake of nutrients such as phosphor, which makes these varieties able to grow on soil that is low in phosphor and no fertilizers are needed. This saves money and makes it possible to obtain improved harvest yields also in countries with poor soil conditions. It is also important to notice that natural resources for phosphor are extremely limited on Earth.

These and many other examples make clear that plant science can provide important solutions for agriculture. Worldwide scientists are doing basic and applied research using model plants like *Arabidopsis thaliana* and high-technology approaches. Over the last 30 years this has led to an enormous amount of information and deep knowledge about plant processes. Although we are at the tip of the iceberg concerning our knowledge of how plant function, still it is possible to use this knowledge for important agricultural innovations. Therefore it is important to realize that basic research lays the fundaments for more applied research and it is worrying that globally funding goes mainly into applied plant research. However, even more worrying is that the total amount of money invested in plant research is dramatically low, which is only about 5% of the total global R&D budget [3]. Bill Gates in his Annual Letter of the Bill & Melinda Gates Foundation in 2012 commented on this as follows: "Given the central role that food plays in human welfare and national stability, it is shocking – not to mention short-sighted and potentially dangerous – how little money is spent on agricultural research" [4].

It is clear that if we want to provide timely solutions for the increasing world population and to make our agriculture globally more

environmentally and economically sustainable we will have to invest in basic and applied plant research and to invest into a young generation of well trained plant researchers. Let's hope that policy makers get this more long-term vision!

Martin Kater is Professor of Genetics. His research group is focusing on the molecular genetic control of flower and seed development.

REFERENCES

[1] Xu et al., Nature 442, 705-708 (2006).
[2] Gamuyao et al. Nature 488, 535-539 (2012).
[3] Food: The growing problem, *Nature* 466, 546-547 (2010).
[4] http://www.gatesfoundation.org/who-we-are/resources-and-media/annual-letters-list/annual-letter-2012.

INDICE DEL VOLUME

BOLOGNESI M., Nuovi approcci teorici e sperimentali alla scoperta di farmaci 3

NASINI G., Biomolecole da basidiomiceti come sorgente di potenziali nuovi farmaci 7

CASARI C.S., Carbon atomic wires: from stars to nanotechnology 17

FANT R., Le mura spagnole di Milano: gli interventi conservativi (2006-2009) e la loro negazione 37

FANT R., Il cantiere della "Ga' Granda". La storia del manufatto nella lettura dell'architetto-conservatore 65

SANSONETTI A., Il cantiere della "Ca' Granda". La lettura dell'esperto scientifico 93

GRASSELLI M., Finite-dimensional global attractor for a nonlocal phase-field system 113

GIORGILLI A., On a theorem of Lyapounov 133

SCANNICCHIO D., Gli acceleratori nucleari, nella terapia dei tumori 161

FUGANTI C., BRENNA M.E., Il profumo di pulito 173

DELLI CARRI A., Caratterizzazione molecolare e funzionale di neuroni striatali derivati da cellule staminali embrionali umane 185

TERENGHI G., Bioengineering the artificial nerve 189

D'ARIANO G.M., Estensione della teoria quantistica di campo a teoria di automa cellulare quantistico 199

VILLAIN J., La sabbia e i fisici 215

PROSPERI G.M., Velocità limite, causalità e test della Relatività Ristretta 223

DE SILVESTRI S., SVELTO O., Come catturare nel tempo gli eventi primari di processi naturali: i laser ad impulsi ultrabrevi 239

GIORGILLI A., On the representation of maps by lie transforms 251

SENECI P., Neurodegenerative diseases and tau: pathological events and molecular targets 279

BRENNA M.E., FUGANTI C., Perché il gusto del caffè è così buono 299

ROSSO A., Avalanche dynamics in disordered systems 311

MAGNAGHI V., MOTTA M., Gaba receptors and pain regulation 337

KATER M.M., Plant Biology, a resource for prosperity 345

INDICE PER MATERIA

Architettura

Restauro

Fant R., Le mura spagnole di Milano: gli interventi conservativi (2006-2009) e la loro negazione 37

Fant R., Il cantiere della "Ga' Granda". La storia del manufatto nella lettura dell'architetto-conservatore 65

Sansonetti A., Il cantiere della "Ca' Granda". La lettura dell'esperto scientifico 93

Matematica

Giorgilli A., On a theorem of Lyapounov 133

Grasselli M., Finite-dimensional global attractor for a nonlocal phase-field system 113

Fisica matematica

Giorgilli A., On the representation of maps by lie transforms 251

Scienze chimiche

Bolognesi M., Nuovi approcci teorici e sperimentali alla scoperta di farmaci 3

Seneci P., Neurodegenerative diseases and tau: pathological events and molecular targets 279

Brenna M.E., Fuganti C., Perché il gusto del caffè è così buono 299

Fuganti C., Brenna M.E., Il profumo di pulito 173

Chimica delle sostanze naturali

NASINI G., Biomolecole da basidiomiceti come sorgente di potenziali nuovi farmaci 7

SCIENZE FISICHE

CASARI C.S., Carbon atomic wires: from stars to nanotechnology 17

DE SILVESTRI S., SVELTO O., Come catturare nel tempo gli eventi primari di processi naturali: i laser ad impulsi ultrabrevi 239

ROSSO A., Avalanche dynamics in disordered systems 311

VILLAIN J., La sabbia e i fisici 215

Adroterapia

SCANNICCHIO D., Gli acceleratori nucleari, nella terapia dei tumori 161

Fisica delle particelle

PROSPERI G.M., Velocità limite, causalità e test della Relatività Ristretta 223

Fisica quantistica

D'ARIANO G.M., Estensione della teoria quantistica di campo a teoria di automa cellulare quantistico 199

SCIENZE MEDICHE

MAGNAGHI V., MOTTA M., Gaba receptors and pain regulation 337

TERENGHI G., Bioengineering the artificial nerve 189

Neuroscienze

DELLI CARRI A., Caratterizzazione molecolare e funzionale di neuroni striatali derivati da cellule staminali embrionali umane 185

SCIENZE NATURALI

KATER M.M., Plant Biology, a resource for prosperity 345

Finito di stampare nel mese di Dicembre 2015

Decreto Pres. Tribunale 10 gennaio 1961, n. 5504 registro cancelleria
Direttore responsabile: Adele Robbiati Bianchi