

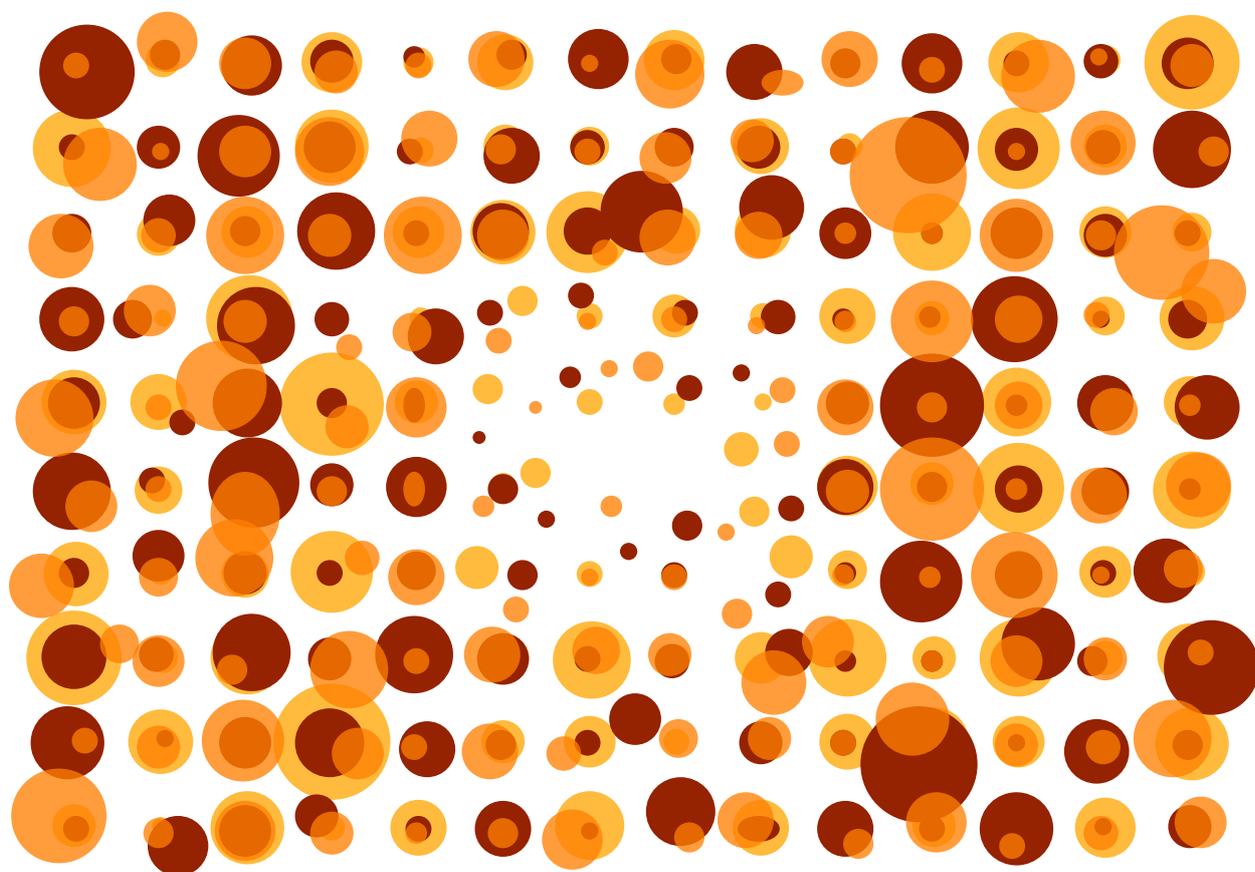
Numero X
Anno 2017


DIPARTIMENTO DI MATEMATICA E FISICA
"Ennio De Giorgi"


UNIVERSITÀ
DEL SALENTO

Viaggio nella Scienza

Ithaca



Il vuoto

Ithaca: Viaggio nella Scienza

Una pubblicazione del Dipartimento di Matematica e Fisica “Ennio De Giorgi” dell’Università del Salento.

Registrazione presso il Tribunale di Lecce n. 6 del 30 Aprile 2013.
e-ISSN: 2282-8079

Direttore Responsabile
Luigi Spedicato.

Ideatore
Giampaolo Co’.

Comitato di Redazione
Adriano Barra,
Rocco Chirivì,
Paolo Ciafaloni,
Maria Luisa De Giorgi,
Vincenzo Flaminio,
Luigi Martina,
Giuseppe Maruccio,
Marco Mazzeo,
Francesco Paparella,
Carlo Sempi.

Segreteria di Redazione
Daniela Dell’Anna.

© 2013-2017 Dipartimento di Matematica e Fisica “*Ennio de Giorgi*”.

© 2017 per i singoli articoli dei rispettivi autori.

Il materiale di questa pubblicazione può essere riprodotto nei limiti stabiliti dalla licenza
“*Creative Commons Attribuzione – Condividi allo stesso modo 3.0 Italia*” (CC BY-SA 3.0 IT).

Per il testo della licenza: <http://creativecommons.org/licenses/by-sa/3.0/it/deed.it>

Ithaca: Viaggio nella Scienza
è disponibile sul sito:

<http://ithaca.unisalento.it/>

Scriveteci all’indirizzo:

ithaca@unisalento.it

3 In questo numero

5 La storia infinita delle pressioni estremamente basse e della loro misura

Anita Calcatelli

23 La tecnologia del vuoto

Gabriele Chiodini

39 L'effetto Casimir, una forza dal vuoto

Piergiorgio Antonini

49 Il vuoto come cristallo birifrangente: l'esperimento PVLAS.

Guido Zavattini

67 La ricerca di nuova fisica nel vuoto quantistico

Marco Incagli

79 All'origine della massa: la QCD e il suo vuoto

Luca Girlanda

- 89** **La più grande cantonata di Einstein**
Paolo Ciafaloni
- 95** **La quantità del nulla**
Lorenzo Dello Schiavo, Anna Baccaglioni-Frank
- 109** **Etica senza verità**
Dario Antiseri
- 113** **Sfere e gruppi topologici**
Francesco Esposito
- La lezione mancata*

In questo numero

Il tema centrale di questo numero di *Ithaca* è il **vuoto**.

Storicamente il concetto di vuoto ha fatto fatica ad affermarsi. Nella visione della scuola Aristotelica il vuoto non poteva esistere. Poiché si supponeva che le velocità dei corpi fossero inversamente proporzionali alla viscosità del mezzo in cui si muovono, se ne derivava che, poichè l'assenza di mezzo implica viscosità nulla, nel vuoto le velocità dei corpi sarebbero infinite, fatto non osservato nemmeno nei corpi celesti. Questo sillogismo, polemicamente presentato contro la visione di Democrito di una natura costituita da atomi fluttuanti nel vuoto, è stata la base culturale su cui tutte le concezioni contrarie al vuoto si sono basate per secoli.

Anche oggi il concetto di vuoto non è ben definito. C'è il vuoto considerato come assenza di materia, e la storia di come, pragmaticamente, questo vuoto sia stato prodotto e reso sempre più spinto è raccontata nell'articolo di Anita Calcatelli. L'articolo di Gabriele Chiodini presenta le moderne tecniche per ottenere il vuoto e le sfide tecnologiche che i grandi apparati sperimentali richiedono per generare, e gestire, il vuoto in volumi sempre più grandi.

D'altra parte questa visione di vuoto concepito come assenza di materia deve essere inglobata nell'attuale descrizione della natura fisica dell'universo riempito dai campi generati dalle quattro interazioni fondamentali, elettromagnetica, nucleare forte, nucleare debole e gravitazione. La meccanica quantistica, ed in particolare il principio di indeterminazione di Heisenberg, impediscono la conoscenza precisa dell'energia del vuoto. Ne risulta che il vuoto quantistico non è assenza, ma è piuttosto lo stato di energia mi-

nima in cui i numeri quantici che lo descrivono hanno, mediamente, valore nullo. Mediamente significa che, nello scorrere del tempo, è possibile che si creino coppie di particelle e antiparticelle i cui numeri quantici si compensano.

Questa descrizione del vuoto non è pura speculazione, ma genera degli effetti fisici misurabili. Storicamente, il primo di questi effetti fu identificato in una modifica dello spettro di eccitazione dell'atomo di idrogeno, effetto noto oggi come *Lamb shift* [1]. Gli articoli di Antonini, Incagli e Zavattini parlano di fenomeni fisici analoghi al *Lamb shift*.

L'articolo di Piergiorgio Antonini descrive gli effetti delle fluttuazioni del vuoto quantistico tra le superfici di due conduttori, un fenomeno che collega il mondo microscopico descritto dalla teoria dei campi quantizzati a quello macroscopico delle misure fatte con apparati usati nell'elettromagnetismo classico.

L'articolo di Guido Zavattini presenta un esperimento che misura gli effetti delle fluttuazioni del vuoto sulla polarizzazione di un raggio LASER come se quest'ultimo passasse attraverso un cristallo birifrangente. Anche in questo caso, fenomeni microscopici hanno conseguenze osservabili su entità macroscopiche.

Da questo punto di vista l'articolo di Marco Incagli è leggermente diverso. In questo caso è evidente che l'accordo tra previsioni della teoria e valore sperimentale del momento magnetico del muone è ottenuto soltanto perchè la teoria considera le correzioni legate alle fluttuazioni del vuoto. D'altra parte, aumentare ulteriormente l'accuratezza della misura potrebbe mettere in luce effetti che l'attuale quadro teorico, detto *Modello Standard*, non prevede.

La struttura del vuoto quantistico ha delle implicazioni teorico-speculative notevoli. Normalmente questi fenomeni sono trattati, e discussi, nell'ambito delle interazioni elettrodeboli, e a queste interazioni sono legati i fenomeni a cui gli articoli di Antonini, Zavattini ed Incagli fanno riferimento. I due articoli successivi si riferiscono alle altre due interazioni fondamentali.

L'articolo di Luca Girlanda mette in evidenza come nel caso delle interazioni forti le proprietà del vuoto siano tali da formare un condensato pronto a generare particelle, meglio dire stati legati, così come un liquido a temperature inferiori a quella di fusione è pronto a cristallizzare.

Paolo Ciafaloni discute dei problemi che nascono nel tentativo di conciliare le proprietà del vuoto quantistico con la teoria della gravitazione universale, e le difficoltà interpretative, ed eventuali incongruenze, che ne conseguono in ambito cosmologico.

Infine, facendo un ulteriore passo verso l'astrazione, Lorenzo Dello Schiavo e Anna Battaglini-Frank discutono della relazione tra concetti di vuoto e di zero nella matematica, da quella babilonese a quella moderna.

Come si vede, si tratta di un vuoto molto pieno di sorprese.

Il numero si chiude con due contributi scollegati dal tema principale: un breve articolo di Dario Antiseri riguardante il rapporto tra Etica e Scienza ed una *lezione mancata* di Francesco Esposito su sfere e gruppi topologici.

Auguriamo una buona lettura,
il Comitato di Redazione.



[1] W. E. Lamb, R. C. Retherford: *Fine Structure of the Hydrogen Atom by a Microwave Method*, Phys. Rev. , **72** (1957) 241.



La storia infinita delle pressioni estremamente basse e della loro misura

When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science. -

William Thomson, Lord Kelvin

Anita Calcatelli

*Già ricercatrice dell'Istituti di Metrologia G. Colonnetti del CNR (ora INRIM)
Strada delle Cacce 73, 10135 Torino, Italia*

Lo scopo della presente nota è quello di dare un panorama storico dello sviluppo della fisica e tecnologia del vuoto e delle inter-conessioni tra la sua misura ed i metodi per ottenere pressioni estremamente basse, quindi prospettare brevemente la situazione attuale.

Introduzione

Prima di iniziare questo breve escursus storico è opportuno stabilire il parametro in base al quale faremo le nostre considerazioni: la pressione limite o ultima raggiungibile in un sistema da vuoto, inteso come un insieme di camera nella quale si vuole instaurare la condizione di vuoto, sistema di pompaggio, misuratore e tutta la componentistica che noi oggi siamo abituati ad usare o che è indispensabile per i processi che si

vogliono realizzare o studiare. La pressione ultima è la più bassa pressione prodotta e misurata in un sistema da vuoto a temperatura ambiente ed è raggiunta quando la variazione di pressione nel tempo è nulla e cioè: $dp/dt = 0$, ossia $(F + Q)/S = p_{lim}$ pressione limite, F = fughe, Q = degassamento totale ed S_p = velocità di pompaggio¹.

Per ridurre p_{lim} si deve aumentare S o rendere minimi L , Q . Il valore di p_{lim} può essere condizionato per esempio dalla diminuzione della velocità di pompaggio al diminuire della pressione o dalla incapacità dei vacuometri di misurare pressioni basse, o dalle interazioni tra le

¹Essendo l'equazione per un sistema da vuoto di volume V : $-V dp/dt = S_p - F - Q$. Q , che è degassamento totale, può avere varie componenti (degassamento da parti riscaldate, diffusione di gas attraverso le pareti coinvolte e quindi emissione nell'ambiente, degassamento stimolato da elettroni, ioni e fotoni).

superficie coinvolte e l'ambiente in vuoto e tra il misuratore stesso e l'ambiente in vuoto.

La creazione artificiale del vuoto è un vecchio problema, ma solo nella seconda metà del ventesimo secolo sono stati compiuti grandissimi progressi nell'ottenimento e nella misurazione di vuoti sempre più spinti. Oggi si raggiungono pressioni anche inferiori a 10^{-10} Pa in particolari ambienti². La forte spinta migliorativa è derivata non solo da esigenze puramente scientifiche ma anche da precise richieste di alcuni settori della tecnologia (come quelli relativi alla produzione di film sottili per l'elettronica, televisori, monitor, camere di simulazione spaziale, conservazione degli alimenti, acceleratori di particelle, applicazione dei plasmi, ecc.). La tecnica di produzione del vuoto si applica quindi ad un grande numero di impianti con scopi e necessità di livelli di vuoto molto diversi. Lo sviluppo nel tempo della tecnologia del vuoto e dei concetti fisici che la sottendono è stato limitato da vari fattori. Nei primi cento anni dopo l'esperienza di Torricelli il maggior problema era rappresentato dalle fughe e gran parte degli sforzi erano rivolti a rendere migliori (e più facili da usare) le pompe piuttosto che ridurre il livello di vuoto. Ciò anche per mancanza di applicazioni. Come scrisse il Prof. Scandone nel lontano 1968 [1] "senza la scienza e la tecnica del vuoto non si sarebbe sviluppato il più dinamico campo della fisica del XX secolo: l'elettronica e le sue numerose applicazioni". In un dato periodo storico, dunque, il vuoto ultimo è principalmente determinato dallo stato della tecnologia esistente in quel periodo; la spinta a migliorare il vuoto limite è la conseguenza delle necessità della ricerca scientifica e delle applicazioni industriali. Fin dall'origine i concetti di vuoto e di pressione atmosferica erano appaiati e gli esperimenti volti a spiegare l'una mostravano anche l'esistenza dell'altro.

²La pressione è una grandezza derivata del Sistema Internazionale di unità di misura definita come forza che agisce sull'unità di superficie e pertanto è misurata in newton al metro quadrato e cioè in pascal (simbolo: Pa).

Gli anni dell'acquisizione dei concetti di base di vuoto e la demolizione alquanto difficile dei preconcetti (1643 - 1669)

Per chi vive nel XX secolo la pressione atmosferica è una nozione elementare, ammessa senza difficoltà. È ugualmente noto il concetto di vuoto, assenza di aria atmosferica, soprattutto nella nostra era spaziale. Ma non è sempre stato così e si è dovuto attendere il secolo XVII perché queste due nozioni complementari fossero messe in evidenza. Prima di quel secolo, e fin dall'antichità, il vuoto era una nozione teorica, oggetto di controversie tra filosofi-scienziati. Presso i greci, come è noto, si ebbero due scuole. Una, rappresentata da Democrito d'Abdera (~ 460-360 a.C.), concepiva l'esistenza del vuoto in un modo molto prossimo a quello dei fisici dei giorni nostri. L'altra, rappresentata da Aristotele (384-322 a.C.), in contrasto con le teorie democritee negava l'esistenza del vuoto in base a considerazioni di meccanica. La sua teoria unitamente alla visione globale dell'universo ebbe un grande successo per più di venti secoli. Vanno ricordati altresì i lavori della scuola Alessandrina sulla meccanica dei fluidi, ad esempio i due libri di "pneumatica" di Erone d'Alessandria (I o II sec d.C.), che fanno supporre che già a quell'epoca fossero state concepite se non addirittura costruite le prime pompe da vuoto. Nei secoli seguenti l'opera di Aristotele divenne una specie di complemento della Bibbia e ci si "dimenticò" di Democrito e degli Alessandrini. Alcuni pensatori intuirono la nozione di pressione atmosferica ma gli scolastici, nella loro interpretazione della "Fisica" di Aristotele, imposero lo slogan "natura abhorret a vacuo".

Il problema del vuoto tornò in primo piano con Galileo e la sua scuola [2, 3, 4, 5, 6]. Qui considereremo dapprima tre personaggi di indiscussa rilevanza per le acquisizioni metodologiche e concettuali: Evangelista Torricelli (1608-1647), Blaise Pascal (1623-1662) e Otto von Guericke (1602-1686) [7], anche se in questo percorso si darà spazio ad altre conquiste, non solo in quel secolo, meraviglioso sotto molti aspetti e precipuamente in quello scientifico, che fu il 1600.

Torricelli, Pascal, von Guericke sono tre no-

mi tra i molti importanti per quanto concerne la comprensione del fenomeno "atmosfera"; ad essi si associano esperimenti che rappresentano pietre miliari per i successivi sviluppi:

- la misura della pressione atmosferica (esperimenti di Torricelli);
- la variazione della pressione atmosferica con l'altitudine (esperimenti di Pascal-Périer sul Puy de Dome);
- l'applicazione di una pompa da vuoto per realizzare condizioni di rarefazione e dimostrare la "forza dell'aria" (vari esperimenti di Otto von Guericke tra cui quello dei famosi emisferi di Magdeburgo)³

Tutti e tre questi esperimenti appartengono al diciassettesimo secolo in cui le due nozioni complementari "pressione atmosferica" e "vuoto" divennero fatti scientifici supportati dall'esperienza e si posero le basi per risolvere la controversia sull'esistenza del vuoto [8]. Infatti gli anni dal 1640 al 1650 furono fondamentali, in questo come in molti altri campi, per la dimostrazione dell'esistenza della pressione atmosferica e della possibilità di produrre condizioni di vuoto.

L'esperimento di Torricelli

L'esperienza di Torricelli (1608-1647) viene in genere localizzata nel 1643 ed è considerata come il punto di partenza per la definizione del vuoto residuo. È ancora incerto chi fu il costruttore del primo barometro; probabilmente il primo esperimento fu quello condotto da Gasparo Berti (1600-1643) [8] che costruì ed utilizzò un barometro ad acqua (Fig. 1). Sicuramente il Viviani ed il Torricelli dimostrarono sperimentalmente l'esistenza della pressione atmosferica e del vuoto.

Le premesse erano già state poste da tempo e tutti gli ingegneri che si occupavano del trasporto dell'acqua (per esempio sulle colline di Genova) o della sua eliminazione dalle miniere avevano a che fare con l'altezza massima della colonna d'acqua in tubature verticali.

³Otto von Guericke non fu solo scienziato multiforme ma anche politico.

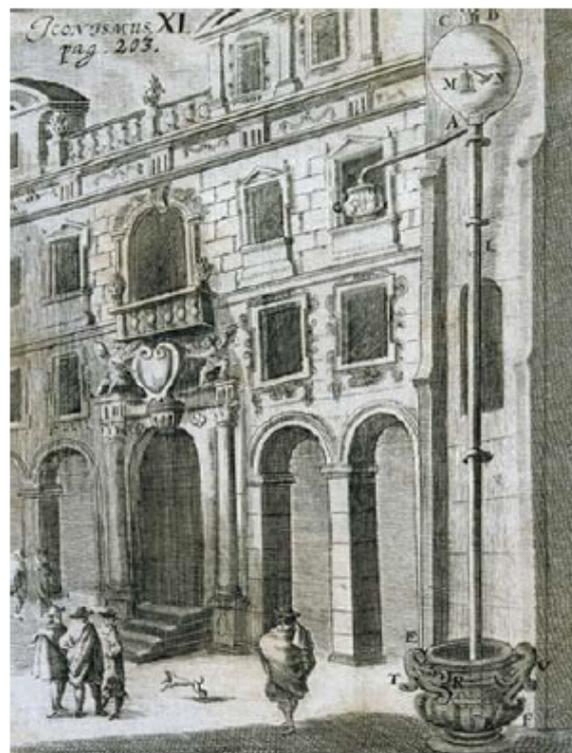


Figura 1: *Esperimento di Gaspare Berti presso il Convento dei Minimi sul Pincio a Roma, descritto da Gaspar Schott, "Technica curiosa, sive Mirabilia artis", Würzburg 1664.*

Molti scienziati ne avevano tentato la spiegazione preparando così il terreno per l'esperimento di Evangelista Torricelli che succedette a Galileo, quand'egli morì l'8 gennaio 1642, come professore di matematica al servizio del Granduca di Toscana [2, 9, 10]. Probabilmente Evangelista Torricelli non fu l'inventore del barometro, ma aveva idee chiare sul concetto di pressione dell'aria.

Egli progettò un esperimento nel 1644 (Fig. 2)⁴, probabilmente eseguito in precedenza da Viviani, in grado di evidenziare variazioni della pressione dell'aria e di produrre il vuoto.

Questo esperimento molto noto dimostrava chiaramente che lo spazio al di sopra del mercurio, quando la colonna veniva capovolta al-

⁴Nella lettera, tra l'altro, si legge: ... Noi abbiamo fatti molti vasi di vetro come i seguenti A e B grossi, e di collo lungo due braccia, questi pieni di argento vivo poi serratagli con un dito la bocca e rivoltati in un vaso dove era l'argento vivo C si vedevano votarsi e non succeder niente nel vaso che si votava; il collo però AD restava sempre pieno all'altezza di un braccio, e un quarto, e un dito in più..



Figura 2: Lettera di Torricelli al suo amico M. Ricci (11 Giugno 1644).



Figura 3: Esperimento di tipo torricelliano eseguito a Roma nel 1645 (da E.Maignan "Cursus philosophicus", Toulouse, 1655 e "Breve storia della Meteorologia a Firenze dalle origini ad oggi", Edito da G. Maracchi, Firenze, 1991, pp. 12).

l'interno di un recipiente contenente anch'esso mercurio, era vuoto.

Il livello del mercurio era indipendente dal volume sopra di esso (Fig. 3). Nonostante l'evidenza molti "aristotelici" non volevano credere

che vi fosse il vuoto⁵.

L'esperimento di Blaise Pascal

Mentre alcuni negavano semplicemente che vi potesse essere il vuoto all'interno del tubo barometrico altri avanzavano l'ipotesi che, nello spazio apparentemente vuoto al di sopra del mercurio, vi fosse una sostanza molto tenue (matière subtile), o "spirito", proveniente dall'interno del mercurio che attraeva il liquido stesso tenendolo sollevato nel tubo. Blaise Pascal (1627-1686) compì nel 1648 un esperimento definito "uno dei più grandi momenti nella storia delle idee" [5, 10]. Nell'esperimento progettato da Pascal, e messo in pratica da suo cognato F. Périer, un barometro e 4 kg di mercurio furono portati sul Puy de Dôme (vulcano estinto situato nel Massif Central in Francia, avente un'altitudine sul livello del mare di 1645 m). Quando il barometro fu sistemato sulla cima della montagna, la colonna di mercurio risultò essere 7,5 cm più corta dell'altezza di una colonna simile posta alla base e tenuta continuamente sotto osservazione. Questo esperimento, ripetuto cinque volte sulla cima e alla base della montagna, costituì una pietra miliare per stabilire che la pressione dell'aria è una realtà ed inoltre che essa varia con l'altitudine con la conseguenza di dimostrare che esiste un "vuoto naturale".

Il 15 novembre 1647 B. Pascal scrisse al cognato F. Pèrrier

Si tratta di fare l'ordinaria esperienza del vuoto diverse volte nello stesso giorno, in un medesimo tubo, con lo stesso argento vivo, sia in basso che sulla cima di una montagna ... per provare se l'altezza dell'argento vivo sospeso nel tubo si troverà uguale o differente ... se avviene che l'altezza dell'argento vivo sia minore in alto che in basso della montagna ne seguirà necessariamente che la pesantezza e la pressione dell'aria è la sola causa di questa sospensione dell'argento vivo e non l'orrore del vuoto, poiché è ben certo che c'è molta più aria che pesa sul piede della montagna che non sulla sommità ...

E il 22 settembre 1648 di F. Pèrrier scrisse a B. Pascal

... noi fummo dunque quel giorno tutti insieme sulle

⁵Per una sintesi del pensiero aristotelico sul concetto di tempo e di spazio vedasi il testo di C. Rovelli, "L'ordine del tempo" ed. Adelphi, 2017 pag. 64.

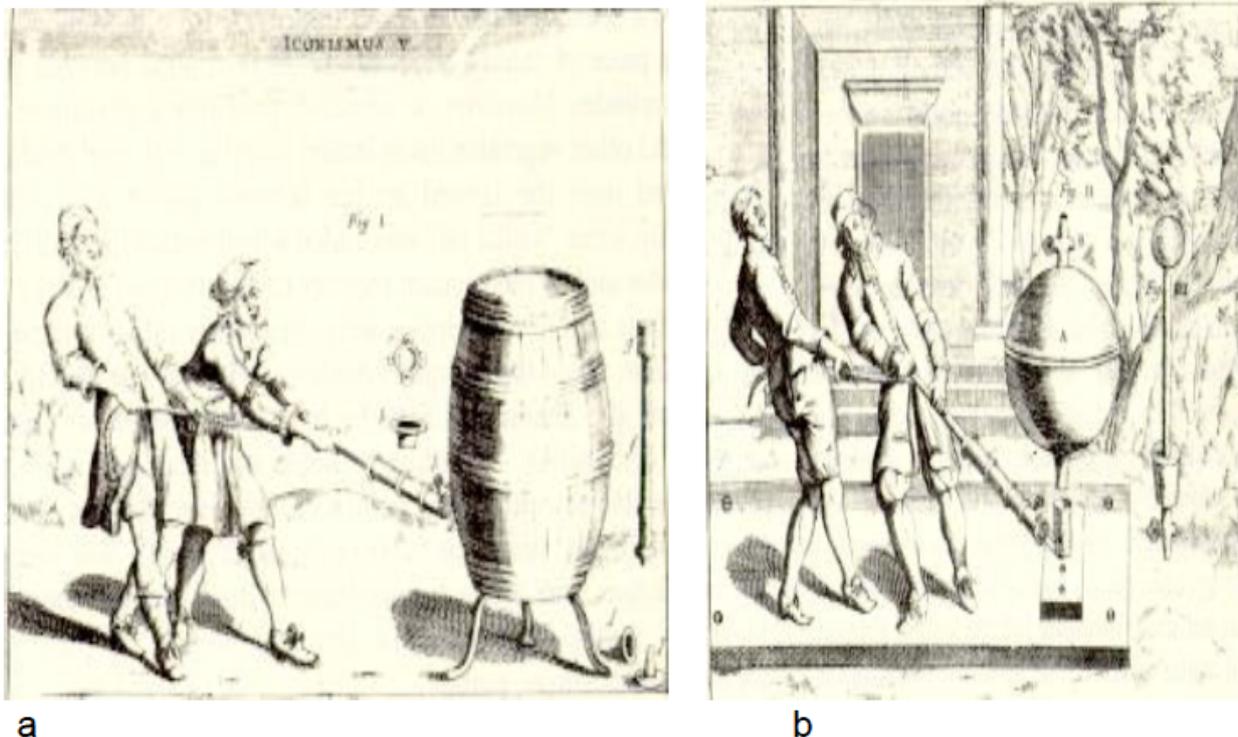


Figura 4: *Esperimenti progettati da Otto von Guericke ed eseguiti nel 1640 (da "History of Vacuum Science and Technology", Edito da T.E. Madey, W.C. Brown, American Vacuum Society, 1984, pp 10): a il vuoto nella botte di legno; b il vuoto in un contenitore sferico di rame.*

otto del mattino nel giardino dei Padri Minimi che è quasi il luogo più basso della città, ove fu iniziata l'esperienza. ... lasciai uno di questi tubi nel suo catino in esperienza continua e segnai sul vetro il livello dell'argento vivo. ... andai con tutti quei signori in alto del Puy de Dome elevato sopra i Minimi circa 500 tese⁶ ... si trovò che in tale tubo l'altezza dell'argento vivo era di 23 pollici 3 linee e mezza ... quindi tra le altezze dell'argento vivo vi era una differenza di 3 pollici, 1 linea e mezza. ... feci ancora altre cinque volte assai esattamente ... e si trovò sempre in tutte queste esperienze la stessa altezza dell'argento vivo ...

Questo è un ottimo esempio di metodo sperimentale. Pascal nei suoi molteplici esperimenti mise in evidenza tra l'altro la dipendenza del livello del liquido dalla densità, già per altro intuita dal Torricelli; infatti, utilizzando mercurio, olio ed acqua, stabilì che il rapporto delle altezze è inversamente proporzionale al rapporto delle densità.

⁶La tesa (toise) unità di misura di lunghezza utilizzata in Francia prima del 1799 ed è pari a circa due metri, 1.949 m

Otto von Guericke: la pompa trasportabile e gli esperimenti di Magdeburgo

L'invenzione della pompa ad aria ad opera di Otto von Guericke fu definita dal W.E.K. Middleton [2, 5] una delle quattro più grandi invenzioni tecniche del secolo (insieme con il telescopio, il microscopio ed il pendolo). Egli condusse esperimenti sul vuoto molto spettacolari anche nelle piazze e li descrisse tutti, unitamente alle sue idee sul cosmo, nell'opera "Experimenta Nova (ut vocantur) Magdeburgica de Vacuo Spatio". Intorno al 1640 egli collegò una pompa ad acqua ad una botte di legno riempita d'acqua per vedere se, dopo aver pompato via l'acqua, rimaneva uno spazio vuoto. Quando, per l'azione della pompa (Fig. 4 a), l'acqua fu completamente estratta si sentì il rumore dell'aria che penetrava attraverso i pori del legno.

In un altro esperimento famoso lo scienziato attaccò la pompa ad un recipiente di rame (Fig. 4 b), questa volta non riempito d'acqua. Quando l'aria fu pompata via la sfera implose e Otto

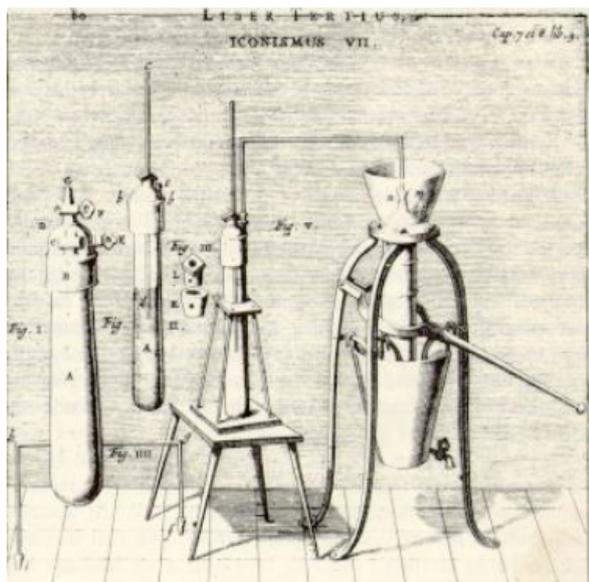


Figura 5: Pompa trasportabile di Otto von Guericke usata per gli esperimenti di Magdeburgo (da "History of Vacuum Science and Technology", Edito da T.E. Madey, W.C. Brown, American Vacuum Society, 1984, pp 18)

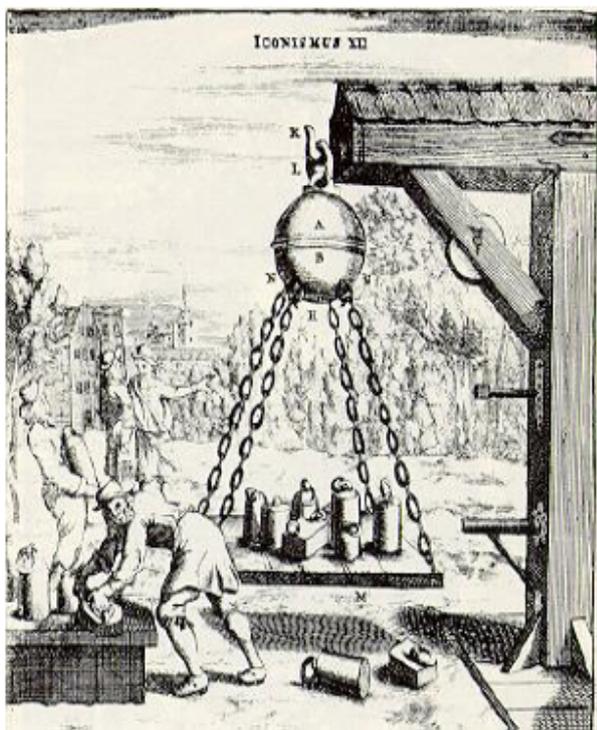


Figura 6: Esperimento progettato da Otto von Guericke per determinare la forza necessaria per separare gli emisferi, ovvero "il peso dell'aria" (da "Adventures in vacuums", Edito da M. J. Sparnaay, Elsevier Science Publishers B.V., 1992, pp. 6)

von Guericke ne attribuì la causa alla pressione

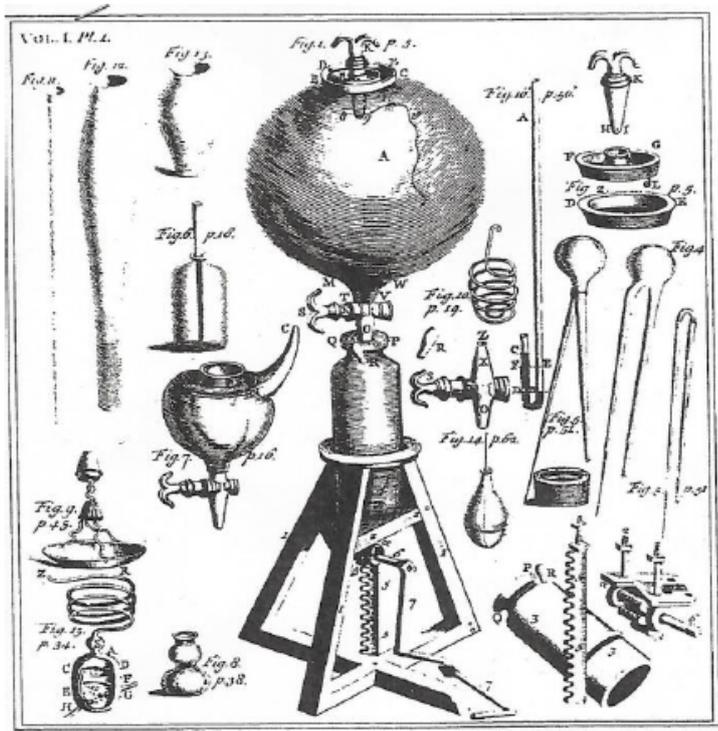
atmosferica e a una non perfetta sfericità del recipiente. Il problema fu risolto costruendo un recipiente "perfettamente" sferico. Egli apportò varie modifiche e miglioramenti alla sua pompa (Fig. 5) e, con i suoi esperimenti, riuscì a formulare delle ipotesi sull'origine di alcuni fenomeni meteorologici, ad esempio che vento e tempeste potessero essere causati da differenze nella pressione atmosferica [2, 10].

Nella Fig. 6 è rappresentato un altro esperimento di Otto von Guericke, cioè il tentativo di separare i due emisferi applicando dei pesi per ricavare anche un valore approssimato delle forze in gioco. Otto von Guericke, informato degli esperimenti di Torricelli, ne condusse per proprio conto osservando il formarsi delle bolle di mercurio che erano ben visibili all'interfaccia tra il vetro ed il mercurio. Egli osservò che bolle di aria fuoriuscivano dal mercurio il cui livello diveniva inferiore di quello iniziale e ciò lo portò ad avanzare l'ipotesi dell'esistenza della "molla dell'aria".

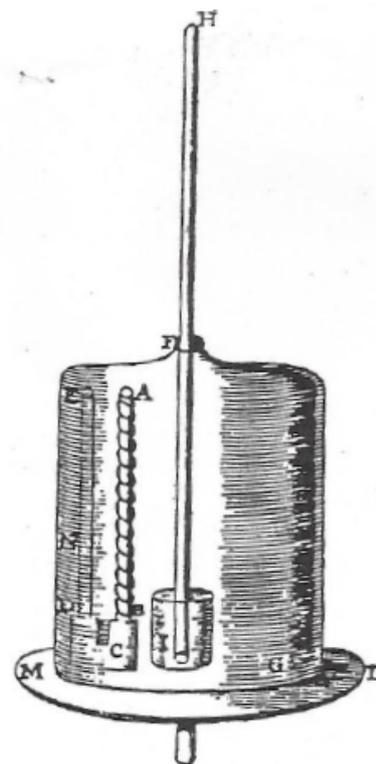
Dal 1660 al circa il 1850

Boyle apportò molte modifiche alla pompa a pistone di von Guericke che utilizzò in molti esperimenti (Fig. 7) che lo portarono a stabilire importanti leggi dei gas tra cui la famosa relazione che lega pressione e volume a temperatura costante dando così l'avvio all'impianto della teoria cinetica dei gas. Infine Robert Boyle misurò il vuoto (≈ 85 Pa) con un manometro a mercurio collegato ad una campana svuotata con una pompa a pistone. 85 Pa era, dunque, la pressione ultima raggiungibile (o misurabile) a quell'epoca. Egli, probabilmente, usò per primo il termine barometro [2, 5].

Gli esperimenti sul vuoto venivano interessando sempre di più il pubblico e quindi venivano eseguiti sulle piazze (emisferi di Magdeburgo) o in riunioni non solo di scienziati dopo cena (le famose after dinner demonstrations); per esempio, nel 1665, Sam Pepys (memorialista) nel suo diario dice di aver partecipato ad una seduta della Royal Society... "there did see a kitlin killed almost quite (but that we could not quite kill her) with sucking away the ayre out of a receiver where she was put- and then the ayre being let in upon her, revives her immediately ... Thence



a



b

Figura 7: a- Pompa a pistone costruita da R.Hook usata anche da R. Boyle al 1660; b- misuratore di vuoto a colonna utilizzato da R. Boyle nelle prime misurazioni di pressioni sub-barometriche: in una campana di vetro, cui era collegata una pompa del tipo a, era saldato un tubo di vetro capovolto in un contenitore di mercurio posto all'interno della campana stessa.

home, and thence to Whetehall where the house full of Dukes going tomorrow; and thence to St James ... " [7]. Era diventato molto comune eseguire questi esperimenti con gli animaletti, tanto che si hanno quadri del seicento in cui vengono messi in vuoto rane o piccoli pappagalli per poi ri-immettere l'aria ambiente appena in tempo per rianimarli (vedasi il quadro Joseph Wright of Derby "Experiment on a bird on the air pump"). In realtà per duecento anni ci si occupò principalmente delle proprietà meccaniche del vuoto e quindi una pressione ultima di qualche centinaio di pascal era sufficiente; sforzi erano piuttosto rivolti a rendere più maneggevoli le pompe piuttosto che ad abbassare il livello di vuoto.

La prima pompa da vuoto a due pistoni fu realizzata da 1704 da Francis Hawksbee(1660-1739 (Fig. 8) [11] ma si dovette aspettare fino al 1850 perché una versione di questa pompa venisse commercializzata; anzi, nel 1850, alla grande esposizione di Londra, fu dato il primo premio

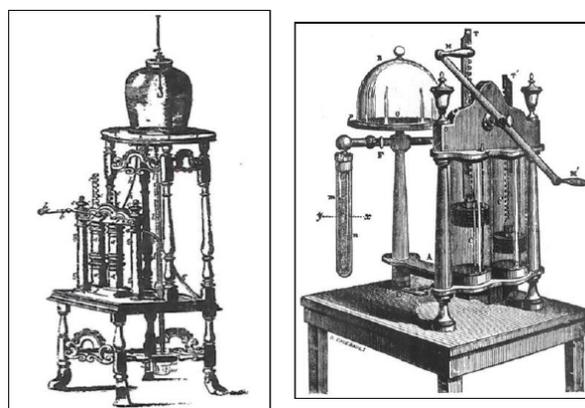


Figura 8: Pompa a due pistoni di Francis Hawksbee (1705). (F. Hawksbee, "Physico-Mechanical Experiments" (London, 1719); b pompa a due pistoni di tipo commerciale [9].

ad una pompa a due pistoni di tipo commerciale che poteva raggiungere un valore di p_{lim} di 13 Pa, solo sei volte più bassa della pressione ultima raggiunta da Boyle.

Altri sperimentatori, tra cui il già citato Robert Hook (1635 - 1702) e Denis Papin (1647 - 1714),

proposero e perfezionarono nuovi tipi di pompe da vuoto. Dotate di uno o due cilindri azionati da un manubrio, una staffa o un sistema di ruote dentate e cremagliera; esse divennero fra gli apparecchi più importanti dei gabinetti di fisica. Contemporaneamente si moltiplicarono le esperienze e le dimostrazioni "pneumatiche" come quella che doveva dimostrare che i suoni non possono propagarsi nel vuoto. Le prime osservazioni sulle scariche elettriche nei gas rarefatti furono rese possibili proprio dalla disponibilità di queste pompe. In quasi duecento anni il vuoto ultimo non si era ridotto di molto per due ragioni:

- non vi erano state richieste di pressioni più basse da parte degli sperimentatori
- il sistema di misura, manometro a mercurio, non consentiva di misurare pressioni inferiori a una sessantina di Pascal.

Intanto incominciavano a comparire i primi sistemi per la misura della pressione arteriosa. Infatti nel 1733 venne pubblicata una descrizione di una misurazione della pressione in una cavalletta in cui un tubo veniva inserito direttamente nella carotide e si misurava l'altezza raggiunta dal sangue de povero animale [12].

Un periodo molto fecondo: dal 1850 al 1900

Tra il 1850 ed il 1900 sotto la spinta di importanti applicazioni emergenti (lampade elettriche, tubi elettronici, ...) si ebbero molti miglioramenti nella tecnologia del vuoto dovuti a parecchi sperimentatori. Nel 1855 Julius Plucker (1801-1868) dell'università di Bonn fece costruire da H. Geissler (1814-1879)⁷ una pompa da vuoto in vetro che usava il mercurio come pistone per fare esperimenti di scarica nei gas a bassa pressione [13]. Questo fu il punto di partenza delle ricerche di metodi e mezzi per raggiungere vuoti sempre più spinti. Infatti nel 1862 A. Toppler apportò modifiche alla pompa di Geissler e nel 1865 H. T. P. Sprengel⁸ (1834-1906) costruì una pompa in cui un treno di gocce di mercurio intrappolava il gas in un tubo di vetro e lo trascinava via. Pro-

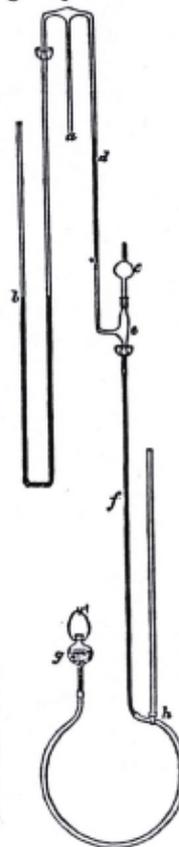
⁷Fu un famoso fisico tedesco e bravissimo soffiatore del vetro.

⁸Sprengel descrisse le sue ricerche in "Researches on the Vacuum" alla Chemical Society in 1865.

XVI. *Apparatus for Measurement of Low Pressures of Gas.* By Professor M'LEOD, Indian Civil-Engineering College, Cooper's Hill*.

THIS apparatus was devised for estimating the pressure of a gas when its tension is so low that the indications of the barometer cannot safely be relied on, unless indeed a very wide barometer and an accurate cathetometer be employed. The method consists in condensing a known volume of the gas into a smaller space and measuring its tension under the new conditions.

The form of the apparatus is the following:—The tube *a* communicates with the Sprengel, and with the apparatus to be exhausted; *b* is a siphon-barometer with a tube about 5 millimetres in diameter; and the principal parts of the measuring-apparatus consist of *c*, a globe of about 48 cubic centims. capacity with the volume-tube at the top, and *d* the pressure-tube; these two are exactly of the same diameter, to avoid error from capillarity. The tube at the bottom of the globe is ground into a funnel-shaped portion at the top of the wide tube *e*; and to the side of the latter the pressure-tube *d* is joined. The volume-tube at the top of the globe is graduated in millimetres from above downwards, the lowest division in this particular apparatus being 45; the pressure-tube *d* is also graduated in millimetres, the 0 being placed at the level of the 45th division on the volume-tube. A ball-and-socket joint connects the bottom of *e* with a vertical tube *f* about 800 millims. long, which is connected at its lower extremity by means of a flexible tube with the mercury-reservoir *g*; a stopcock at *h* permits the regulation of the flow of mercury into the apparatus; this may be conveniently turned by a rod, so that the operator may watch the rise of



* Read before the Physical Society, June 13, 1874. Communicated by the Society.

Figura 9: Descrizione fatta nel 1874 da McLeod del suo misuratore di vuoto a compressione.

prio una pompa di Sprengel fu utilizzata (1870) da W. Crookes [14] per raggiungere un vuoto da lui definito come "approaching perfection". Naturalmente egli agì anche sui materiali e infatti sostituì tubi di gomma con sistemi tutti in vetro.

Fino a questo momento gli strumenti di misura erano essenzialmente basati sulla colonna torricelliana, ma nel 1874 H. McLeod pubblicò [15, 16] la descrizione (Fig. 9) del vacuometro a compressione poi noto semplicemente come McLeod in teoria in grado di misurare vuoti anche dell'ordine di $(10^{-2} - 10^{-3})$ Pa. Questo misuratore è basato sulla compressione del gas residuo in un volume noto e la pressione così incrementata può essere misurata con la colonna. La pressione incognita si ricava applicando la legge di Boyle. Alla fine del 1870 si localizza l'inizio delle grandi applicazioni industriali [17]; Edison costruì le lampade a incandescenza che necessitavano di

condizioni di vuoto (con due pompe di Sprengel e una pompa di Geissler raggiunte vuoti dell'ordine di 10^{-1} Pa, misurato con un McLeod). Incominciarono a presentarsi problemi con i gas condensabili. Nel 1892 H. Fleuss (1851-1932) costruì pompe a pistone solido con tenute a olio e nel 1894 G. W. A. Kahlbaum ottenne pressioni dell'ordine di 10^{-4} Pa. Così in pochi anni intorno al 1900 si migliorarono le pompe a pistone metallico lubrificato a olio (che fa anche tenuta), si costruirono valvole, si incominciarono ad azionare queste pompe con motori elettrici da poco inventati. Agli inizi del 1900 si realizzarono pressioni di circa 10^{-3} - 10^{-4} Pa, sempre misurate con dei McLeod, cioè p_{lim} fu ridotta di sei ordini di grandezza in cinquant'anni. Il miglioramento continuo era in gran parte il risultato dei bisogni dell'industria delle lampade elettriche; d'altra parte alcune fondamentali scoperte scientifiche furono possibili grazie al raggiungimento di vuoti "spinti", per esempio la scoperta, nel 1897, dell'elettrone da parte di J. J. Thomson (1865-1940).

In questo periodo il grande miglioramento nella pressione ultima raggiungibile è dovuto a:

1. riduzione delle fughe mediante l'impiego di sistemi completamente in vetro;
2. riduzione del degassaggio mediante trattamento termico dei sistemi in vuoto;
3. miglioramento delle caratteristiche delle pompe.

Intorno alla metà del 1800 si situano studi e realizzazioni di misuratori di tipo meccanico come i manometri di tipo Bourdon⁹ [18] e, all'inizio del ventesimo secolo, i misuratori meccanici a membrana che, con lo sviluppo dell'elettronica sia per i sensori sia per le unità di controllo, divennero i ben noti CDG (Capacitance Diaphragm Gauges) che sono i più diffusi misuratori di vuoto di elevata precisione [19] in grado di coprire l'intervallo di pressioni dall'atmosfera a circa 10^{-2} Pa con vari tipi di sensori. Questi misuratori, resi disponibili da alcune industrie, sono oggi ampiamente usati come strumenti secondari di

⁹Eugène Bourdon (1808 -1884), ingegnere e orologiaio francese, inventò il misuratore di pressione che da lui prese nome

riferimento [20] per confrontare tra loro impianti primari di istituti metrologici [21].

Dal 1900 a oggi: le tecnologie del vuoto entrano nel commercio

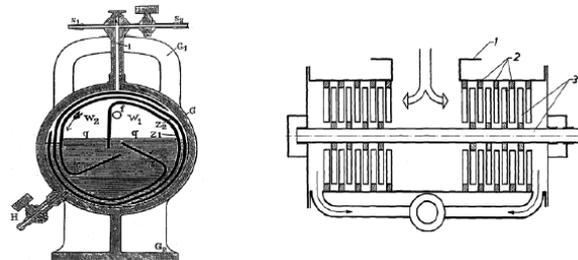


Figura 10: Pompe ideate da Gaede: A sinistra pompa rotativa a mercurio (1905), a destra schema di pompa molecolare (1913) [7].

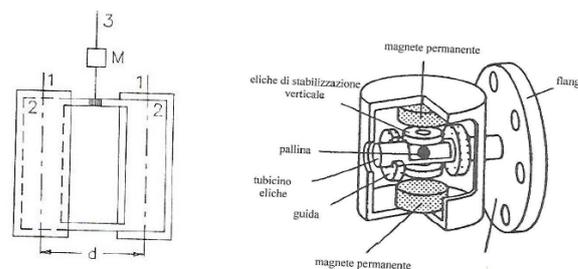


Figura 11: Sinistra: schema di vacuometro radiometrico di Knudsen (3 filo di sospensione, M specchietto, 1 due piatti molto sottili collegati tra loro e sospesi con il filo, 2 piatti simili a 1 ma riscaldabili). Destra: schema di spinning rotor attualmente in uso [25, 26].

Nei primi venti anni del ventesimo secolo il grande avanzamento nella tecnologia del vuoto è dovuto a due personaggi come W. Gaede (1878-1945) [22] in Germania e I. Langmuir in USA (1881-1957) [2, 3].

Infatti, nel 1905 Gaede inventò e brevettò la pompa rotativa a mercurio, che poteva produrre vuoti dell'ordine di 10^{-3} Pa (per l'industria delle lampadine elettriche e dei tubi elettronici) e nel 1914 inventò la pompa molecolare con la quale si potevano raggiungere pressioni nel campo di 10^{-5} Pa. Egli così scriveva: "Il gas viene trascinato fuori dal recipiente verso il vuoto preliminare mediante un cilindro rotante ad alta velocità all'interno di un involucro ermeticamente sigillato" (Fig. 10).

Il vuoto misurato da S. Dushman (1883-1954)¹⁰ era di 5×10^{-5} Pa. [23].

L'inizio del ventesimo secolo portò molte invenzioni anche nel campo delle misurazioni di vuoto perché il fisico danese M.H.C. Knudsen (1871-1949), oltre a dare tanti contributi alla teoria cinetica dei gas [24], inventò il misuratore detto appunto di Knudsen¹¹ (cosiddetto misuratore a frizione) [25].

Esso è basato sul trasferimento di quantità di moto dalle particelle del gas ad una superficie sospesa all'interno della camera in cui si vuol misurare il valore della pressione: una piastra sottile è sospesa tra due elementi uno dei quali è riscaldato mentre l'altro è mantenuto alla temperatura dell'involucro in cui è inserito il vacuometro (Fig. 11 a). Le molecole del gas vicine all'elemento riscaldante colpiscono il lato della piastrina con energia maggiore delle altre e quindi l'elemento sensibile è in rotazione [26]. Dalla frequenza di rotazione si ricava il valore della pressione¹²

Questo misuratore è l'antenato di tutta una serie di misuratori con elementi in sospensione e degli ormai famosi "spinning rotor gauges" (Fig. 11 b) che sono stati dapprima studiati da W. Beams e al. [25] e poi da K. Fremerey [27], in questo caso la sospensione e la rotazione di una sferetta metallica sono ottenute con una combinazione di campi magnetici. Sono strumenti molto precisi e per questo studiati negli istituti

metrologici [28] e sono utilizzati come campioni secondari di riferimento e di trasferimento per confrontare tra loro sistemi di taratura di vacuometri [29].

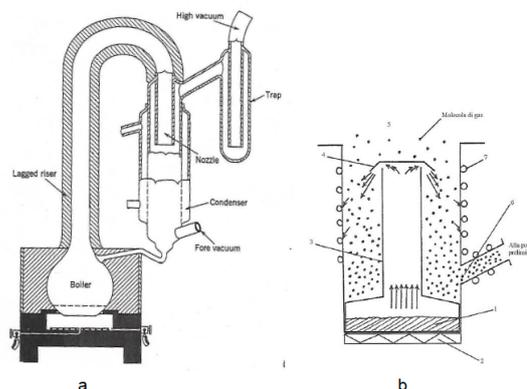


Figura 12: Pompe a diffusione di mercurio di Langmuir: a con ugello cilindrico [9], b con ugello a ombrella (1 fluido motore, 2 fornello riscaldante, 3 camino, 4 ugello attraverso il quale escono a ombrello i vapori risaliti nel camino che trascinano il gas residuo che viene poi pompato da una pompa ausiliaria.

Nel 1906, Marcello Pirani (1880-1968) sviluppò la sua più importante invenzione, un nuovo tipo di vacuometro. Si basa sulla perdita di calore da parte di un filamento per trasferimento al gas circostante e alle pareti del misuratore e, cioè, si osserva il cambiamento di resistività di un filamento di tantalio e platino (oggi si usano tungsteno, platino e nichel) al variare della sua temperatura [30], a sua volta legata alla pressione del gas. Si determina quindi la pressione misurando una grandezza ad essa correlata come la resistenza elettrica. È il ben noto ed ancora oggi molto usato vacuometro a conducibilità termica conosciuto semplicemente come "il Pirani". Pirani lavorò per molti anni presso la Siemens&Halske di Berlino dove si occupava della promozione della produzione di lampadine elettriche. Utilizzava per la misura della pressione nell'impianto di produzione dei McLeod che causavano molti problemi in parte dovuti a contaminazione da mercurio a causa della sua elevata tensione di vapore ed erano soggetti a frequenti rotture essendo in vetro. Con l'introduzione dei vacuometri a conducibilità termica i McLeod venivano utilizzati soltanto per la taratura degli strumenti di lavoro e trattati quindi con maggior cura.

¹⁰È molto noto e tutt'ora usato il suo libro "Scientific Foundations of Vacuum Technique" (1922 e 1949). Questo testo fu completamente revisionato nel 1961 dal suo collega James Lafferty. Le principali ricerche di Dushman riguardano la meccanica quantistica, le forze elettromotrici, la struttura atomica e le emissioni elettroniche.

¹¹Fu ricercatore alla Technical University of Denmark, dove ricevette una medaglia d'oro nel 1895 e divenne professore nel 1912. Nel 1934 pubblicò "The Kinetic Theory of Gases" uno dei testi fondamentali. A lui si deve il numero di Knudsen (un numero adimensionale che rappresenta il rapporto tra il cammino libero medio di una molecola e lo spazio entro cui tale molecola si muove), lo strato di Knudsen (che rappresenta l'interfase tra liquido e vapore), il gas di Knudsen (un modello teorico che ignora le collisioni tra le molecole del gas), il manometro assoluto di Knudsen e la pompa di Knudsen.

¹²Per il modo di funzionare di tutti i vacuometri cui si fecero e che sono ampiamente utilizzati oggi si possono consultare i testi in bibliografia o le dispense di corsi di fisica anche on line come per esempio:
http://www.roma1.infn.it/people/luci/labo_termo/Vuoto_parte2.pdf.

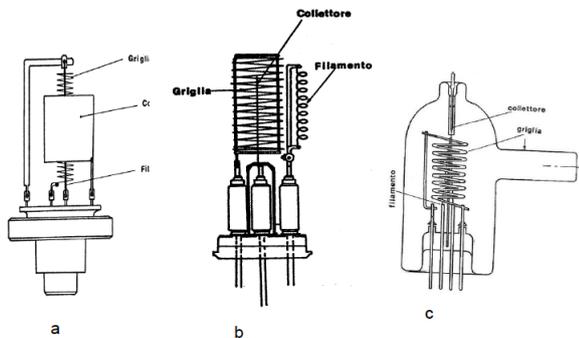


Figura 13: Schema di vacuometri a ionizzazione a confronto: a Geometria del vacuometro a ionizzazione tradizionale, b Bayard-Alpert nudo, c Bayard-Alpert in bulbo [36]. Questo limite si riteneva che fosse dovuto non al sistema di misura ma alle pompe.

Questo limite si riteneva che fosse dovuto non al sistema di misura ma alle pompe utilizzate, cioè le pompe a diffusione la cui velocità di pompaggio doveva azzerarsi a 10^{-6} Pa. Alla fine del 1930 non si era capaci di misurare pressioni inferiori a 10^{-6} Pa e noi, oggi, ne conosciamo il motivo (presenza di raggi X molli) che fu ipotizzato nel 1947 da Wayne Nottingham che, al primo incontro post-bellico "Physical Electronics Conference" tenuto al MIT, ipotizzò che il limite di 10^{-6} Pa non fosse dovuto alle pompe bensì al sistema di misura e cioè che fosse legato ai raggi X emessi dalla griglia per bombardamento degli elettroni provenienti dal filamento; questi raggi X colpendo il collettore provocano l'emissione di altri elettroni; il collettore così privato di cariche negative evidenzia una corrente positiva non distinguibile da quella dovuta agli ioni del gas residuo. La soluzione del problema risiede, dunque, nella conoscenza del modo di operare dei vacuometri e quindi dell'interazione dello strumento di misura con la grandezza da misurare.

Seguono vari tentativi in diversi paesi (USA, UK, Polonia, URSS, ...) [33] per ridurre l'effetto dei raggi X ma la più brillante e pratica soluzione fu prospettata nel 1950 da Bayard e Alpert, che, alla "Physical Electronics conference" al MIT, descrissero un metodo elegante per ridurre i raggi X, permettendo di misurare pressioni ultime di 10^{-9} Pa [34], ed inventarono una nuova geometria in cui le dimensioni del collettore sono fortemente ridotte (Fig. 13 b). Tuttavia anche

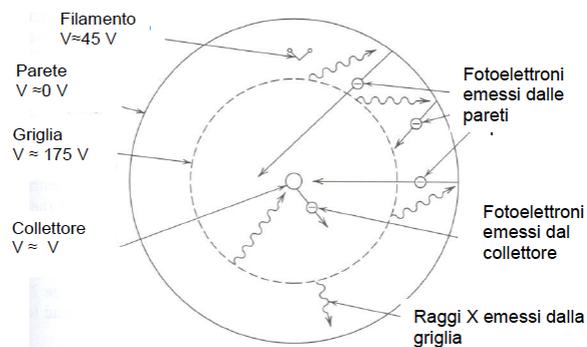


Figura 14: Interazioni del misuratore con l'ambiente in vuoto.

in questo caso si hanno fenomeni fisici che influiscono sul risultato della misurazione, come illustrato nella Fig. 14.

Questa nuova possibilità di misura ha rappresentato una vera pietra miliare permettendo lo sviluppo di moltissimi sistemi da vuoto in vetro, vetro-metallo, tutto metallo. Ha trascinato lo sviluppo di nuove pompe, di componenti e lo studio del comportamento dei materiali. Si riavviò il processo di ideazione e sviluppo di nuove pompe con scoperte ed impieghi di principi fisici del tutto nuovi come per esempio la scoperta delle cosiddette pompe a cattura, che intrappolano il gas sulla loro superficie tra le quali le ormai notissime [24, 31, 35, 36]:

1. pompe ioniche
2. pompe basate sulla chemisorzione (getter)
3. pompe basate sull'assorbimento fisico (crio-pompe)
4. miglioramento delle pompe turbomolecolari.

che appartengono ai giorni nostri e sono da molto tempo a catalogo delle principali industrie costruttrici di componenti per ultra-alto vuoto. Intorno alla metà del ventesimo secolo incominciano a comparire articoli che descrivono sistemi in grado di raggiungere vuoti molto spinti così come quello presentato nel 1958 da A. Venema e M. Badringa [37] che descrive un sistema in vetro fornito di pompa a diffusione di mercurio in grado di raggiungere pressioni dell'ordine di 10^{-10} Pa.

Intorno al 1960 si situa lo sviluppo significativo della tecnologia dell'ultra alto vuoto. Diventano

disponibili commercialmente flange metalliche con guarnizioni di rame (la famosa conflat) e pompe turbomolecolari, oltre alle ioniche ed alle criogeniche.

Diventa importante non solo misurare la pressione totale ma anche conoscere i tipi di molecole presenti nel sistema; si miniaturizzano spettrometri di massa a deflessione magnetica (1960)¹³ Nel 1958 W. Paul¹⁴ [38] descrive per primo lo spettrometro di massa a campo quadrupolare per l'analisi del gas residuo [39, 40, 41]. A. Davis descrive un tipico sistema per ultra-alto vuoto costituito da un sistema tutto metallico, pompa sputter ion, spettrometro di massa a deflessione magnetica con moltiplicatore di elettroni. Si raggiungevano pressioni totali ultime dell'ordine di 10^{-11} Pa, e pressioni parziali dell'ordine di 10^{-14} Pa. Un'altra data importante è il 1963, anno in cui viene identificato un altro fenomeno che limita la pressione ultima misurabile: il desorbimento di neutri e ioni positivi causato dal bombardamento elettronico degli strati di gas adsorbito sul materiale della griglia (o della sorgente ionica degli analizzatori di gas residuo). Si tratta del noto ESD= Electron Stimulated Desorption. Tali ioni non possono essere distinti dagli ioni dovuti al gas residuo.

Tra la fine degli anni sessanta e l'inizio degli anni settanta si costruiscono impianti, validi ancora oggi, equipaggiati con pompe ioniche e criogeniche, dotati di Bayard-Alpert e con guarnizioni in rame o anche in oro, per la produzione di film sottili. Ovviamente si tratta di impianti interamente metallici e in molti casi riscaldabili e sovente sono forniti di spettrometro quadrupolare e, nel caso che siano usati per la produzione di film sottili, di sorgenti di evaporazione a cannone elettronico. In seguito la ricerca di vuoti sempre più spinti (UHV=Ultra High Vacuum e XHV=Extremely High Vacuum) è stata stimolata dalle ricerche collegate con gli acceleratori di particelle e nel 1977 C. Benvenuti descrive un si-

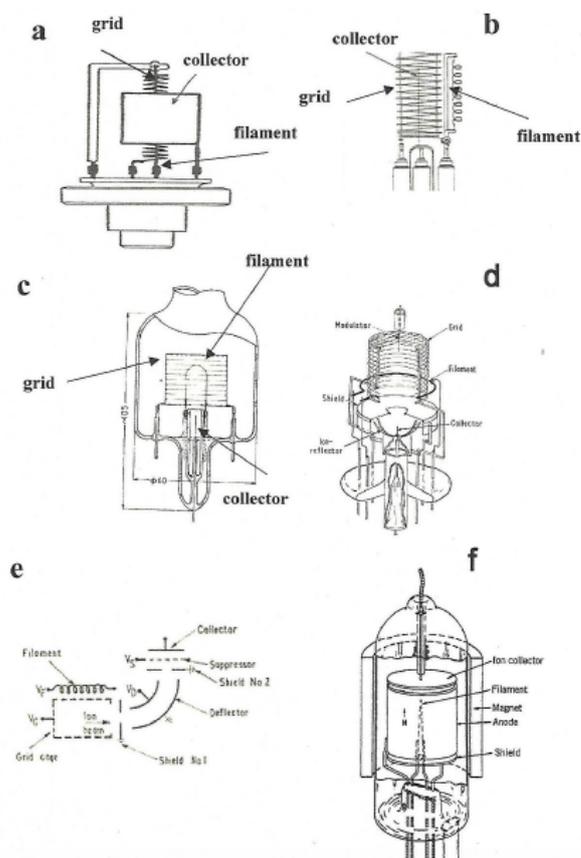


Figura 15: Confronto tra alcune geometrie di Vacuometri a ionizzazione a catodo caldo: a triodo tradizionale, b B-A, c vacuometro di Groskowsky, d vacuometro a estrattore di Redhead, e vacuometro di Helmer, f vacuometro a catodo caldo di Lafferty [29].

stema (e vacuometro) che raggiunge condizioni di XHV di 10^{-12} Pa, costituito da combinazioni di varie pompe (sputter-ion, criogeniche, getter non evaporabili) e vacuometri a ionizzazione speciale (Bent-beam).

Come si è visto, i nuovi campi di misura hanno reso possibili molte nuove realizzazioni di sistemi, componenti, nuove pompe basate su nuovi principi fisici. Moltissime pubblicazioni prendono in esame materiali e metodi per la loro pulizia, da quella chimica al bombardamento ionico, ma il limite di vuoto ultimo è ancora oggi dovuto agli strumenti di misura; infatti vacuometri a ionizzazione e spettrometri di massa sono sorgenti di gas e perturbano la misura.

Con lo scopo di estendere sempre più il limite delle più basse pressioni misurabili, tra il 1960 ed 1980 si situano molti miglioramenti dei vacuometri esistenti e la realizzazione di nuovi. Se ne citano solo alcuni. P. A. Redhead e al. nel

¹³Gli spettrometri di massa furono introdotti nel 1912 da Joseph Thomson e da Francis Aston. Entrambi ebbero il premio Nobel: Thomson nel 1906 e Aston nel 1922. Gli spettrometri a deflessione magnetica si basano sulla deviazione subita da un fascio di ioni in presenza di un campo elettrico e di un campo magnetico e sono poi rivelati in vario modo, all'inizio su lastre fotografiche.

¹⁴Nel 1989 divise il premio Nobel con N. F. Ramsey e H. G. Dehmelt, per questa invenzione.

1960 aggiunse un elettrodo¹⁵ al Bayard-Alpert simile al collettore e posto nello spazio interno alla griglia per servire come modulatore per misurare la corrente residua e valutare la corrente vera di ioni e J. M. Lafferty descrisse un vacuometro a catodo caldo con moltiplicatore di elettroni (SEM=Secondary Electron Multiplier) capace di misurare pressioni fino a 10-15 Pa. Nel 1966 Helmer rese commerciale il suo vacuometro, ma lo è stato per un breve periodo [42].

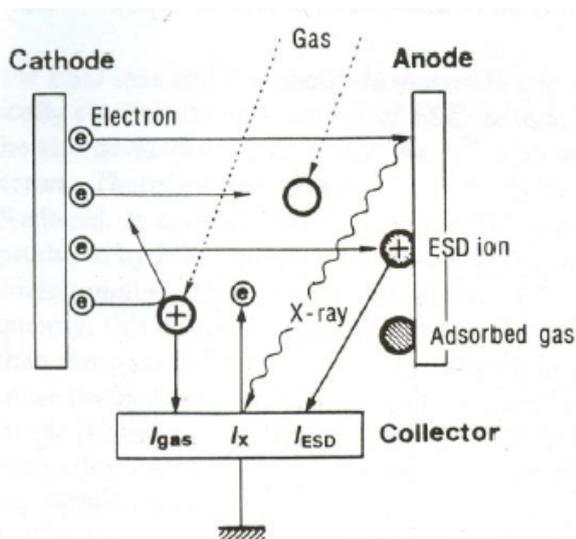


Figura 16: Fenomeni fisici in un vacuometro a ionizzazione, oltre ai raggi X, si hanno ioni prodotti per stimolazione elettronica [52].

C. Benvenuti e M.H. Hauer nel 1980 presentarono all'Ottavo Congresso Internazionale del Vuoto i risultati di loro numerosi studi, eseguiti al CERN di Ginevra, su sistemi ad ultra alto vuoto in cui si raggiungevano pressioni fino a 2.5×10^{-12} Pa misurate con un vacuometro tipo Helmer migliorato con un catodo di tungsteno toriato e con il diametro della griglia allargato ed altri miglioramenti [43]. Le realizzazioni di impianti vanno di pari passo con lo studio di misuratori atti a rivelare sempre più ampi campi di pressione. Infatti, molti studi sono stati condotti sui limiti sia ad alta sia a bassa pressione di va-

¹⁵La più bassa pressione misurabile con un Bayard-Alper è limitata dalla corrente residua che è indipendente dalla pressione. Essa è il risultato della fotoemissione dal collettore di ioni causata dai raggi X molli e da radiazioni ultra violette emesse dal filamento. L'introduzione del nuovo elettrodo che può essere o alla tensione di griglia o quella del collettore. Si veda al proposito la breve nota di P.A. Redhead, Modulated Bayard-Alpert gauge su Rev. Sci. Instrum, 1959, 31, 343-344.

cuometri sia a catodo caldo [44, 45, 46, 47, 48, 49] sia a catodo freddo del tipo Penning e magnetron che, con l'introduzione di un campo magnetico, consentono di allungare il cammino degli elettroni e ridurre la corrente di emissione dal filamento [50].

Si è così passati dalla tecnologia dell'ultra-alto vuoto (UHV, $\approx 10^{-6}$ Pa $> p > 10^{-10}$ Pa), a quella del vuoto estremamente basso (XHV, $\approx 10^{-10}$ Pa $> p > 10^{-13}$ Pa). Qui si situa, oltre ai lavori già citati a cavallo tra i due intervalli di pressione, la serie di lavori di Fumio Watanabe che analizzò le varie realizzazioni di misuratori di vuoto nella sua "never-ending story towards XHV pressure measurements" [51]. Egli condusse numerosi studi sul desorbimento da vari materiali e sui vari fenomeni fisici (Fig. 15) che influenzano le misure di vuoto. Partendo dalla considerazione che il limite di raggi X del vacuometro B-A è stato ridotto di 500 volte riducendo le dimensioni del collettore, ha pensato che si potesse ridurre ulteriormente tale limite diminuendo ancora la dimensione del collettore fino a "zero". Dal momento che è impossibile realizzare un collettore di dimensioni nulle, F. Watanabe [53, 54, 55, 56, 57] ha adottato il collettore virtuale o puntiforme e la disposizione di collettore, griglia e filamento non allineati. Tuttavia è giunto alla conclusione che non fosse possibile ridurre il limite inferiore di pressione oltre 10^{-13} Pa perché continuano ad essere presenti effetti dovuti ai raggi X così come il desorbimento dai materiali presenti nel sistema ed anche emissione di ioni per stimolazione elettronica.

In tempi recenti è stata realizzata la miniaturizzazione di vacuometri a scala millimetrica o anche micrometrica, anche se in alcuni casi i principi fondamentali della fisica non consentono una drastica miniaturizzazione. Come descritto da S. Wilfert e Chr. Edelman [58] possono essere considerati due metodi per ridurre drasticamente le dimensioni dei vacuometri:

- disporre di misuratori in cui l'unità di controllo e misura è montata sul sensore stesso, per ridurre l'influenza dei cavi;
- ridurre il volume del vacuometro per montarne più di uno nella camera di prova e diversi sensori nello stesso sistema e produrre con le moderne tecnologie un gran numero di misuratori e, conseguentemente, anche abbassarne il prezzo.

Mentre può essere relativamente facile produrre vacuometri miniaturizzati per pressione fino a 1 Pa (diaframma), 10^{-3} (spinning rotor) o anche 10^{-5} Pa (sensori termici a film sottile) appare meno praticabile miniaturizzare i vacuometri a ionizzazione senza una riduzione drastica della loro sensibilità [33].

Necessità di tarature

Tutti i misuratori che sono stati presi in esame sono di tipo secondario¹⁶, eccetto le colonne di liquido, quindi necessitano di riferimenti primari. Inoltre tutti i vacuometri sia a ionizzazione sia a conducibilità termica variano il loro comportamento (sensibilità) con il tipo di gas presente nel sistema. Knudsen ha tarato il suo vacuometro con un sistema basato sull'espansione statica. Da allora i sistemi primari sono stati costruiti a livello industriale o di laboratorio di ricerca, ma attualmente sono pressoché disponibili solo presso gli Istituti Nazionali di Metrologia (NMI) che sviluppano strumentazione in cui la pressione viene misurata direttamente come una forza su una superficie o in sistemi in cui la pressione viene calcolata dalle leggi del gas da misure di grandezze di base del Sistema Internazionale di unità di misura [22, 59, 60, 61, 62]. Per pressioni inferiori a 10^{-3} Pa fino a 10^{-4} Pa, sono disponibili diversi sistemi di NMI (Germania, India, Italia, Corea, Regno Unito, ecc.) basati sull'espansione statica, o metodo di Knudsen, mediante espansione singola o multipla. Con espansione multipla la pressione varia da pochi Pascal fino a 10^{-5} – 10^{-6} Pa [59] con incertezza relativa di poche parti in 10^{-4} o 10^{-3} , il cui valore aumenta con il numero dei volumi coinvolti. Per pressioni inferiori a 10^{-1} Pa fino a 10^{-7} Pa sono disponibili sistemi basati sull'espansione continua o dinamica descritta per la prima volta da Normand [61]. Tali dispositivi sono generalmente costituiti da una camera di taratura collegata ad una camera di pompaggio attraverso

¹⁶Si intende per secondario uno strumento in cui la grandezza da misurare è derivata da misurazioni di altre grandezze da cui essa dipende, per esempio la deformazione di una membrana, la variazione di resistenza elettrica di un filamento o la corrente di ioni. Pertanto per la loro taratura è necessario costruire sistemi riferiti direttamente alle grandezze del Sistema Internazionale di unità di misura.

o una conduttanza. Il sistema di pompaggio è spesso una combinazione di pompe turbomolecolari e meccaniche e la conduttanza è più simile possibile ad un foro in parete sottile per poterne calcolare il valore mediante metodi analitici o di Monte Carlo [29, 63, 64]. Questi impianti devono essere dotati di un sistema di generazione e misurazione del flusso di gas che spesso è basato sul metodo a volume variabile e pressione costante [65, 66]. Il limite inferiore può essere compreso tra 10^{-6} e 10^{-7} Pa. Per pressioni più basse, dal punto di vista storico va menzionato il sistema di fasci molecolari del PTB (Physikalisch-Technische Bundesanstalt) realizzato molti anni fa da G. Messer e G. Grosse [67]. Per tarature nel campo di XHV per il momento l'unica possibilità è collegata all'utilizzo del metodo di espansione continua con un divisore di flusso gas [59, 69]. Ma anche con il divisore di flusso i sistemi primari non coprono il campo di misura completo di tutti i misuratori a ionizzazione disponibili. Per studiare il comportamento di vacuometri a ionizzazione sono stati costruiti diversi piccoli sistemi raggiungendo pressioni [69] inferiori a 10^{-11} Pa. A questi livelli uno dei principali inconvenienti è rappresentato dal degassamento dalle pareti della camera. Sono stati eseguiti studi con camere ricoperte di TiN e pompate con pompe turbomolecolari in grado di raggiungere vuoti nel campo di 1×10^{-9} Pa.

Conclusioni

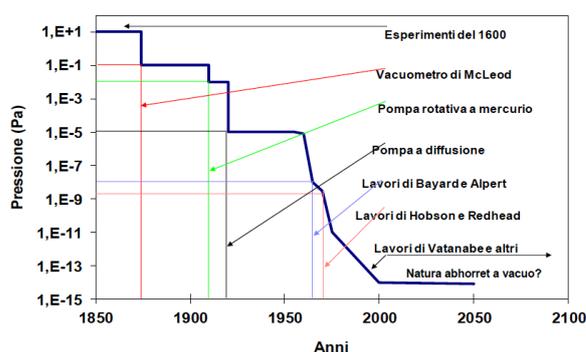


Figura 17: I periodi più importanti nello sviluppo della tecnologia del vuoto.

I tre citati esperimenti, insieme con molti altri, sono ancora oggi considerati da quanti misurano la pressione atmosferica, realizzano e misurano basse pressioni delle vere e proprie pietre miliari.

Non è esagerato dire che quegli esperimenti, che a noi oggi appaiono quasi elementari, furono altamente innovativi e complessi; essi diedero l'avvio a moltissimi altri esperimenti e speculazioni teoriche che hanno portato al grande impianto della teoria cinetica dei gas ed alle grandi realizzazioni tecnologiche dei secoli successivi. Anche nel nostro secolo vi sono state molte figure di scienziati, da Clausius a Dushman, a Gaede, a Knudsen, a Langmuir, a Pirani, a Alpert, per citarne solo alcuni, che hanno reso possibile con i loro studi ed esperienze la realizzazione di vuoti sempre più spinti e la possibilità di misurare pressioni sempre più basse a partire dall'atmosfera. Tuttavia per molto tempo la tecnologia languì (Fig.17) fino alla fine del diciannovesimo secolo per mancanza di applicazioni.

Soltanto quando si svilupparono industrie che richiedevano atmosfere sempre più rarefatte, come per esempio per la produzione delle lampade o, poi, per lo sviluppo dell'industria elettronica si ebbero delle nuove invenzioni. Poi, via via, si dovette rispondere alle esigenze del mondo della ricerca e si fecero dei salti in avanti nel raggiungimento di vuoti sempre più spinti e nella necessità di poterli quantificare in modo adeguato. Come si può vedere dalla Fig. 17 tra la fine del diciannovesimo e l'inizio del ventesimo secolo c'è un gran fiorire di ricerche e realizzazioni volte a misurare il vuoto dapprima non molto spinto e via via verso l'ultra alto vuoto. Infatti nel 1874 Mcleod inventò il famoso vacuometro a compressione che rimase fin quasi ai giorni nostri il più diffuso misuratore primario. Il secondo grande impulso, nel nostro secolo, venne dall'industria nucleare e della televisione e, nel campo della ricerca, dalla necessità di costruire acceleratori di particelle con vuoti ultra-spinti e impianti per la fusione termonucleare controllata.

Va infine notato che a mano a mano che i sistemi si facevano più sofisticati si rendeva necessario sviluppare metodi e mezzi per la ricerca delle perdite. Per fare ciò nel campo delle tecnologie qui considerate sono strumenti utili sia i vacuometri sia gli spettrometri di massa. Ma la ricerca delle perdite in un più ampio campo di applicazioni costituisce oggi una disciplina a parte con sue regole e norme.



Tipo di vacuometro	Intervallo di pressione (Pa)	Primo sviluppo
Colonne a liquido manobarometro a mercurio	$10^5 - 10$	metà 1600
Meccanici diaframmi vari	$10^5 - 10$	Leibniz 1702
Bourdon	$10^5 - 3000$	Bourdon metà 1800
diaframmi capacitivi	$10^5 - 10^{-2}$	Olsen e Hirst 1929
Compressione McLeod	$1 - 10^{-5}$	McLeod 1874
Viscosità vacuometro a torsione	$100 - 10^{-2}$	Sutherland 1897
fibra oscillante	$100 - 10^{-2}$	Langmuir 1913
Trasferimento di quantità di moto radiometro di Knudsen	$1000 - 10^{-1}$	Knudsen 1910
Conducibilità termica Pirani	atmosfera - 10^{-1}	Pirani 1906
termocoppia	atmosfera - 10^{-1}	Voegé 1906
Ionizzazione catodo caldo		
con collettore grande	$10^{-1} - 10^{-6}$	Von Bayer 1909
catodo freddo	$10^{-1} - 10^{-4}$	Pening 1937

Figura 18: Dal 1660 al 1945

Tipo di vacuometro	Intervallo di pressione (Pa)	Primo sviluppo
Catodo caldo Bayard - Alpert	$10^{-2} - 10^{-9}$	Bayard - Alpert 1950
Bayard - Alpert modulato	$10^{-2} - 10^{-11}$	Redhead 1960
Estrattore	$10^{-2} - 10^{-10}$	Redhead 1966
Helmer	$10^{-3} - 10^{-12}$	Helmer - Hayward 1964
Magnetron a catodo caldo	$10^{-4} - 10^{-16}$	Conn - Dalgligh 1954
Realizzazioni varie	$10^{-2} - 10^{-12}$	da Groszkowski (1960-1970) a Watanabe (dopo 1990)
Catodo freddo Magnetron invertito	$10^{-2} - 10^{-10}$	Beck - Brisbane 1952
Magnetron	$10^{-3} - 10^{-11}$	Redhead 1959
Radioattivi	$300 - 10^{-2}$	Dowing - Hellen 1946
Trasferimento di quantità di rotazione (spinning rotor)	$1 - 10^{-3}$	Fremerey dopo il 1970
Diaframmi	$10^5 - 10$	autori vari fino ad oggi
Conducibilità termica	$10^4 - 10^{-1}$	autori vari fino ad oggi

Figura 19: Dal 1945 ad oggi

- [1] F. Scandone, Introduzione al primo numero della rivista specializzata "il vuoto, scienza e tecnica", V.1, N.1 Sett. (1968).
- [2] J. Sparnaay, Adventures in vacuum, North Holland, (1992).
- [3] E. Madey, Early applications of vacuum, J.V.Sci.Technol, A2 (2), (1984), 110-117
- [4] J.M. Lafferty, History of American Vacuum Society, J. Vac. Sci. Tech. A(2)2, 1984, pp. 104 - 109

- [5] W.E.K. Middleton, *The history of the Barometer*, Johns Hopkins University, Baltimore, 1964, pp. 3 - 5 (libro)
- [6] R. N. Peacock, *Vacuum gauges in Foundation of vacuum science and technology* ed. Da J.M. Lafferty, John Wiley and Sons New York 1997
- [7] M.Bergoglio, A. Calcatelli, *L'eredità di Torricelli, la misura delle grandezze fisiche*, Atti della Società Torricelliana di Scienze e Lettere, Faenza, 1997
- [8] *Breve storia della meteorologia a Firenze dalle origini a oggi*, edito da G. Macchi, Firenze 1991
- [9] P.A. Redhead, *The Ultimate Vacuum*, *Vacuum*, V.53, 1999, 137-149
- [10] F. Affronti, *Atmosfera e Meteorologia*, S.T.E.M.-Mucchi, Modena, 1977
- [11] *Notes and records of published on line (2012)*, the Royal Society, Francis Hauksbee and his air pump
- [12] L..A. Geddes, *The direct and indirect measurement of blood pressure*, Year book medical publishers, Chicago 1970
- [13] M. H. Hablanian, *Comments on history of vacuum pumps*, *J. Vac. Sci. Technol.* A2(2),1984, 118-125
- [14] R.H. Dekosky, *William Crookes and the Quest for absolute vacuum in 1870s*, in *History of Vacuum Sci. and Technol.* Pubblicato da American Vacuum Society, 1983, 84-101
- [15] H. McLeod, *Apparatus for measurements of pressure of gas*, *Phyl. Mag.* 48, 1874, 3-8
- [16] J. H. Leck, *Pressure measurements in Vacuum Systems*, Chapman&hall, 1964 (libro)
- [17] R. K. Waits, *Edison's vacuum technology patents*, *J.Vac:Sci. Technol.* A2(1), 4, 2003, 881-891
- [18] brevetto francese n. 1849, brevetto U.S.A. 1852
- [19] J. Sullivan, *development of variable capacitance pressure transducers for vacuum applications*, *J. Vac. Sci. Technol.* A3 (3), 1985, 1721-1730
- [20] A. Calcatelli et al. *Results of the regional Key comparison EUROMET.M.P.K1.a in the pressure range from 0,1 Pa to 1000 Pa*, *Metrologia* V. 42, 07004, 2005
- [21] M. Bergoglio, A. Calcatelli, *The physical measurement of pressure in SI units and the extension to lower pressure*, *Vacuum* V64, 2002, 153-162
- [22] M. Dunkel, *Memories of Wolfgang Gaede on occasion of 100th anniversary of his birth*, *Vacuum*, V. 29, N.1, 11979, 3-8
- [23] *High vacuum pumps in History of vacuum Sci. and Technol.* Pubblicato da American Vacuum Society, 1983, 122-136
- [24] M. Knudsen, *The Kinetic Theory of Gases*, Methuen &Co Ltd, London, 1934
- [25] W. Steckelmacher, *Knudsen Gauges*, *Vacuum*, 1957, V.1, N. 4, 266-282
- [26] W. Beams, D. M. Spitzer and J.P. Wade, *Spinning rotor pressure gauge*, *Rev.Sci.Instrum.* V. 33, 1962, 151-155
- [27] J. K. Fremerey, *Spinning rotor vacuum gauges*, *Vacuum*, V. 32, N.10/11, 1982, 685-690
- [28] P. Rohl and W. Jitschin, *Performances of the spinning rotor gauge with novel transport device as a transfer standard for high vacuum*, *Vacuum*, V.88 N.7, 1988, 507-509
- [29] A. Calcatelli, *Vacuum Measurements and gauge calibration*, *V.39,7-8*, 1989, 799-807
- [30] H. Adam and W. Steckelmacher, *Marcello Piani(1880-1968). History of vacuum science and Technology*, V.2, A volume commemorating the 40h anniversary of the American Vacuum Society, edited by P.A. Redhead y AVS, 1994
- [31] *Foundations of vacuum science and technology*, ed. J. M. Lafferty, 1997 (libro)
- [32] O.E. Buckley, *An ionization manometer*, *Proc. Natl. Acad. Sci.*, 2, 1916, 683-685
- [33] A. Calcatelli, *The development of vacuum measurements down to extremely high vacuum*, *Measurement*, 46, 2013, 1029-1039
- [34] R. Bayard and D. Alpert, *extension of the low pressure range of ionization gauge*, *Rev. Sci. Instrum.*, V. 21, 1950, 570-572
- [35] *Handbook of vacuum technology*, edito da K. Jousten, Wiley-Blackwell, 2008, edizione inglese
- [36] B. Ferrario, *Introduzione alla tecnologia del vuoto*, Paravia, Bologna 1999
- [37] A.Venema, M. Bandringa, *The production and measurements of ultra-high vacua*, *Philips Techn. Rev.* 1958/59, V. 20, N. 6, 145- 176
- [38] W. Paul, *Electromagnetic traps for charged and neutral particles*, Nobel Lecture, December 8, 1989
- [39] P. H. Dawson, *Quadrupole mass spectrometry and its applications*. New York, AIP, 1995.
- [40] J. Roboz, *Introduction to mass spectrometry*, interscience publishers, New York, 1968
- [41] *Autori vari, Quadrupole mass spectrometry and its application*, Ed P. Dowson, Elsevier Scientific Publishing Co., New York, 1976
- [42] P.A. Redhead, J.P.Hobson, E.U.Kornelsen, *The physical Basis of Ultra-high vacuum*, American Institute of Physics, 1994 (libro)
- [43] C. Benvenuti and M.H.Hauer, *Proceeding of Eight International Vacuum Congree*, Cannes, 1980, V.II,199
- [44] H. Kuo, *An approach to the non-linearity of an ionization vacuum gauge at the upper limit of the measured pressure*, *Vacuum* V. 31N. 7, 1981, 303-308
- [45] G.K.T. Conn and al., *A thermionic ionization gauge of high sensitivity employing a magnetic field*, *Rev. Sci. Instrum.*, 31, 1954, 412-416
- [46] J. M. Lafferty, *Boride cathodes*, *J. Appl. Phys.*, V. 22 N. 3, 1951, 299-309
- [47] H. U. Becker and G. Messer, *Sensitivity dependence on collector surface properties in ion gauges*, *Proc 8th Intern. Vac. Congr. Cannes*, 1980, 234-237
- [48] P. E. Gear, *The choice of cathode material in a hot cathode ionization gauge*, *Vacuum*, Vol 26 N.1, 1976, 3-10

- [49] P. A. Redhead, The magnetron gauge: a cold-cathode vacuum gauge, *Ca. J. Phys.* Vol. 37, 1959, 1260-1271
- [50] F. Watanabe, My never-ending story towards XHV pressure measurements, *Vacuum*, 53 (1999), 151
- [51] F. Watanabe, My never-ending story towards XHV pressure measurements, *Vacuum*, 53 (1999), 151
- [52] F. Watanabe, Total pressure measurements down to 10-12 Pa without electron stimulated desorption ions, *J. Vac. Sci. Technol.*, A11 (4), Ju/aug 1993, 1620
- [53] F. Watanabe, Point collector ionization gauge with spherical grid for measuring pressures lower than 10-11 Pa, *J. Vac. Sci. Technol.*, A 5 (2), 1987, 242-248
- [54] F. Watanabe, Ion spectroscopy gauge: Total pressure measurements down to 10-12 Pa with discrimination against electron-stimulated-desorption ions *J. Vac. Sci. Technol.*, A 10 (5), 1992, 3333-3339
- [55] F. Watanabe, Dominance of electron-stimulated desorption neutral species in ultra-high vacuum pressure measurements, *Vacuum*, 52, 1999, 333-338
- [56] F. Watanabe, H. Ishimaru, Separation of gas phase and electron-stimulated desorption ions in the modulated-ion-current pressure gauge, *J. Vac. Sci. Technol.*, A 5 (5), 1987, 2924-2926
- [57] F. Watanabe, S. Hiramatsu and H. Ishimaru, Modulating ion current pressure gauge, *Vacuum* V. 33 N. 5, 1983, 571-1578
- [58] St. Wilfert, Chr. Edelmann, Miniaturized vacuum gauges, *J. Vac. Sci. Technol.*, A 22(2), 2004, 309-3020
- [59] K. Jousten, Calibration and standards, in *Foundations of vacuum science and technology*, edited by J. M. Lafferty, John Wiley and sons, USA, 1997, 657-699
- [60] J. Greenwood, The reference gauge technique for static expansion ratio applied to NPL medium vacuum standard SEA3, *Vacuum*, V. 81, N. 4, 2006, 427-433
- [61] C. E. Normand, Use of a standard orifice in the calibration of vacuum gauges, *Trans. VIII AVS Annual Symp.*, 1, 1961., 534-543
- [62] C. R. Tilford, S. Dittmann and K.E. McCulloh, The national Bureau of Standards primary high-vacuum standard, *J. vac. Sci. Technol.*, A6(5), 1988, 2853-2859
- [63] B. P. Buttler, V. Music, F.J. Redgrave, Influence of the orifice geometry on the simplification of transmission probability calculations, *Vacuum* V.53, 1999, 163-166
- [64] M. Nniewinski, P. Szwemin, A. Calcatelli and M. Bergoglio, Evaluation of the conductance of the orifice of the new CNR-IMGC dynamic expander, *Metrologia*, v. 36, n.6, 1999, 555-559
- [65] A. Calcatelli, G. Reiteri and G. Rumiano, Gas flow measurements connected with the continuous expansion system, *Proc. Int. Symp. On pressure/vacuum*, IMEKO Tc 16, Sept. 2003, Beijing, Acta Metrologica Sinica Press, 29-35
- [66] K. Jousten, H. Menzer, D. Wandrey and R. Nieprashk, New fully automated, primary standard for generating vacuum pressures between 10-10 Pa and 3×10^{-2} Pa with respect to residual pressure, *Metrologia* V. 36, N. 6, 1999, 493-497
- [67] G. Grosse and G. Messer, Calibration of vacuum gauges at pressure below 10^{-9} mbar with a molecular beam method, *Proc. Eight Int. congress*, Sept 22-26, 1980, Cannes, vol II, 255-258]
- [68] S. S. Hong, Y. H. Shin and K. H. Chung, measurement uncertainties for vacuum standards at Korea Institute of Standard and Science, *J. Vac. Sci. Technol.* A24(5), 2006, 1831-1838 27
- [69] H. Akimichi, M. Hirata, Generation and pressure measurement of extremely high vacuum (XHV) using a TiN coated chamber, *Metrologia*, 42, 2005, S184-S186

Anita Calcatelli: Laureata in fisica presso l'Università di Torino con una tesi sulla particelle elementari, ha lavorato nell'industria metalmeccanica prima di essere assunta dal CNR, dove si è occupata prima delle proprietà di strati sottili e poi di scienza della misura nell'ambito della metrologia delle basse pressioni nel campo del vuoto e, in generale, applicazioni della fisica e tecnologia del vuoto, anche con incarico di responsabile di reparto e con svariati incarichi internazionali, pubblicando numerosi articoli sulle riviste del settore. L'attività didattica si svolge con corsi di tipo specialistico sia per enti vari sia in ambito universitario. Ha svolto parecchi stage in laboratori stranieri in Canada, USA, India, Cina, Corea etc. prima come discente e poi come docente anche per conto di organismi internazionali (per es. UNIDO) L'attività è descritta in più di 150 articoli su riviste specializzate, monografie, contributi a congressi internazionali. Attualmente, in pensione, svolge attività nel campo della formazione principalmente di insegnanti di materie scientifiche, nell'ambito della scienza della misura. Ha fatto parte di gruppi femministi, tra cui il Gruppo Donne e Scienza di Torino e il Coordinamento Nazionale Donne e Scienza, interessandosi in particolare del rapporto fra le donne e la tecnologia con la presentazione di relazioni a convegni dedicati al tema donne e scienza. È socia dell'Associazione Donne e Scienza sin dalla nascita di questa organizzazione.

La tecnologia del vuoto

Gabriele Chiodini

Istituto Nazionale di Fisica Nucleare Sezione di Lecce

Negli ultimi decenni l'avanzamento della tecnologia del vuoto è stato enorme ed ha comportato l'utilizzo di tecnologie e materiali avanzati sviluppati inizialmente nell'ambito della ricerca pura. Ormai il raggiungimento di condizioni di vuoto sempre più spinto ed in impianti di dimensioni sempre più grandi è diventato indispensabile per moltissimi settori industriali, tecnologici e scientifici.

La rivoluzione scientifica moderna è stata possibile anche grazie all'avanzamento della tecnologia del vuoto che ha permesso di osservare nuovi fenomeni (ad esempio l'effetto fotoelettrico), sviluppare tecniche di analisi rivoluzionarie (ad esempio la spettroscopia di massa) e costruire macchinari d'indagine senza precedenti (ad esempio gli acceleratori di particelle¹).

Gli obiettivi scientifici più ambiziosi richiedono un costante miglioramento delle tecniche di vuoto e ciò ha portato a grandiose realizzazioni, impensabili sino a qualche anno fa.

¹Nel 1923 Wideröe progettò il betatrone scoprendo la famosa regola 2:1 tra campo accelerante e campo guida che garantisce un incremento di energia con orbita costante per un fascio di elettroni iniettato in una ciambella priva di aria. Wideröe non riuscì a costruire un betatrone funzionante per problemi di vuoto e solo nel 1940 Kerst reinventandolo ne realizzò uno accelerando elettroni fino a 2.2 milioni di elettronvolt. Il betatrone sebbene sviluppato per l'indagine di fenomeni nucleari è attualmente impiegato in medicina e nell'industria soprattutto come sorgente portatile di raggi X per eseguire radiografie.

Oggi, si raggiungono pressioni di 10^{-9} mbar (1 Atmosfera $\approx 10^3$ mbar) su grandi volumi ($> 100 \text{ m}^3$), come nelle macchine acceleratrici moderne [1]. Condizioni estreme di pressione, fino a circa 10^{-12} mbar, confrontabili con quelle presenti nello spazio siderale, sono riprodotte su piccoli volumi ($< 0.1 \text{ m}^3$) nella ricerca sui materiali. In Fig. 1 sono riportati i range di pressione impiegati nei vari ambiti di ricerca e nella parte alta la classificazione dei vari regimi di vuoto.

La forte spinta migliorativa nasce non solo da esigenze puramente scientifiche, ma anche da precise richieste di alcuni settori della tecnologia. La tecnica di produzione del vuoto si applica infatti ad un grande numero di impianti con scopi e necessità di livello di vuoto anche molto diversi come si vede nella Fig. 2.

Le ragioni per cui si desidera produrre il vuoto sono legate alla natura dell'applicazione considerata e quindi possono essere molteplici:

- impedire processi chimici e fisici causati dall'azione dei gas atmosferici, come combustione e scariche elettriche in filamenti caldi,
- rallentare i processi di decomposizione organica dovuti ad agenti aerobici (imballaggio sotto vuoto),
- aumentare il libero cammino medio di elettroni, ioni, atomi e molecole per permettere loro di raggiungere superfici o bersagli opportuni, senza urti con molecole estranee, ad esempio nella metallizzazione sotto vuoto e nelle macchine acceleratrici di particelle,

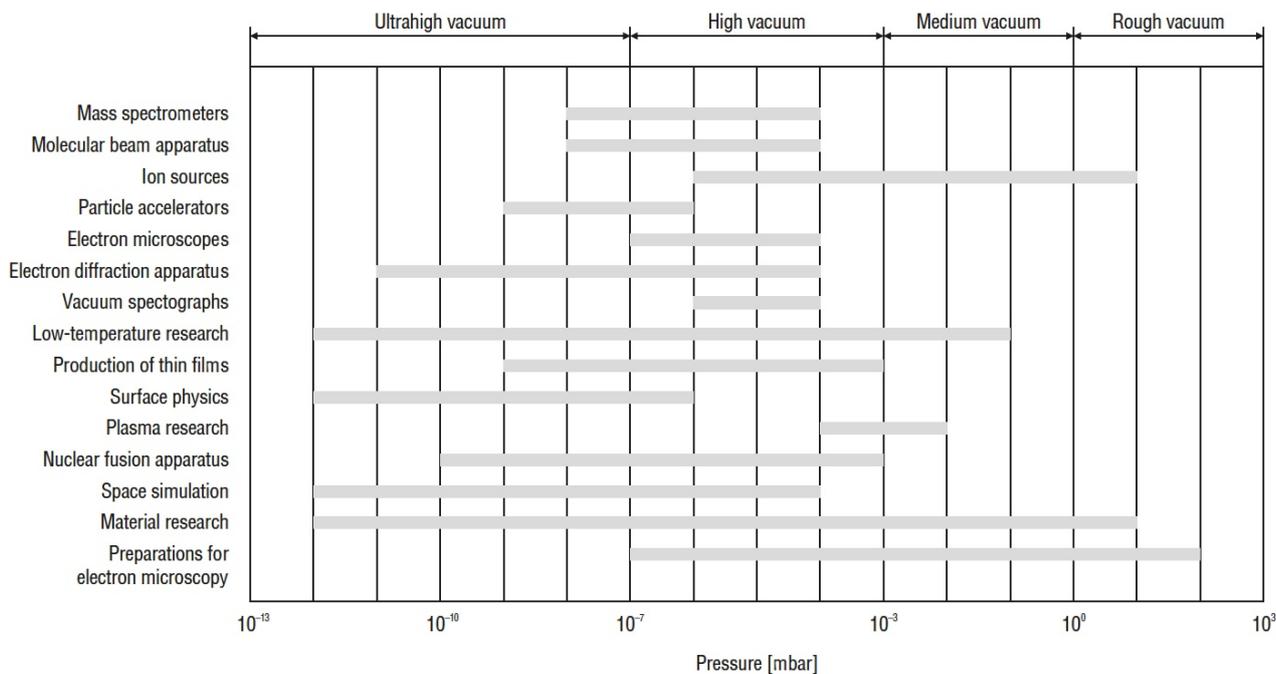


Figura 1: Range di pressione nei metodi di indagine ed analisi fisici e chimici. Credit: Leybold.

- ridurre la frequenza di collisione di molecole e atomi che compongono il gas con le superfici per allungare i tempi di contaminazione delle superfici stesse (studi di struttura e composizione di superfici solide; preparazione di film sottili),
- aumentare l'isolamento termico, ad esempio nei dewar e nei dispositivi a bassissima temperatura,
- eliminare i gas e vapori contenuti in un materiale, per esempio degasaggio di oli e liofilizzazione di materiali organici,
- simulare particolari situazioni fisiche come quelle incontrate dai satelliti e navicelle nello spazio.

L'ultra-alto-vuoto

Nel linguaggio comune si parla di vuoto quando c'è una depressione rispetto alla pressione atmosferica. Invece nella tecnologia del vuoto ci si riferisce a condizioni di alto vuoto (HV) e/o di ultra-alto-vuoto (UHV), quando il gas si trova in regime molecolare, cioè quando esso è così rarefatto che le collisioni di una molecola con le altre, sono molto meno frequenti delle collisioni con le pareti della camera che contiene il gas (vedi Fig. 3).

I livelli pratici di alto vuoto raggiungono valori di pressione di circa 10^{-6} mbar mentre i livelli di ultra-alto-vuoto sono caratterizzati da pressioni inferiori a 10^{-9} mbar. Infatti nei laboratori di ricerca spesso si lavora con livelli di vuoto fino a circa 10^{-5} mbar che si possono raggiungere con una pompa meccanica primaria seguita da una turbomolecolare secondaria.

Per pompare la camera a livelli di ultra-alto-vuoto oltre alla pompa turbomolecolare si usano una o più dei seguenti tipi di pompe: turbomolecolare, ionica, a sublimazione di titanio, getter non evaporativo e criopompa. Le pressioni UHV vengono misurate mediante misuratori di ioni, ad esempio un filamento caldo o di tipo magnetronico invertito.

Il raggiungimento del regime di ultra-alto-vuoto richiede diversi accorgimenti, l'impiego di materiali speciali e tecniche di processamento specifiche, come il preriscaldamento (*bake-out*) dell'intero sistema per diverse ore prima dell'uso che permette di rimuovere tracce residue di acqua e altri gas che sono adsorbite nelle superfici della camera.

Bake-out

La camera a vuoto è racchiusa in un materiale resistente al calore e riscaldato ad una temperatura di circa $180\text{ }^{\circ}\text{C}$ mediante dei forni. Dopo 24 ore di cottura, i forni vengono rimossi e la

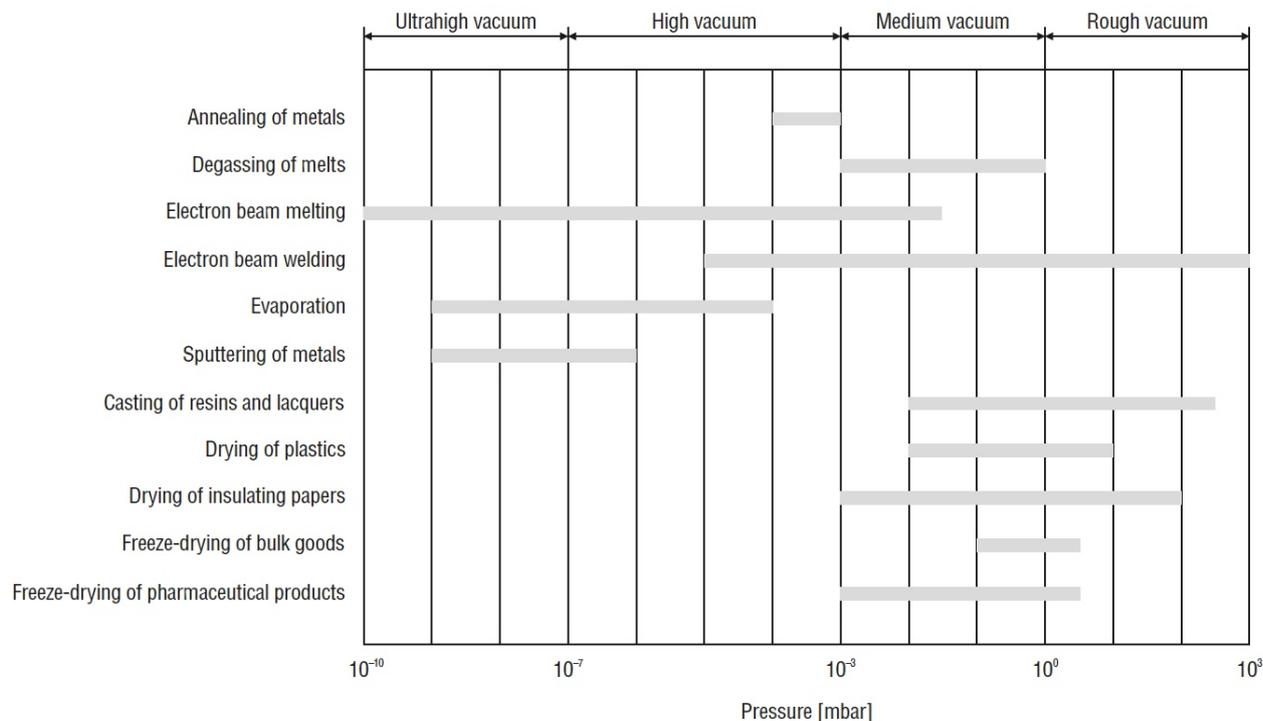


Figura 2: Range di pressione nei processi in vuoto industriali. Credit: Leybold.

camera si raffredda. Una volta a temperatura ambiente, la camera può raggiungere una pressione nella regione UHV.

Il processo di riscaldamento elimina gli atomi di gas dalle superfici delle pareti della camera (se la camera non fosse sottoposta a questo procedimento si impiegherebbero mesi per raggiungere le condizioni di UHV).

Materiali per UHV

Le guarnizioni utilizzate tra i componenti in un sistema UHV devono impedire qualsiasi minima perdita. Le guarnizioni interamente metalliche sono le uniche in grado di mantenere l'integrità del UHV. Queste guarnizioni hanno dei bordi a lama di coltello su entrambe le facce ritagliate da un metallo morbido, tipicamente rame.

Le camere a vuoto fabbricate in acciaio inox sono ampiamente utilizzate sia in regime di alto vuoto che di ultra-alto-vuoto. Tuttavia, ci sono altri materiali da costruzione che possono avere dei vantaggi: alluminio, titanio, rame e metalli specializzati. Ad esempio, l'alluminio rilascia meno idrogeno, vapore acqueo e vapori di idrocarburi nell'ambiente di vuoto consentendo cicli più rapidi e con meno pompaggio. Le camere da vuoto in rame o in leghe di rame vengono utilizzate per applicazioni non magnetiche, radiofrequenze (RF) e ad alta conducibilità termica. Le camere da vuoto in titanio, con i loro bassi livelli di permeazione all'idrogeno, offrono proprietà di gettering (vedi pompa getter). Spesso ciascun componente dell'impianto è trattato a parte prima dell'assemblaggio per renderlo idoneo all'ambiente di ultra-alto-vuoto.

Regimi di flusso

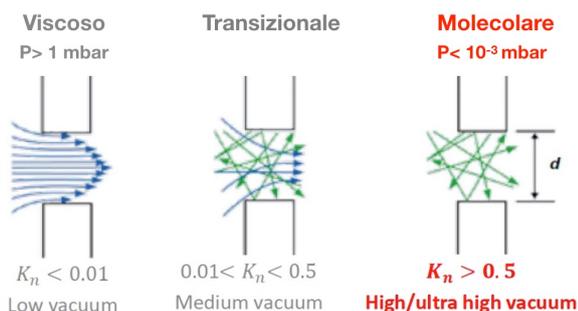


Figura 3: I regimi di vuoto sono distinti dal numero di Knudsen $K_n = \lambda/d$ definito dal rapporto tra libero cammino medio delle molecole λ e le dimensioni lineari del recipiente d .

Perdite in un impianto a vuoto

La fuoriuscita in generale di sostanze gassose (da superfici interne o da materiali di costruzione) è un problema significativo per i sistemi UHV. La perdita di gas da materiali utilizzati all'interno di una camera da vuoto viene ridotta al minimo grazie a un'attenta selezione di quelli con basse pressioni di vapore (ad esempio vetro, acciaio inossidabile, ceramica). Le plastiche e le colle sono spesso sorgenti di gas e devono aver subito un processamento di degasaggio prima di essere usate in ambiente UHV.

La fuoriuscita dalle superfici della camera è un problema più sottile. A pressioni estremamente basse, cresce la probabilità che molecole di gas siano adsorbite dalle pareti della camera, quindi la superficie totale interna di una camera diventa un parametro più critico del volume della stessa.

Il vapore acqueo è una fonte significativa di gas residuo, particolarmente quando la camera viene aperta in aria poiché qualsiasi vapore acqueo presente assorbe altre contaminazioni ed evapora dalle superfici troppo lentamente per essere rimosso completamente da un pompaggio a temperatura ambiente, ma abbastanza velocemente da indurre un livello continuo di contaminazione di fondo. La rimozione dell'acqua e dei gas simili richiede generalmente il riscaldamento da 180 °C a 400 °C mentre le pompe a vuoto sono in funzione.

Durante l'uso, le pareti della camera possono essere raffreddate usando azoto liquido per ridurre ulteriormente il degasaggio. L'idrogeno e il monossido di carbonio sono i gas residui più comuni presenti dopo il riscaldamento perché rilasciati da materiali quali acciai inossidabili.

Precauzioni utili

Accorgimenti importanti per ottenere una pressione di ultra-alto-vuoto sono i seguenti:

- ridurre al minimo le dimensioni della camera ed in particolare la sua superficie interna;
- utilizzare alte velocità di pompaggio mediante più pompe a vuoto collegate in serie e/o in parallelo;
- impiegare un tubo ad alta conducibilità (grande diametro, circuito con ostruzioni minime, valvole, ecc.);

- utilizzare materiali a basso rilascio di gas dalle superfici (acciaio inossidabile, alluminio, titanio, ecc.);
- evitare di creare pozzi di gas intrappolati dietro bulloni, saldature di vuoti, ecc.;
- elettrolucidare tutte le parti metalliche interne dopo la lavorazione o la saldatura;
- usare materiali a bassa pressione di vapore (ceramica, vetro, metalli, teflon se non scalfito);
- scaldare il sistema per rimuovere acqua o idrocarburi adsorbiti alle pareti;
- raffreddare le pareti della camera da vuoto a temperature criogeniche durante l'uso;
- usare guanti per evitare tutte le tracce di idrocarburi presenti sulle mani.

I fondamenti del vuoto

La comprensione del regime di ultra-alto-vuoto nelle applicazioni reali richiede nozioni specifiche [2, 3, 4]. A queste basse pressioni il percorso libero medio di una molecola di gas è di decine di km, quindi le molecole del gas residuo si scontrano con le pareti della camera più frequentemente di quanto si scontrino tra loro. Conseguentemente, quasi tutte le interazioni del gas si svolgono sulle superfici della camera.

In queste condizioni il regime di flusso del gas è molecolare (opposto al flusso laminare dei fluidi) e gli effetti di superficie sono fondamentali nel determinare il vuoto limite.

La legge di Ohm del vuoto

Nei sistemi in cui si pratica l'alto vuoto si instaura una condizione di regime dinamico tra la continua immissione di gas (dovuta alla permeazione, ad eventuali microperdite ed al rilascio di molecole dalle pareti, detto *fenomeno di degasaggio*) e l'evacuazione continua da parte delle pompe da vuoto connesse alla camera.

Questo stato stazionario è descritto utilizzando due quantità:

- il flusso di gas S , misurato in m^3s^{-1} , detto anche velocità di pompaggio $S = dV/dt$ dove dV è volume di materia elementare che attraversa la superficie nel condotto nell'intervallo di tempo elementare dt ,

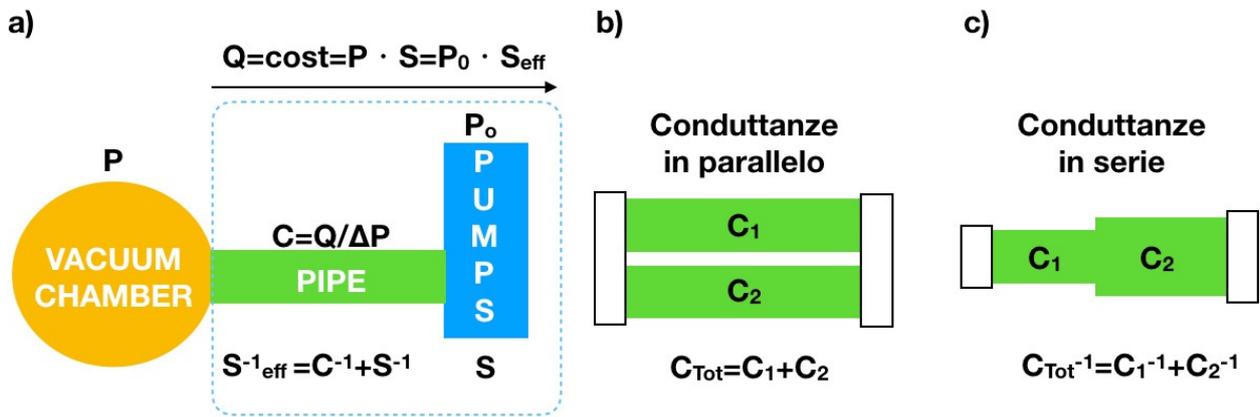


Figura 4: a) Definizione di conduttanza e di velocità di pompaggio efficace. b) e c) Conduttanza in serie e parallelo.

- la portata Q (*throughput*) misurata in $\text{m}^3 \cdot \text{mbar} \cdot \text{s}^{-1}$ ed espressa dalla relazione $Q = P \cdot S$ dove P è la pressione del gas. Essa è direttamente connessa con il flusso di materia nel recipiente a pressione P .

Supponiamo di avere un condotto che collega due recipienti alle pressioni P_1 e P_2 , rispettivamente. Knudsen ha dimostrato che in condizioni di regime molecolare, assumendo una distribuzione Maxwelliana di velocità, la portata Q dipende dalla geometria del condotto e dalla differenza di pressione, cioè $Q = C(P_1 - P_2)$ dove il coefficiente C prende il nome di *conduttanza* e ha le stesse dimensioni della velocità di pompaggio S .

Tale legge è nota anche come la legge di Ohm della fluidodinamica per la somiglianza formale con la legge che lega la differenza di potenziale applicata ai capi di una resistenza elettrica e la corrente che vi scorre (Fig. 4).

Risulta molto pratico utilizzare il concetto di velocità di pompaggio efficace S_{eff} che consente di includere nella velocità di pompaggio della pompa anche l'effetto della presenza delle conduttanze dei tubi di connessione alla camera da vuoto.

La relazione $Q = P S_{\text{eff}} = S_{\text{pompa}} P_{\text{pompa}}$ definisce la velocità di pompaggio efficace quella che si ha alla pressione della camera a vuoto P a parità di throughput. Dalla legge di Ohm della fluidodinamica otteniamo $Q = C(P - P_{\text{pompa}})$. Eliminando la pressione della pompa dall'espressione precedente si ottiene $S_{\text{eff}}^{-1} = C^{-1} + S_{\text{pompa}}^{-1}$. Da questa relazione risulta chiaro quanto la conduttanza della connessione riduca l'effettiva portata della pompa.

Effetti di superficie

La quantità di gas rilasciata dalle pareti determina la pressione finale di lavoro a partire da condizioni di vuoto medio (Fig. 5a).

L'evacuazione del gas di volume è un processo relativamente rapido e, superato il vuoto medio, il pompaggio delle camere subisce un rallentamento dettato dalla frequenza con cui le pareti rilasciano molecole di gas e quest'ultimo fenomeno determina la pressione finale ottenibile (Fig. 5b).

In Fig. 6 sono schematizzate le diverse interazioni tra le molecole del gas residuo e le pareti della camera a vuoto che verranno ora discusse.

Tra tutte le molecole che urtano la superficie di una camera da vuoto solo una certa frazione vi aderisce concorrendo a formare un monostato di ricopertura. Diremo allora che le molecole che aderiscono alla superficie sono state soggette al fenomeno di *adsorbimento*. Se poi le molecole del gas si legano debolmente con le molecole della stessa specie già adsorbite, allora concorrono a creare altri strati depositati sulla parete: parleremo in tal caso di fenomeno di *condensazione*. In pratica si ha un cambiamento di fase dallo stato gassoso a quello liquido o solido. In questi processi gioca un ruolo fondamentale la temperatura della superficie.

Infine, se le molecole del gas diffondono all'interno della matrice cristallina della parete o penetrano nelle sue porosità a tal punto da essere trattenute, allora parleremo di fenomeno di *assorbimento* (o più semplicemente *assorbimento*). La quantità di gas che può essere assorbita da una parete dipende dal coefficiente di solubilità e dalla costante di diffusione del gas nel solido. In realtà pochi gas diffondono signifi-

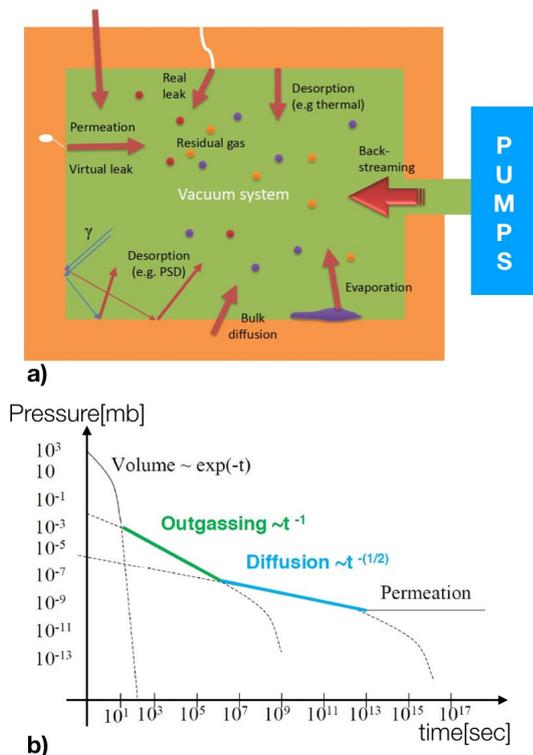


Figura 5: a) Le sorgenti di gas residuo in una camera a vuoto. b) Andamento nel tempo della pressione in una camera a vuoto dall'inizio dell'evacuazione del gas. L'evacuazione del gas di volume è rapida (legge di decadimento esponenziale) mentre l'evacuazione dei gas residui è più lenta e caratterizzata da una legge di potenza specifica della sorgente. La permeazione, sempre presente, ed eventuali microperdite, invece, determinano la pressione asintotica finale del sistema.

cattivamente nei materiali solidi, fatta eccezione per l'idrogeno e l'elio.

La molecola che aderisce alla superficie è catturata da un potenziale attrattivo generato dall'interazione con la parete. Se la buca di potenziale ha dimensioni di 0.1-0.3 nm classifichiamo il processo come adsorbimento chimico (*chemisorzione*), mentre se ha dimensioni dell'ordine di 0.4 nm come adsorbimento fisico (*fisisorzione*).

Nella chemisorzione l'adesione della molecola alla parete implica la messa in comune di elettroni di valenza dei costituenti del gas e della parete. Tale processo è selettivo e dipende fortemente dalla natura del gas e delle pareti con cui interagisce. Inoltre l'adsorbimento chimico è responsabile della formazione di uno strato mono molecolare soltanto, visto che esso può aver luogo sino a che vi siano legami di

valenza disponibili. La chemisorzione è spesso accompagnata da rotture di legami chimici con conseguente dissociazione molecolare: a tale processo è associata un'energia di attivazione E_a .

L'adsorbimento (sia chimico che fisico) comporta una riduzione dei gradi di libertà delle particelle del gas e quindi il processo d'adsorbimento è esotermico. Il fenomeno opposto, cioè il rilascio di molecole dalle pareti, detto *desorbimento* o degasaggio, avviene con assorbimento di calore da parte del gas.

Dalle considerazioni fatte in precedenza è evidente che, per accelerare il rilascio delle molecole adsorbite occorre fornire energia alle pareti. Ciò viene fatto nei modi più disparati, quali il bombardamento della superficie tramite particelle cariche (ad esempio ad LHC) e radiazione elettromagnetica. Tuttavia, come già anticipato, il modo più semplice e di utilizzo comune nei sistemi d'alto vuoto, è il riscaldamento delle pareti dell'impianto.

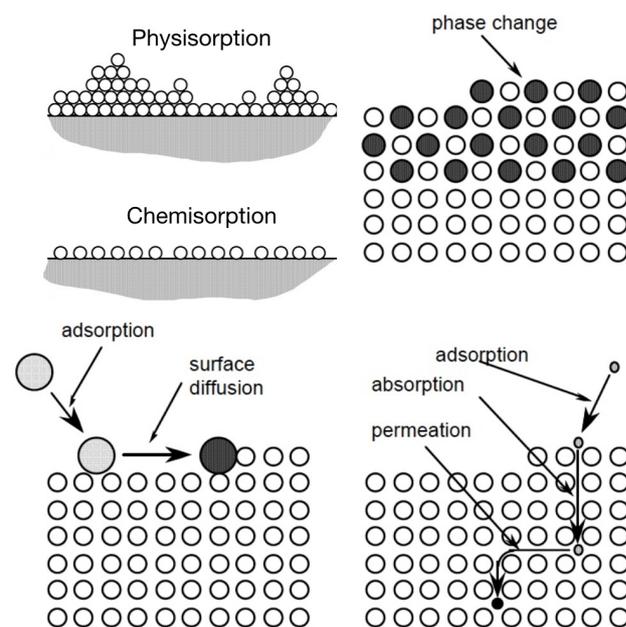


Figura 6: Fenomeni fisici responsabili della cattura sulla superficie della parete di una camera a vuoto delle molecole del gas residuo in condizioni di regime molecolare.

Tuttavia, se da un lato l'adsorbimento costituisce un aspetto negativo per il raggiungimento di un alto vuoto, dall'altro esso è abilmente sfruttato in alcuni casi fondamentali.

L'adsorbimento fisico è alla base delle pompe chiamate comunemente criogeniche (vedi paragrafo successivo). Quando si vuole evitare

l'utilizzo di pompe che possono lasciare negli impianti tracce di idrocarburi dovuti all'olio utilizzato nelle pompe, allora si opta per questi sistemi basati sull'utilizzo di polveri o granuli di materiale che presenta un alto rapporto superficie/volume e che viene raffreddato a bassa temperatura. Ad esempio le Zeoliti sono portate a 77 °K (temperatura d'ebollizione di azoto a pressione atmosferica), mentre le temperature d'esercizio delle pompe a carbone attivo sono tipicamente più basse (da 4 a 20 °K). L'uso delle basse temperature è cruciale perché i legami dovuti alle forze di Van der Waals della fisisorzione sono deboli e quindi l'energia termica delle molecole deve essere ridotta in modo da consentire l'intrappolamento nella buca di potenziale, poco profonda, del sito d'adsorbimento.

L'adsorbimento chimico trova applicazione in diversi tipi di pompe dette getter, la cui utilità è notevole soprattutto se i gas da adsorbire sono attivi. Tali pompe sono costituite da sottilissimi strati di metallo quali Ti, Ba, Zr, Ta o leghe come Zr-Al, Ti-Zr che vengono portate ad una temperatura d'esercizio tra i 20 e i 400 °C. Il titanio è spesso impiegato in sistemi che sono destinati a sopportare numerosi cicli di pressione tra la condizione di vuoto e quella di pressione atmosferica, mentre il bario è spesso depositato all'interno di piccoli sistemi da vuoto, come le valvole elettroniche e i tubi a raggi catodici che, una volta costruiti, restano sotto vuoto tutta la loro vita. Infine ricordiamo che l'adsorbimento chimico è un processo selettivo e questa proprietà è sfruttata nel processo di selezione dei gas rari. Questi ultimi sono tipicamente gas nobili che non reagiscono con i getter che al contrario adsorbono le impurezze presenti nel gas.

Le pompe da vuoto

Poiché il vuoto è indispensabile per molte applicazioni è necessario produrlo in ambienti o recipienti adatti attraverso opportuni dispositivi dette *pompe da vuoto* (Fig. 7).

Esistono diversi tipi di pompe da vuoto che coprono differenti intervalli di pressione. Ogni tipo di pompa è caratterizzato dai valori della velocità di pompaggio e della pressione limite. La velocità di pompaggio dipende dalla pressione all'ingresso della pompa e presenta sempre un andamento decrescente al diminuire della pressione di lavoro. Solo in alcuni casi

si ha una diminuzione della velocità di pompaggio a pressioni elevate. È utile sapere che la pressione finale P_f di un impianto a vuoto non coincide con la pressione limite P_l della pompa. Infatti la conduttanza C dei tubi che connettono il sistema alla pompa a vuoto causa una caduta di pressione inevitabile proporzionale alla portata Q secondo la relazione $Q = C(P_f - P_l)$ necessaria a far fluire il gas dal sistema alla pompa (vedi paragrafo sulla Legge di Ohm della fluidodinamica).

Le pompe rotative

Sono le pompe meccaniche comunemente usate per produrre il vuoto primario (basso e medio vuoto) partendo dalla pressione atmosferica e ne esistono di due tipi: a palette ed a pistone rotante.

La pompa rotativa a palette è costituita da un rotore provvisto di palette mobili che ruota eccentricamente in una cavità cilindrica connessa al condotto di ingresso e di uscita. Le palette vengono tenute a contatto con la superficie interna della cavità cilindrica da una molla ed un velo d'olio garantisce la tenuta. La variazione di volume delle camere crea depressioni (fase di aspirazione) e compressioni dell'aria (fase di scarico). Nella fase di aspirazione il gas viene estratto dalla camera da vuoto tramite il manicotto di aspirazione, successivamente grazie all'eccentricità del rotore la camera continua ad aumentare di volume. Una volta raggiunto il volume massimo il manicotto di aspirazione viene chiuso da una seconda palette mentre la camera della pompa comincia a diminuire di volume consentendo ai gas di essere espulsi dal manicotto di scarico (Figura 8).

In generale, nelle pompe meccaniche, la lubrificazione delle parti in moto e la tenuta da vuoto del sistema sono assicurate da oli speciali a bassa tensione di vapore. La pressione di scarico dei gas compressi deve essere più elevata della pressione atmosferica, per cui sono necessari rapporti di compressione (rapporti tra pressione di uscita e di ingresso) molto grandi (dell'ordine di 10^5) per ottenere una pressione finale di 10^{-2} mbar all'ingresso della pompa. La pressione finale delle pompe rotative ad un singolo stadio ricade tipicamente nell'intervallo che si estende da 10^{-2} mbar (per le pompe di piccola velocità di pompaggio $S < 50 \text{ m}^3/\text{h}$) a 10^{-3} mbar (per le pompe con velocità di pompaggio

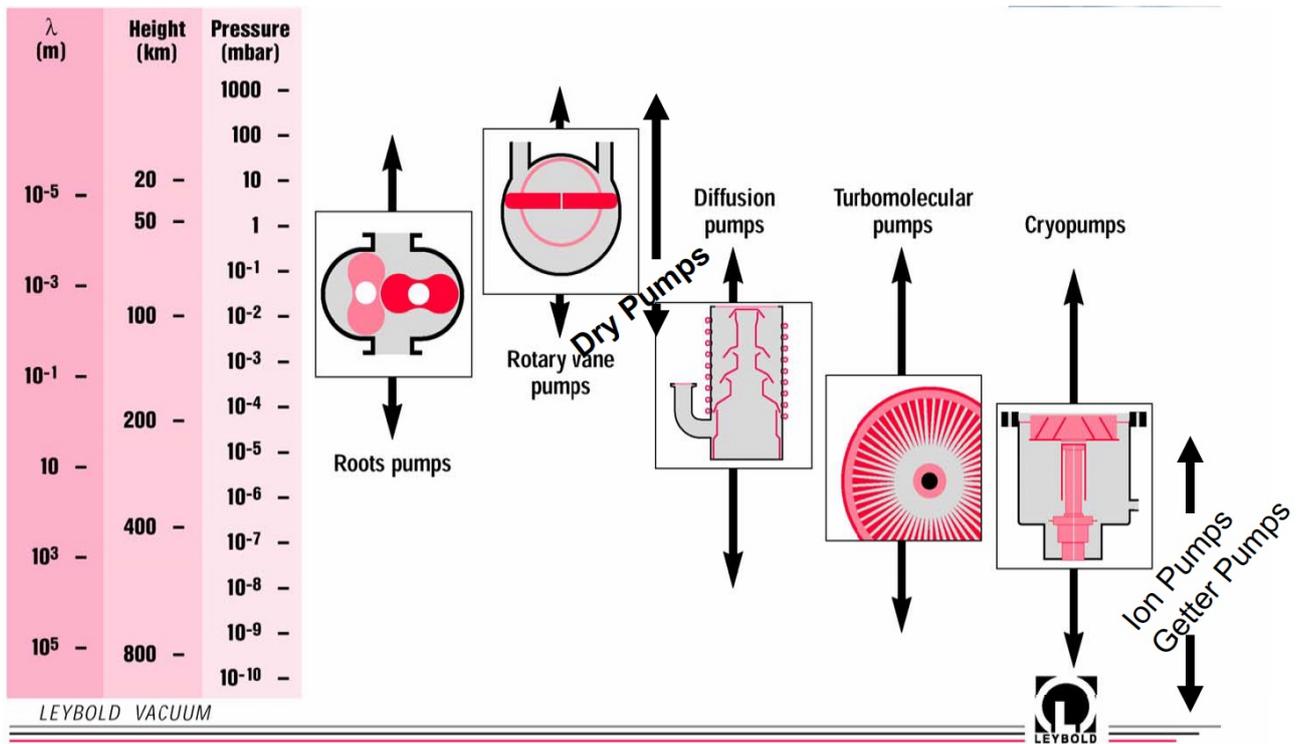


Figura 7: Tecnologie di pompe a vuoto per differenti range di pressione riportati nella terza colonna di sinistra. Nella prima e seconda colonna di sinistra sono riportati i valori del libero cammino medio molecolare e dell'altezza atmosferica corrispondenti a quei range di pressione.

più elevata). Tale limite non è determinato dal meccanismo di pompaggio, ma è imposto dalla solubilità dei gas nell'olio lubrificante. Infatti l'olio espulso attraverso la valvola di scarico è saturo dei gas o vapori che si sono in esso disciolti durante la fase di compressione, ritorna nel serbatoio dell'olio e successivamente è immesso di nuovo nella cavità cilindrica. Lì l'olio rilascia parte dei gas o vapori in esso disciolti ed in questo processo si possono determinare le condizioni per un sensibile peggioramento delle prestazioni della pompa. In particolare, quando i gas aspirati sono in prevalenza vapori che si condensano nella fase di compressione (ad esempio vapori di acqua), per evitare il verificarsi di questa circostanza si ricorre ad un particolare accorgimento: zavorrare la pompa con l'aria dell'ambiente (apertura del *gas ballast*).

Per avere pompe meccaniche con velocità di pompaggio ancora più elevata è stata concepita la pompa rotativa a pistone rotante. In essa l'albero di rotazione del rotore è coassiale rispetto alla cavità cilindrica mentre il corpo del rotore (camma) è eccentrico rispetto all'albero. Un pistone cavo trascinato dalla camma eccentrica pone in comunicazione il recipiente da evacuare con la cavità cilindrica. La camma

nella sua rotazione fa sì che il pistone scorra con moto alternativo nella guida (oscillante nella sua sede) e metta alternativamente la camera di compressione in comunicazione con il recipiente da evacuare oppure la isoli. Il rotore durante il suo moto comprime i gas fino ad espellerli nell'atmosfera attraverso la valvola di scarico. Notiamo infine che le pompe a pistone rotante per le loro caratteristiche costruttive sono più adatte delle pompe a palette per le grandi portate volumetriche (50-500 m³/h).

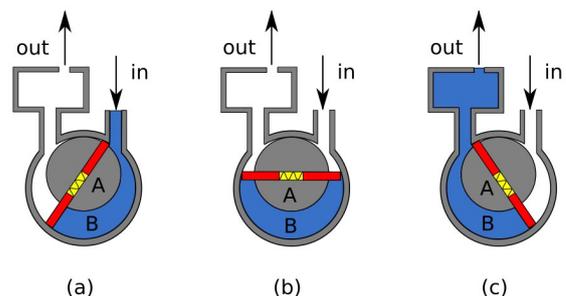


Figura 8: Schema di funzionamento di una pompa rotativa. In azzurro viene indicato il volume di fluido spostato durante un ciclo.

Le pompe a diffusione

Le pompe a diffusione non possono immettere i gas aspirati direttamente nell'atmosfera come nel caso delle pompe rotative, perchè la loro pressione di lavoro si estende da qualche 10^{-3} a 10^{-10} mbar per tutti i gas. Occorre prevedere allora che il recipiente da svuotare sia inizialmente connesso ad una pompa meccanica (pompa primaria) che porti il vuoto a valori tali da cadere nell'intervallo operativo della pompa a diffusione.

Nelle pompe a diffusione si utilizza un getto di molecole ad alta velocità di un vapore che ha la funzione di trasferire per urto quantità di moto alle molecole del gas da evacuare in direzione della bocca di pompaggio. Nelle pompe di prima generazione si utilizzavano vapori di mercurio, nelle pompe attuali si utilizzano oli speciali con tensioni di vapore a temperatura ambiente di $5 \cdot 10^{-9}$ - 10^{-10} mbar.

Le molecole di vapore, provenienti dal liquido in ebollizione localizzato alla base della pompa, passando attraverso una strozzatura anulare a forma di becco, acquistano una velocità di alcune centinaia di m/s diretta verso il basso, dove si trova la bocca di aspirazione connessa alla pompa meccanica che assicura il pre-vuoto, e verso la parete fredda della pompa. Le molecole del gas che arrivano alla bocca della pompa a diffusione vengono trascinate nel getto di vapore ed acquistano una quantità di moto diretta verso il basso. Quando le molecole di vapore del getto incidono sulla parete fredda vengono condensate e ritornano sotto forma di liquido nel bagno che sta alla base della pompa, assicurando la continuità del meccanismo di pompaggio. Il trasferimento di una quantità di moto netta alle molecole dei gas che attraversano il getto si traduce in una differenza di densità molecolare e quindi di pressione tra le due regioni separate dal getto di vapore (Figura 9).

Il peso molecolare degli oli impiegati è notevolmente superiore al peso molecolare dei gas residui (400 - 500 masse atomiche), ed il rapporto di compressione è più piccolo per gas residui di massa molecolare più bassa, quindi questo tipo di pompa è meno efficiente per i gas leggeri come l'idrogeno e l'elio.

Un tipico schema di pompa a diffusione a più stadi consiste di un cilindro con una flangia superiore per il collegamento al recipiente da evacuare e con un fondo contenente una resi-

stenza da fornello elettrico per produrre i vapori del fluido di pompaggio (negli attuali sistemi commerciali più diffusi si tratta di olio minerale particolare). La camicia esterna della pompa è raffreddata per circa due terzi della sua altezza con una serpentina in cui circola acqua.

Una bocca d'uscita posta in prossimità del fondo permette di scaricare i gas a pressione più elevata verso la pompa rotativa. Sulla bocca di connessione al recipiente sul condotto con la pompa meccanica sono poste trappole ben raffreddate (ad acqua o con freon o con azoto liquido) per evitare sia la diffusione del vapore d'olio nel recipiente da evacuare sia la diffusione del fluido nella pompa meccanica. La velocità di pompaggio in una pompa a diffusione varia con il tipo di gas. Notiamo che la pressione finale è determinata principalmente dalla tensione di vapore del fluido di pompaggio e più specificatamente dai prodotti che si formano attraverso il processo di rottura delle molecole del fluido stesso (*cracking*).

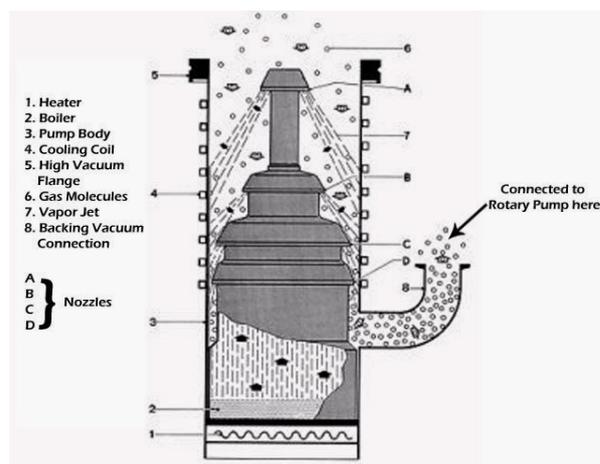


Figura 9: Schema di una pompa a diffusione.

Le pompe turbomolecolari

Le pompe turbomolecolari sono disponibili commercialmente dalla fine degli anni '50. Per il loro funzionamento è necessario disporre di una pompa meccanica primaria che mantenga una pressione di 10^{-3} mbar nella regione a pressione più elevata. Si ottengono così pressioni finali dell'ordine di 10^{-9} mbar e la velocità di pompaggio presenta un andamento praticamente costante in tutto l'intervallo operativo.

In questo tipo di pompe le molecole dei gas sono trascinate verso la bocca di evacuazione

da un sistema di palette opportunamente distanziate ed inclinate e poste in rapidissima rotazione all'interno di una cavità cilindrica. Le molecole che si trovano nelle due regioni del volume della pompa turbomolecolare separate dal sistema di palette in moto, hanno probabilità diverse di diffondere nei due sensi, ciò provoca densità molecolari differenti nelle due regioni e quindi un gradiente di pressione. In pratica le palette in rotazione urtano le molecole d'aria, spingendole verso le palette dello statore successivo che sono ferme, l'urto contro queste, grazie al loro orientamento, fa passare le molecole alle palette del rotore successivo, dove il processo si ripete. Queste palette sono orientate in maniera progressiva in modo da adattarsi alle differenti pressioni presenti all'interno della pompa turbomolecolare stessa (Figura 10).

Per avere un rapporto di pressione elevato ed un flusso netto elevato (requisiti base per una pompa da vuoto) si agisce sull'angolo di inclinazione delle palette e sul rapporto tra la distanza tra due palette successive e la loro larghezza. Nelle pompe turbomolecolari moderne la velocità di rotazione delle palette è di decine di migliaia di giri al minuto e la velocità periferica raggiunge centinaia di metri al secondo. Disponendo più sistemi di palette in cascata si ottengono rapporti di pressione sino a 10^6 . In una pompa turbo occorre assicurare l'assoluta assenza di agenti inquinanti del vuoto. Per questo motivo l'asse di rotazione della pompa è sospeso con cuscinetti magnetici che limitano l'attrito eliminando allo stesso tempo la necessità della lubrificazione.

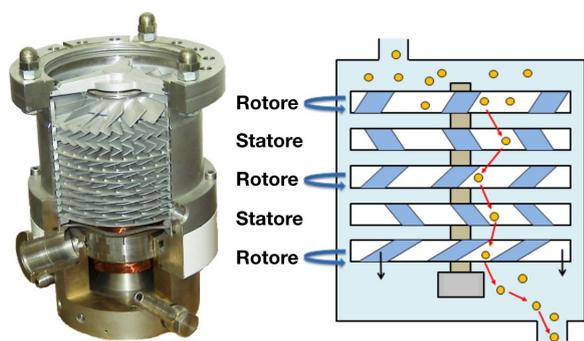


Figura 10: A sinistra la foto di una pompa turbomolecolare reale ed a destra il suo schema di funzionamento.

Le pompe a getter e ioniche

Queste pompe sono state sviluppate e commercializzate agli inizi degli anni '60. Esse sfruttano due processi distinti per evacuare le varie specie di gas presenti in un impianto a vuoto: l'adsorbimento chimico per i gas reattivi (ad esempio H_2 , N_2 , O_2 , H_2O , CO , CO_2) e l'assorbimento per i gas nobili o poco reattivi (He , Ne , Ar , CH_4 , ...).

Noi abbiamo già parlato nei paragrafi precedenti di adsorbimento chimico dei gas reattivi (*gettering*). Qui ricordiamo come alcuni materiali come il titanio, lo zirconio ed il tantalio posseggono notevoli capacità di adsorbimento chimico a temperatura ambiente nei confronti dei gas reattivi presenti nell'atmosfera.

Il titanio è tipicamente preferito per le applicazioni del vuoto a causa della relativamente alta tensione di vapore e della sua reattività chimica a largo spettro. In una pompa ionica la sorgente di titanio è riscaldata sino alla temperatura di sublimazione ($1500\text{ }^\circ\text{C}$) in maniera che produca un deposito (film) di titanio altamente reattivo sulle pareti che circondano la sorgente. Le molecole dei gas attivi che incidono sul film di titanio reagiscono chimicamente formando composti stabili quali ossidi, idruri e nitruri di titanio: si ha così una efficace azione di pompaggio nei confronti di tali gas.

Gli idruri di titanio devono trovarsi a temperatura ambiente o a temperatura inferiore per limitare a valori accettabili le pressioni di equilibrio per il processo di dissociazione. Negli impianti a vuoto il film di titanio è in grado di pompare i gas attivi per un tempo limitato (sino a quando la superficie dei film non è saturata) e quindi si rendono necessarie frequenti deposizioni e, se la portata dei gas è elevata, è richiesta una deposizione continua di titanio. La velocità di pompaggio cresce al diminuire della pressione. Al contrario, a pressioni sufficientemente basse ($P < 10^{-7}$ mbar) la velocità di pompaggio tende ad un valore limite indipendente dalla pressione.

In alcuni tipi di pompe ioniche il pompaggio dei gas nobili e dei gas meno reattivi (ad esempio il metano) è attuato mediante ionizzazione delle molecole con un fascio di elettroni. Gli ioni del gas nobile, prodotti per ionizzazione, sono indirizzati da intensi campi elettrici sul film di titanio dove vengono intrappolati e successivamente ricoperti dal deposito di nuovi strati di titanio. Il fascio di elettroni, prodotto da un

filamento caldo, ha un'efficienza di ionizzazione piuttosto limitata. Per questa ragione, come per i vacuometri a ionizzazione, si è preferito il procedimento di scarica di tipo Penning, cioè in presenza di un campo magnetico (Figura 11).

La pompa ionica in questo caso è costituita da un sistema di celle di Penning adiacenti immerse in un campo magnetico diretto assialmente d'intensità 0.1-0.2 T. Le differenze di potenziale tra la struttura anodica e quella catodica possono variare tra 2 e 10 kV. La scarica si autosostiene attraverso la ionizzazione multipla dei gas residui e attraverso l'estrazione di elettroni dal catodo freddo che viene bombardato dagli ioni prodotti nella scarica stessa. Gli ioni che incidono sul catodo con energie di parecchie decine di eV possono liberare, oltre agli elettroni, anche gli atomi superficiali del materiale (effetto di *sputtering*) e penetrare in profondità rimanendo intrappolati nel reticolo cristallino del catodo. Costruendo il catodo in titanio si ottiene, oltre all'assorbimento dei gas nobili ionizzati, la formazione continua di un film di titanio che provvede all'adsorbimento chimico dei gas reattivi.

Il processo di assorbimento dei gas nobili è in realtà più complesso in quanto, a causa dello sputtering che erode progressivamente il catodo, gli atomi imprigionati nel reticolo possono essere rimessi in circolazione. In pratica si realizzano forme degli elettrodi e del catodo tali da ottimizzare il processo di assorbimento degli ioni.

Anche le pompe ioniche hanno bisogno ovviamente di un sistema di pompe ausiliario che produca un vuoto primario compreso tra 10^{-1} e 10^{-3} mbar, così da consentire alle pompe ioniche di arrivare a vuoti dell'ordine di 10^{-9} mbar.

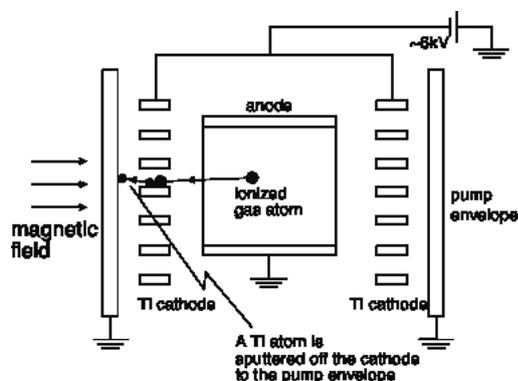
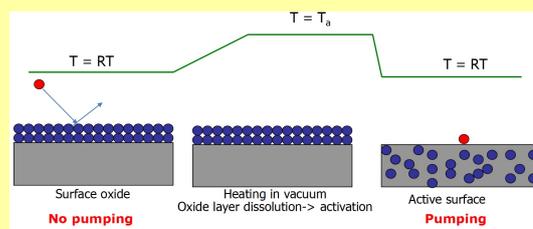


Figura 11: Schema di funzionamento di una pompa ionica.

Il getter non evaporativo (NEG)



Uno dei successi più spettacolari delle pompe a getter è rappresentato dal raggiungimento delle condizioni di vuoto estremo nella ciambella di 27 km di lunghezza della macchina acceleratrice di particelle LEP (acronimo di Large Electron Positron collider) sita presso il CERN di Ginevra. In questo caso sono state usate delle pompe getter denominate NEG (Non Evaporated Getter). Si tratta di strisce di ZrAl (zirconio-alluminio) che formano composti stabili con i prodotti desorbiti dalle pareti d'alluminio della lunga camera da vuoto di LEP. Quando la superficie del getter è saturata, può essere rigenerata scaldando il materiale così da consentire la diffusione del composto stabile all'interno della matrice solida del getter.

Le pompe criogeniche

Una parete fredda esplica un'azione di pompaggio nei confronti di quei vapori o gas che hanno un'elevata temperatura d'ebollizione e presentano quindi una piccola tensione di vapore alla temperatura della parete fredda. Questo è il principio sfruttato dalle pompe criogeniche. È importante osservare che con le pompe criogeniche la pressione ultima di un gas è sempre superiore alla tensione di vapore saturo dei gas alla temperatura della parete fredda.

L'acqua e l'anidride carbonica raggiungono valori inferiori a 10^{-9} mbar a 77 °K, mentre per il metano, l'argon, l'azoto e l'ossido di carbonio occorre arrivare sino a 20 °K per avere simili valori della tensione di vapore e per il neon a circa 7 °K. L'idrogeno e l'elio rappresentano il maggior problema. Infatti l'idrogeno solidifica a temperature inferiori a quella d'ebollizione dell'elio liquido e quindi il suo crio-assorbimento

viene ottenuto abbassando la temperatura della parete fredda sotto 4 °K o sfruttando particolari meccanismi quali l'intrappolamento criogenico (*cryotrapping*) e l'assorbimento criogenico (*cryosorption*). La solidificazione dell'elio non è osservabile regolando semplicemente la temperatura e quindi la percentuale d'elio negli impianti deve essere abbassata utilizzando altri metodi.

In realtà, per valutare la velocità effettiva di pompaggio nel caso di pompe operanti alla temperatura dell'elio liquido, bisogna tener conto della presenza di uno schermo termico. Quest'ultimo è normalmente raffreddato con azoto liquido ed ha la funzione di ridurre la potenza d'irraggiamento termico incidente sulla parete fredda e di ridurre quindi il consumo di elio liquido. Tale schermo riduce la velocità effettiva di pompaggio sia attraverso una diminuzione della conduttanza sia attraverso una riduzione della velocità media delle molecole e quindi del loro flusso.

La quantità di gas che può essere pompata da una parete fredda è limitata dal fatto che lo spessore di solido depositato sulla parete riduce progressivamente l'efficacia di pompaggio. Per comprendere come ciò si produca è sufficiente tener presente che la temperatura della superficie libera del solido si discosta sempre di più da quella della parete fredda al crescere dello spessore della sostanza condensata. Questo è dovuto al fatto che la potenza depositata sulla superficie libera del solido (per irraggiamento da parte delle pareti del recipiente o dello schermo ed energia termica delle molecole catturate) deve essere evacuata attraverso lo strato solido che possiede una conducibilità termica generalmente abbastanza piccola. Per questa ragione la pompa criogenica deve essere attivata soltanto dopo aver effettuato un pre-vuoto nel recipiente (tipicamente 10^{-3} mbar) con un'altra pompa e deve essere periodicamente rigenerata. La rigenerazione consiste nell'isolare la pompa dal recipiente da evacuare e nel portare la superficie fredda a temperatura ambiente in maniera che lo strato solido evapori ed i vapori che ne risultano vengano evacuati da una pompa meccanica.

La pompa criogenica può essere realizzata utilizzando come elemento refrigerante un bagno di liquido (N_2 , H_2 , He) oppure un gas freddo (in genere He) che viene fatto circolare in maniera continua. Lo schema di pompa a bagno che è

più semplice da realizzare e meno costosa ha una parete del recipiente più interno, riempito con elio liquido che serve come superficie di condensazione. Per limitare il consumo di elio liquido il recipiente è circondato da uno schermo raffreddato con azoto liquido che riduce, come si è detto precedentemente, la velocità di pompaggio ma agisce come pompa criogenica dei gas o vapori che hanno una elevata temperatura di ebollizione (ad esempio CO_2). I vantaggi offerti da una pompa criogenica sono rappresentati da elevate velocità di pompaggio (alcune migliaia di l/s) ottenute con dimensioni della pompa abbastanza contenute, dalla possibilità di sagomare la superficie di condensazione secondo le esigenze sperimentali, dall'assenza di vibrazioni (nel caso delle pompe criogeniche a bagno).

Gli svantaggi sono rappresentati dalla ridotta autonomia (necessità di rabboccare i liquidi criogenici evaporati e di rigenerare la superficie fredda in maniera periodica), dalla scarsa efficacia nel pompare l'idrogeno, dall'impossibilità di pompare l'elio. Nelle pompe a flusso continuo di gas, la superficie fredda è uno scambiatore di calore in cui evapora tipicamente dell'elio liquido pompato continuamente in esso. In anni recenti si sono diffuse attualmente criopompe in cui la parete fredda è realizzata sfruttando la potenza refrigerante prodotta da criogeneratori. I criogeneratori hanno bisogno di un robusto compressore di elio che raggiunge pressioni dell'ordine di un centinaio di atmosfere, connesso alla macchina termica tramite conduttanze flessibili di acciaio inox. Per talune applicazioni il limite principale di questa soluzione è rappresentato dalle vibrazioni associate al compressore.

Il vuoto nelle grandi macchine della ricerca

Tra i progetti di ricerca della *big science* ben tre hanno bisogno di creare un ambiente di ultra-alto-vuoto in elevati volumi (migliaia di metri cubi) e con controlli e monitoraggi sofisticatissimi.

Questi esperimenti colossali sono frutto di finanziamenti notevoli e distribuiti su lunghi archi di tempo, a cui lavorano gruppi numerosi e coordinati di scienziati e tecnici, dotati di grandi laboratori e apparecchiature d'avanguardia.

Nei paragrafi seguenti illustreremo le particolarità relative ai sistemi di vuoto di queste grandi macchine da ricerca.

Il più potente acceleratore di particelle: LHC

Il Large Hadron Collider (LHC) è l'acceleratore di particelle più potente al mondo situato presso il CERN di Ginevra [5]. Esso accelera due fasci di protoni fino ad un'energia equivalente a 7.5 mila miliardi di elettronvolt ed è utilizzato per le ricerche sperimentali nel campo della fisica delle particelle. È grazie a LHC che i due esperimenti, denominati ATLAS e CMS, hanno scoperto il bosone di Higgs tra le collisioni dei due fasci di protoni, la particella responsabile della massa di tutte le particelle conosciute.

LHC ha la particolarità di non avere uno solo, ma tre sistemi di vuoto:

1. vuoto per i due fasci di protoni,
2. vuoto di isolamento termico per i magneti superconduttori,
3. vuoto di isolamento termico per la linea di distribuzione dell'elio liquido.

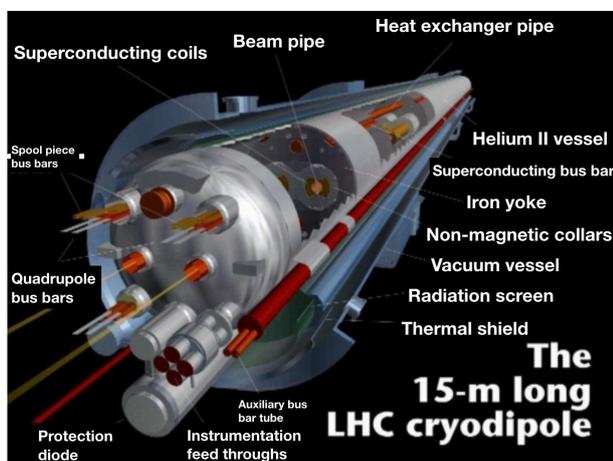


Figura 12: Disegno di uno dei 1010 dipoli superconduttori di LHC lunghi 15 m e con un campo magnetico nominale di 8 T.

LHC è in grado di mantenere in orbita i fasci di protoni a quella energia grazie a migliaia di dipoli e quadrupoli superconduttori (Fig. 12). Questi crio-magneti sono collocati in 16 sezioni d'arco lunghe 48 km e mantenuti a 1.9 K mediante un bagno di elio superfluido. Le sezioni

d'arco sono connesse tra loro con 6 km di sezioni dritte, mantenute a temperatura ambiente, dove si trovano i sistemi di controllo del fascio e le regioni di inserimento per gli esperimenti.

È necessario un ultra-alto-vuoto per i tubi in cui i fasci di particelle si muovono. Negli archi, l'ultra-alto-vuoto è mantenuto dal pompaggio criogenico. Poiché i tubi del fascio sono raffreddati a temperature estremamente basse, essi costituiscono una criopompa quasi perfetta ed i gas si condensano e aderiscono alle pareti molto facilmente. Appena due settimane di pompaggio sono necessarie per portare la pressione a valori inferiori a 10^{-10} mbar.

L'ultra-alto-vuoto nelle sezioni a temperatura ambiente è mantenuto grazie a due tecniche: una d'avanguardia ed una tipica dei sistemi di ultra-alto-vuoto. In primo luogo queste sezioni utilizzano diffusamente un *getter coating non evaporabile* o NEG - sviluppato e industrializzato al CERN - che assorbe le molecole residue quando viene riscaldato. Esse sono rivestite da un *coating* di lega di titanio-zirconio-vanadio depositata all'interno dei tubi del fascio. Funziona come un sistema di pompaggio distribuito, efficace per rimuovere tutti i gas, ad eccezione del metano e dei gas nobili che vengono invece rimossi da 780 pompe ioniche. In secondo luogo, le sezioni a temperatura ambiente consentono di fare il *baking* di tutti i componenti a 300 °C. Questa operazione deve essere eseguita a intervalli regolari per mantenere il vuoto alla pressione bassa desiderata.

Il volume più grande da pompare a LHC è il vuoto di isolamento termico necessario per mantenere le parti criogeniche a 1.9 K e ridurre la quantità di calore che proviene dall'ambiente circostante posto a temperatura ambiente. Il vuoto di isolamento ha una pressione di 10^{-6} mbar per un volume complessivo di circa 9000 m³ per i magneti superconduttori e di 5000 m³ per la linea di distribuzione d'elio liquido. In confronto, il volume del vuoto dei fasci di particelle accelerate è piccolo, infatti esso è costituito da due anelli con un raggio di 3 cm ed una circonferenza di circa 27 km per un volume totale di 150 m³.

In LHC il vuoto arriva a circa 10^{-10} mbar nel tubo del fascio a temperatura criogenica ed a 10^{-11} mbar vicino ai punti di interazione dove si verificano le collisioni che si vogliono studiare. Questi livelli di pressione sono circa cento volte

inferiori a quelli dell'atmosfera lunare; LHC è in effetti lo spazio più vuoto del sistema solare.

Con l'accensione dei fasci nel 2008, il Large Hadron Collider (LHC) è diventato il più grande sistema di vuoto operativo al mondo che utilizza una vasta gamma di tecnologie di vuoto. La costruzione di questo sistema di vuoto ha richiesto più di 250.000 giunti saldati e 18.000 guarnizioni a vuoto. I sistemi di vuoto di LHC sono dotati inoltre di 170 misuratori di ionizzazione Bayard-Alpert e di 1084 strumenti di tipo Pirani e Penning per monitorare la pressione del vuoto.

Gli interferometri per onde gravitazionali: LIGO e VIRGO

L'osservazione delle onde gravitazionali generate dalla fusione di coppie di buchi neri e di stelle di neutroni è sicuramente l'evento scientifico più importante di questi ultimi anni, che apre una nuova finestra d'osservazione sull'universo.

Questo è stato reso possibile grazie ai sofisticati miglioramenti approntati negli ultimi anni ai tre grandi interferometri di Michelson realizzati per rilevare le onde gravitazionali: il primo denominato Advanced Virgo, con bracci lunghi 3 km e situato nel comune di Cascina (Pi) e gli altri due identici denominati Advanced LIGO, con bracci lunghi 4 km situati negli USA [6].

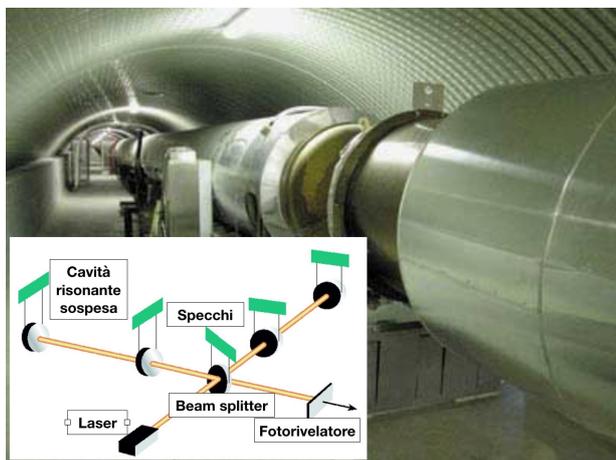


Figura 13: Una parte di uno dei tubi di acciaio di Advanced Virgo in cui si vedono gli anelli di supporto saldati al tubo a spirale per aumentare notevolmente l'integrità strutturale del cilindro. Nel riquadro è mostrato lo schema di interferometro di Michelson.

I componenti ottici di questi enormi interferometri sono collocati in tubi a vuoto dal diametro

superiore al metro e di lunghezza di alcuni chilometri, per un volume di circa una decina di migliaia di metri cubi, e tenuti a pressioni dell'ordine di 10^{-9} mbar (Fig. 13). L'area centrale, il cuore dell'interferometro è un immenso tavolo ottico sotto ultra-alto-vuoto. Ogni elemento ottico è sospeso ad un sistema di isolamento sismico contenuto in una torre da vuoto. Le torri contengono delle cavità Fabry-Perot e sono connesse tra loro da tubi a vuoto.

Creare e mantenere questo vuoto è assolutamente essenziale per il funzionamento di questi apparati. L'ambiente in condizioni di vuoto impedisce alle onde sonore di produrre vibrazioni sugli specchi, poiché il suono non può attraversare il vuoto. Le variazioni di temperatura dell'aria, se presente all'interno dei tubi del fascio, altererebbero la forma dell'ottica ad un livello tale da distruggere la qualità del fascio laser. Inoltre le correnti d'aria all'interno dei tubi potrebbero anche causare il riflesso del laser rendendo impossibile la rilevazione dell'onda gravitazionale. Entrambe queste fonti di rumore possono essere eliminate operando in vuoto.

Ci sono voluti 40 giorni di pompaggio costante per evacuare i colossali tubi fino alla loro pressione operativa. In questo tempo, le pompe turbomolecolari hanno rimosso la maggior parte dell'aria nei tubi mentre i tubi stessi sono stati riscaldati a 150-170 °C per rimuovere i gas residui delle pareti interne. Il mantenimento di questo vuoto richiede il funzionamento costante di pompe ioniche, che estraggono le molecole fuoriuscite dai tubi e dalle altre strutture all'interno dei sistemi di vuoto, e di criopompe a corrente liquida che rimuovono le molecole di acqua. Per raggiungere i bassissimi valori di pressione richiesti sono stati sviluppati processi metallurgici speciali per la fabbricazione dei componenti, come il desorbimento dell'idrogeno a 400 °C.

I tubi a vuoto con un diametro di 1.20 m sono stati costruiti in segmenti di circa 20 m, con acciaio inox 304L di soli 3 mm di spessore, e saldati a spirale continua. Con il suo contenuto di carbonio relativamente basso questo acciaio è resistente alla corrosione, soprattutto nelle zone saldate. La ruggine si è inevitabilmente prodotta all'interno dei tubi a vuoto essendo la loro fabbricazione avvenuta negli anni '90. Quindi prima della installazione, le superfici interne dei tubi sono state meticolosamente lucidate e pulite per rimuovere la ruggine, al fine di evitare

che granelli di ossido cadano sul fascio laser o migrino su superfici ottiche.

I tubi a vuoto sono situati all'interno di tunnel che devono rimanere molto stabili su lunghi periodi di tempo. I tunnel sono fatti di elementi prefabbricati di cemento di circa 20 m di lunghezza sostenuti da un migliaio di pilastri che raggiungono gli strati più stabili del terreno.

Prossimo reattore per la fusione nucleare calda: ITER

The International Thermonuclear Experimental Reactor (ITER) è un progetto internazionale che si propone di realizzare un reattore a fusione nucleare di tipo sperimentale in grado di produrre più energia di quanta il processo ne consumi per l'innesco ed il sostentamento della reazione di fusione [7]. Nello specifico, ITER è un reattore deuterio-trizio in cui il confinamento del plasma è ottenuto in un campo magnetico generato da un magnete superconduttore all'interno di una macchina a ciambella denominata Tokamak. ITER sarà il più grande Tokamak del mondo con un plasma di raggio 6.2 m ed un volume di 840 m³ (Fig. 14a).

La ciambella sottovuoto, ed il criostato per il magnete di ITER, sono tra i più grandi sistemi di vuoto mai costruiti con i rispettivi volumi di 1.400 m³ e di 8.500 m³ e pressioni di 10⁻⁸ mbar e < 10⁻⁶ mbar. Saranno necessarie tecniche sofisticate per il monitoraggio e la manutenzione di questi sistemi. Una volta in esercizio, non sarà più possibile accedere alla macchina.

Prima di iniziare la reazione di fusione è necessario eseguire il pompaggio a vuoto per eliminare tutte le sorgenti di molecole organiche che altrimenti sarebbero dissociate nel plasma caldo. Pompe meccaniche e pompe criogeniche potenti devono evacuare l'aria dalla ciambella e dal criostato fino a quando la pressione interna è scesa ai livelli richiesti. Considerando il volume di ITER, questa operazione richiederà 24-48 ore.

I principali sistemi di pompaggio sono le sei pompe di scarico della ciambella, le quattro criopompe per i sistemi di iniezione a fascio neutro, utilizzate nel riscaldamento del plasma, e le due criopompe per il criostato necessario per il raffreddamento dei magneti superconduttori. Le complesse criopompe sono state progettate per le esigenze specifiche di ITER. Sono tutte basate su criopaneli, raffreddati con elio supercriti-

co e ricoperti di carbone attivo come materiale assorbente. La ricerca e lo sviluppo hanno dimostrato che il carbone da gusci di cocco finemente macinato ha la giusta densità e porosità per imprigionare le particelle di elio in ITER.

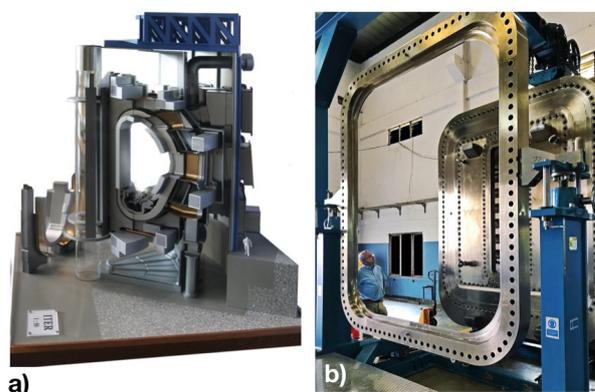


Figura 14: a) Modello artistico di una porzione di ITER. b) La più grande flangia da vuoto con guarnizione al mondo pronta per i test di accettazione dopo la costruzione, ideale per i molteplici accessi nel vuoto di ITER. Notare le dimensioni del toroide e della porta d'accesso paragonandole alle dimensioni degli uomini in figura.

Una cinquantina di porte, nella ciambella di ITER consentono l'accesso a operazioni di movimentazione remota, sistemi di diagnostica, riscaldamento e pompaggio del vuoto. Le porte d'accesso più grandi hanno superfici di 5 m², ma un'area non sigillata della larghezza di un capello umano è sufficiente a distruggere il vuoto e fermare le prestazioni necessaria per la fusione termonucleare.

È stata progettata e prodotta una replica di dimensioni maggiori delle più grandi porte equatoriali per testare la sigillatura a vuoto di quella che sarà la porta più grande mai costruita su un Tokamak: 4.8 × 3.3 × 4.8 m³ (Fig. 14b). Il grande test di tenuta è pronto in una fabbrica di Bangalore, in India. L'impianto di prova è stato appena completato ed è di dimensioni impressionanti (circa 5 metri di altezza e di lunghezza) e dal peso di circa 19 tonnellate. Dopo il test sarà spedito a ITER, dove verrà utilizzato per testare l'assemblaggio e la funzionalità di una serie di grandi guarnizioni a vuoto e convalidare le tecnologie di sigillatura proposte per il primo plasma e le successive fasi operative di ITER. La guarnizione di tenuta non solo consente di testare le più grandi guarnizioni

rettangolari smontabili, ma permette anche di preparare tecniche di installazione che saranno fondamentali per ottenere la qualità del vuoto richiesta. L'impianto può essere riscaldato a 240°C per simulare il baking del serbatoio di ITER ed ha anche la possibilità di applicare carichi alle flange per confermare la solidità della tenuta.

Conclusioni

Nel lavoro di ricerca non sempre c'è bisogno di conoscere i principi fondamentali del vuoto, tuttavia, spesso bisogna sapere come i diversi parametri dell'apparato sperimentale influenzeranno il vuoto.

La progettazione di un sistema di vuoto si basa essenzialmente nel determinare outgassing, pressione, conducibilità e velocità di pompaggio. In particolare, le sorgenti di gas devono essere sotto controllo in tutte le fasi del processo e/o dell'esperimento, al fine di ottenere la pressione necessaria.

La caratteristica dell'ultra-alto-vuoto è determinata in gran parte dai fenomeni superficiali, cioè dall'interfaccia tra vuoto e pareti. Le pompe per ultra-alto-vuoto agiscono mediante l'assorbimento fisico o chimico dei gas, la sepoltura di ioni nelle superfici o la condensazione su criopannelli. Le superfici delle pareti della camera devono essere accuratamente condizionate da un bakeout o da un altro trattamento per avere un desorbimento statico o stimolato del gas sufficientemente basso.

Le tecnologie del vuoto più avanzate sono state sviluppate per la ricerca fondamentale, ciò nonostante esse hanno trovato uso anche nel quotidiano: ad esempio la tecnologia del UHV ha migliorato notevolmente le prestazioni dei pannelli solari termici.

Ringrazio la Prof.ssa Stefania Spagnolo per l'accurata lettura del manoscritto e i suoi suggerimenti che mi hanno permesso di migliorare notevolmente la scrittura e la comprensibilità dell'articolo.



[1] CAS CERN Accelerator School : Vacuum Technology, Iñigo Aro, Spain, Ed.: Daniel A. Brandt CERN - 2007 - 003, 16 - 24 May 2006.

The US Particle Accelerator School, 2017, Vacuum Science and Technology for Accelerator Vacuum Systems Yulin Li and Xianghong Liu.

- [2] M. W. Zemansky, *Calore e Termodinamica*, Zanichelli (1970).
- [3] B. Ferrario, *Introduzione alla Tecnologia del Vuoto*, edizione riveduta da A. Calcatelli, Patron Editore - Bologna (1999).
- [4] Fulvio Ricci, *Dispense di Fisica e Tecnologia del Vuoto*, Dipartimento di Fisica, Università di Roma La Sapienza, Roma, Italia.
- [5] *The Large Hadron Collider. Conceptual Design*. CERN/AC-95-05 (LHC) . 20 October 1995.
- [6] <http://www.ligo.org>, <https://www.ego-gw.it> .
- [7] <https://www.iter.org> .



Gabriele Chiodini: Nato a Bergamo nel 1967. Si è laureato a Milano nel 1995 con una tesi sui rivelatori a pixel al Silicio per ATLAS a LHC e nel 1998 ha conseguito il PhD con uno studio sperimentale sulla turbolenza nei plasmi. Dal 1999 al 2002 è stato ricercatore al Fermi National Accelerator Laboratory di Chicago dove si è occupato dello studio dei quark pesanti. Dal 2003 è ricercatore presso la sezione INFN di Lecce e si occupa dello studio del Modello Standard e della ricerca di nuova Fisica nell'esperimento ATLAS a LHC. È ideatore e responsabile del bersaglio attivo di diamante con elettrodi grafitati dell'esperimento PADME ai Laboratori Nazionali di Frascati finalizzato principalmente alla ricerca di un ipotetico fotone oscuro.

L'effetto Casimir, una forza dal vuoto

Piergiorgio Antonini

Laboratori Nazionali di Legnaro - INFN

La forza di Casimir prende il suo nome dal fisico olandese Hendrik Casimir, che per primo ne ha prevista l'esistenza nel 1948. Si tratta di una forza che attrae due superfici di conduttori posti allo stesso potenziale. Non è una forza elettrostatica, non è attrazione gravitazionale. Se fra le due superfici non c'è nulla, la forza deve essere generata dal vuoto stesso. Vediamo in dettaglio di che cosa si tratta, e come la si può misurare.

Introduzione

Il fisico olandese Hendrik Casimir (Fig. 1) pubblica nel 1948 un articolo [1] nel quale prevede una forza di attrazione fra due superfici conduttrici poste a piccola distanza fra loro, dove per piccola si intende rispetto alla lunghezza d'onda della banda di assorbimento del materiale di cui è costituito il solido. Questa forza dipende solo dalla distanza fra le superfici, dall'area comune e da costanti fondamentali. L'articolo è un ampliamento di una idea già pubblicata da Casimir e Polder pochi mesi prima [2], nel quale studiando la stabilità di sistemi colloidali [3] introduce l'influenza del ritardo nell'energia di interazione fra due atomi neutri nelle forze di Van der Waals-London.



Figura 1: H. Casimir, from *Society for the Nationaal Archief*.

La forza di Casimir

L'articolo di Casimir parte dall'interazione fra un atomo (o una molecola) e una superficie conduttrice:

$$\delta E = -\frac{3\hbar c}{8\pi} \frac{\alpha}{R^4} \quad (1)$$

La forza di Casimir

È il vuoto, ovvero lo stato fondamentale dell'elettrodinamica classica (QED) a generare la forza di Casimir: in meccanica quantistica i livelli energetici dell'oscillatore armonico sono dati da

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right)$$

con $n = 0, 1, 2, \dots$ e \hbar la costante di Planck. Il primo livello,

$$E_0 = \frac{\hbar\omega}{2}$$

è quello da cui si fa derivare la forza di Casimir. Si tratta di una manifestazione macroscopica della meccanica quantistica.

dove α è la polarizzabilità statica dell'atomo (o molecola) e R la distanza fra particella e superficie, nell'ipotesi che questa distanza non sia piccola. Mentre nella derivazione originale [2] la teoria era stata sviluppata a partire da forze di Van der Waals-London correggendola con effetti di ritardo nell'energia di interazione fra i due atomi neutri, ora Casimir la deriva da modifiche del punto di energia zero del campo elettromagnetico. La forza può essere interpretata come pressione delle onde elettromagnetiche al punto zero. La derivazione è semplice, e considera la differenza fra i modi elettromagnetici possibili fra queste due configurazioni mostrate in Fig. 2, ovvero:

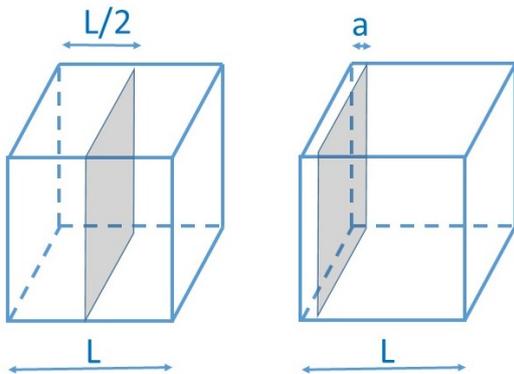


Figura 2: Geometria utilizzata da Casimir per la derivazione dell'espressione della forza.

I) un cubo a facce conduttrici di lato L nel quale sia posta una superficie conduttiva a distanza grande, ad esempio $L/2$ da una qualunque delle facce del cubo;

II) lo stesso cubo nel quale la superficie aggiuntiva sia a distanza piccola a .

In entrambi i casi $\frac{1}{2}(\sum_k \hbar\omega_k)$, dove la somma è fatta su tutte le possibili frequenze di risonanza della cavità diverge, ma non ne diverge la differenza

$$\frac{1}{2}\left(\sum_k \hbar\omega_k\right)_I - \frac{1}{2}\left(\sum_k \hbar\omega_k\right)_{II} \quad (2)$$

I modi possibili di una cavità definita dalla geometria

$$0 \leq x \leq L$$

$$0 \leq y \leq L$$

$$0 \leq z \leq a$$

hanno numeri d'onda

$$k_x = \frac{\pi}{L}n_x$$

$$k_y = \frac{\pi}{L}n_y$$

$$k_z = \frac{\pi}{L}n_z$$

dove n_x, n_y, n_z sono interi positivi e

$$k = \sqrt{(k_x^2 + k_y^2) + k_z^2}.$$

Con L grande la sommatoria può essere sostituita da un integrale

$$\frac{1}{2} \sum \hbar\omega = \hbar c \frac{L^2}{\pi^2} \int_0^\infty \int_0^\infty \left[\frac{1}{2} \sqrt{k_x^2 + k_y^2 + \frac{\pi^2}{a^2}} \right] dk_x dk_y, \quad (3)$$

e qui si vede come le costanti \hbar e c entrano nell'espressione della forza.

Con uno sviluppo in serie si ottiene la formula finale

$$F = \hbar c \frac{\pi^2}{240a^4}, \quad (4)$$

la cui conseguenza è una forza per unità di superficie che può essere interpretata come pressione del punto zero delle onde elettromagnetiche.

Le due superfici costituiscono una cavità, nella quale possono esistere solo modi che hanno lunghezza d'onda sottomultiplo della distanza fra le due superfici. L'energia di punto zero (zero point energy) all'interno del risonatore è minore dell'energia di punto zero dello spazio libero. La differenza di energia fra lo spazio vuoto, che possiamo pensare essere quello fra le superfici quando queste sono a distanza infinita, (dove quindi tutti i modi sono possibili) e il medesimo spazio in presenza di superfici a distanza a fra loro è:

$$U(a) = E(a) - E(\infty) = -\frac{\pi^2 \hbar c A}{720 a^3} \quad (5)$$

dove A è l'area delle superfici, e a la loro distanza. La differenza di pressione generata dallo sbilanciamento fra i modi interni e quelli esterni porta alla forza di Casimir, come è artisticamente raffigurato in Fig. 3.

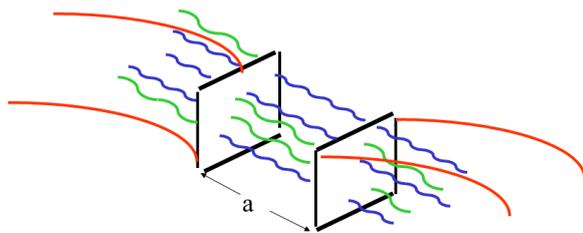


Figura 3: Un modo per immaginare la forza di Casimir è considerare lo sbilanciamento fra i modi esistenti all'esterno dello spazio delimitato dalle due superfici rispetto a quelli possibili inclusi fra le superfici. Quelli esterni sono di più, generando la forza.

$$F = -\frac{\pi^2 \hbar c A}{240 a^4} \quad (6)$$

Per fare un esempio, ad una distanza $d = 100\text{nm}$, la pressione è dell'ordine di 10 N/m^2 [4].

Derivazione della forza di Casimir

Grazie alla teoria di Lifshitz [8] si può spiegare la forza di Casimir a partire dalle forze di interazione fra due atomi neutri a distanza R . Per distanze piccole, rispetto alla lunghezza d'onda della transizione fra lo stato fondamentale e quello eccitato dell'atomo, ritrovando i risultati di Casimir e Polder [2].

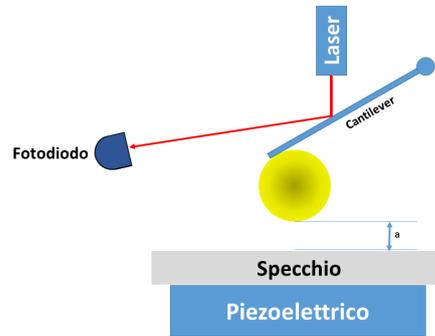


Figura 4: Schema della misura piano-sfera. Le superfici della sfera e dello specchio vengono ricoperte da uno strato metallico, solitamente alluminio oppure oro. La forza di Casimir piega il cantilever, e un fascio laser viene deflesso. Un fotodiodo (PD) sensibile alla posizione permette di ricavare l'angolo e quindi la forza.

La presenza di forze di attrazione fra atomi neutri porta ad ipotizzare simili forze fra corpi macroscopici a piccola distanza. Lifshitz affronta il problema dal punto di vista macroscopico. L'interazione fra gli oggetti avviene quindi attraverso il campo elettromagnetico presente all'interno del materiale, e che si estende anche al suo esterno sia in forma di onde che si propagano che di onde stazionarie. Questo campo, come sottolinea Lifshitz, non si annulla nemmeno alla temperatura nulla, venendo associato alle vibrazioni del campo di punto zero.

Nel caso di distanze grandi, alla temperatura $T = 0$, la forza dipende da $1/a^4$, dove a è la distanza, e dalla costante dielettrica dei materiali interessati. Nel caso di conduttori l'espressione si semplifica riottenendo esattamente l'espressione di Casimir (6), ed è ottenuta considerando i modi normali del campo compreso fra le due superfici.

La forza di Casimir per diverse geometrie

Piano Sfera

La prima misura che abbia confermato l'effetto Casimir è stata fatta con la geometria piano-sfera [9]. Questa geometria, sebbene sia diversa da quella inizialmente calcolata da Casimir, ha il vantaggio di essere sperimentalmente più semplice non presentando i problemi di allineamento delle due superfici piane. Lo schema di una misura piano-sfera si può vedere in Fig. 4.

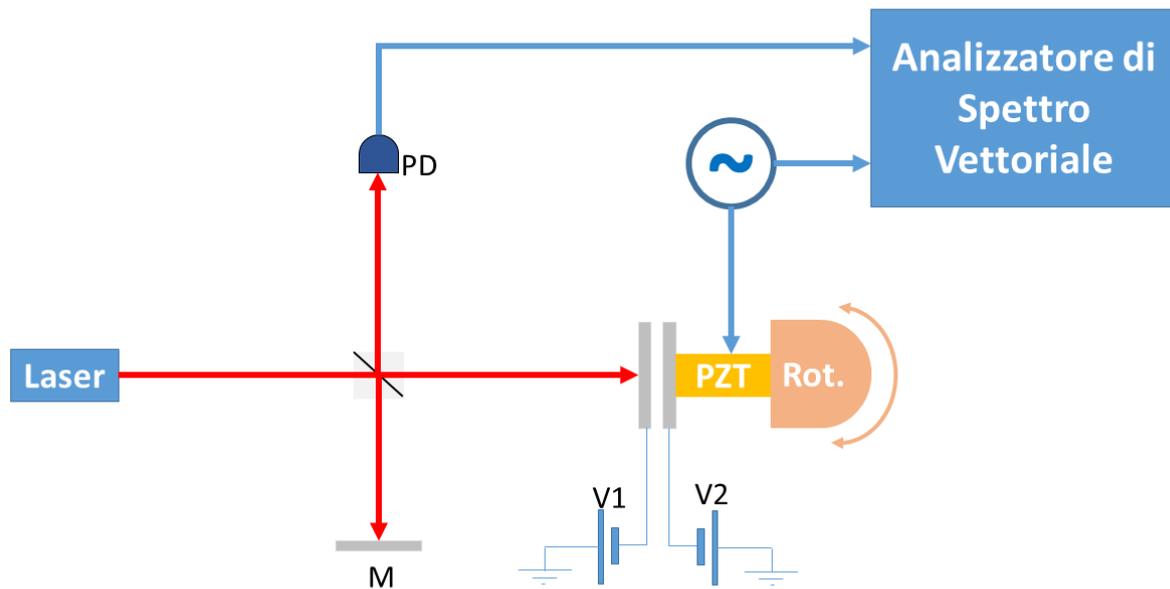


Figura 5: Set-up sperimentale per la misura della forza di Casimir fra superfici piane parallele. PD: fotodiodo per misurare le frange di interferenza, M: specchio fisso dell'interferometro, PZT: attuatore piezoelettrico per la variazione della distanza, Rot: attuatore per la rotazione della superficie mobile per ottenere l'allineamento. V1, V2: alimentatori di tensione per la soppressione della tensione parassita.

Un attuatore piezoelettrico sposta lo specchio verso la sfera. La forza di Casimir flette il cantilever. La flessione è proporzionale alla forza secondo la nota legge di Hooke. Un fotodiodo sensibile alla posizione misura la flessione di un fascio laser riflesso permettendo di risalire all'angolo e quindi alla forza.

Per questa geometria la forza di Casimir ha una dipendenza dal cubo della distanza, anziché dalla quarta potenza. La forma generale è quella dell'equazione seguente [10]:

$$F(a) = \frac{\pi^3}{360} R \frac{\hbar c}{a^3}, \quad (7)$$

dove R è il raggio della sfera. In generale la forma esatta della forza di Casimir per questa geometria viene calcolata tramite simulazioni numeriche della geometria considerata, e il risultato della misura in funzione della deviazione dalla forza prevista.

Superfici parallele

La prima misura della forza di Casimir nella configurazione originale di due superfici piane parallele è stata fatta utilizzando la tecnica di misura

omodina[11]. Un esempio di set-up sperimentale per la misura omodina è visibile in Fig. 5 [12], mentre la misura originale [11] è stata eseguita usando una fibra ottica.

Nella tecnica omodina la misura della forza non è fatta in condizioni statiche, ma dinamiche: anziché misurare la forza di attrazione fra le due superfici a diverse distanze usando ad esempio la deflessione di una bilancia di torsione, una delle due superfici viene fatta muovere mediante un attuatore piezometrico ad una frequenza nota. L'altra superficie costituisce lo specchio di un interferometro. L'interferenza fra il raggio laser che viene riflesso dalla superficie e il raggio di riferimento produce su un fotodiodo (PD in figura) una corrente elettrica, la cui variazione nel tempo δi dipende dal movimento dello specchio δl , oltre che dalla variazione della lunghezza d'onda del laser $\delta \lambda$. Se il laser è sufficientemente stabile questo ultimo contributo può essere trascurato dalla espressione:

$$\frac{\Delta i}{i_0} = 4\pi v \left(\frac{\delta l}{l} + l \frac{\delta \lambda}{\lambda} \right) \quad (8)$$

dove v è la visibilità della frangia di interferenza.

Muovendo uno dei due specchi tramite l'attuatore piezoelettrico, anche la superficie che fa da specchio per l'interferometro si muove, a causa della forza fra le due superfici. Questa superficie è mantenuta parallela all'altra tramite un telaio, la cui deformazione costituisce un oscillatore. Dato che fra le due superfici è presente una forza $F(a)$ che dipende dalla loro distanza, l'equazione del moto dello specchio è:

$$m\ddot{x}_r(t) = -m\omega_r^2 x_r(t) + F(a), \quad (9)$$

dove $x_r(t)$ è la posizione dello specchio, $\omega_r = 2\pi\nu_r$ la frequenza angolare propria, e m la massa. Riscrivendo la distanza a fra le due superfici come somma della parte stazionaria e di quella oscillante, $a = a_0 + x_s(t)$, per una forza della forma $F(a) = C/a^n$, si ha:

$$F(x_s, a_0) \simeq \frac{C}{a_0^n} \left[1 - n \frac{x_s}{a_0} + \frac{n(n+1)}{2} \left(\frac{x_s}{a_0} \right)^2 \right]. \quad (10)$$

Queste componenti della forza modificano l'ampiezza, la frequenza e la fase del moto libero dell'oscillatore.

Una modulazione periodica della posizione di una delle due superfici, $x_s = x_s^0 \cos \omega_s t$ fornisce, al primo ordine, una forza della forma:

$$F(t) = \frac{C}{a_0^n} \left(1 - n \frac{x_s^0}{a_0} \cos \omega_s t \right) \quad (11)$$

Dato che i due specchi si muovono alla stessa frequenza si ottiene che lo spettro di Fourier della soluzione dell'equazione del moto (8) ha un picco alla frequenza $\nu_s = \omega_s/2\pi$ la cui ampiezza è:

$$A_s^0 = n \frac{C x_s^0}{m \omega_r^2 a_0^{n+1}}. \quad (12)$$

Quindi in questo caso la forza ha una dipendenza dalla quinta potenza della distanza:

$$A_C = \frac{\pi^2 \hbar c S}{60 m \omega_r^2 a_0^5} x_s^0 \quad (13)$$

Fattori correttivi

La forma originale della forza di Casimir è calcolata supponendo che la misura sia eseguita fra due superfici che siano:

- perfettamente conduttive,

- perfettamente piane,
- perfettamente parallele
- alla temperatura di zero assoluto, $T = 0$ K.

La pratica sperimentale chiaramente non rispetta queste condizioni, e deviazioni dalle stesse ha conseguenze misurabili sulla forza di Casimir.

Effetto termico

Le formule della forza di Casimir fino a qui considerate sono state calcolate alla temperatura di zero assoluto, ovvero ignorando la presenza di fotoni termici. Per distanze piccole l'effetto termico è trascurabile, ma a distanze grandi a causa dello spettro di corpo nero questo effetto non può più essere trascurato.

Per

$$\lambda_T \geq \frac{2\pi c}{\omega} = \frac{\hbar c}{k_B T}, \quad (14)$$

dove k_B è la costante di Boltzmann e T la temperatura in gradi Kelvin, l'effetto termico diventa dominante. Circa $7 \mu\text{m}$ a temperatura ambiente. In particolare la forza in questo caso dipende da $1/a^3$ [8]:

$$F = \frac{T}{8\pi a^3} \left(\frac{\epsilon_0 - 1}{\epsilon_0 + 1} \right)^2, \quad (15)$$

e dipende solo dalla temperatura e dalla costante dielettrica ϵ_0 del materiale.

Rugosità e conducibilità finita

Prendiamo ad esempio la forza di Casimir nel caso piano-sfera, Eq. (7). L'espressione della forza nel caso ideale $F(a)$ deve essere corretta nel caso che la conducibilità del materiale non sia infinita. Seguendo [10], se ω_p è la frequenza di plasma, oltre la quale il metallo è trasparente, la correzione alla forza di Casimir è:

$$F_c^{(p)}(a) = F(a) \left[1 - \frac{4c}{a\omega_p} + \frac{72}{5} \left(\frac{c}{a\omega_p} \right)^2 \right]. \quad (16)$$

La rugosità del materiale modifica ulteriormente l'espressione della forza, aggiungendo il fattore correttivo seguente:

$$F^R(a) = F_c^{(p)}(a) \left[1 + 6 \left(\frac{A_r}{a} \right)^2 \right], \quad (17)$$

dove A_r è la rugosità media, considerata uguale per le due superfici, in questo caso.

Patch effect

Per patch effect si intende la forza di attrazione fra le due superfici causata dal fatto che sebbene esse siano mantenute allo stesso potenziale, questo sia vero in media su tutta la superficie. In realtà la superficie di un metallo non può essere tutta equipotenziale, essendo costituita da microcristalli di diverse funzioni lavoro (*work function*). L'effetto può essere calcolato, dopo aver misurato la distribuzione dei patch ad esempio con un *atomic force microscope*. Un tipico risultato è mostrato in Fig. 6.

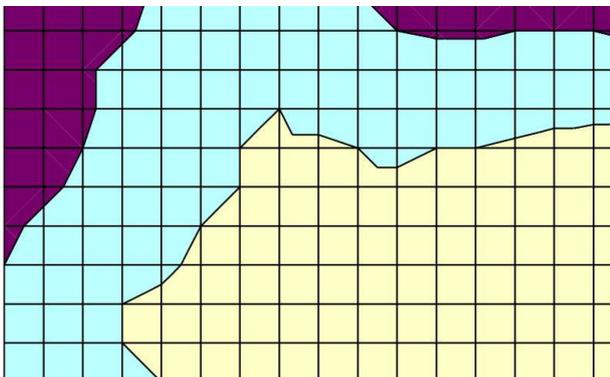
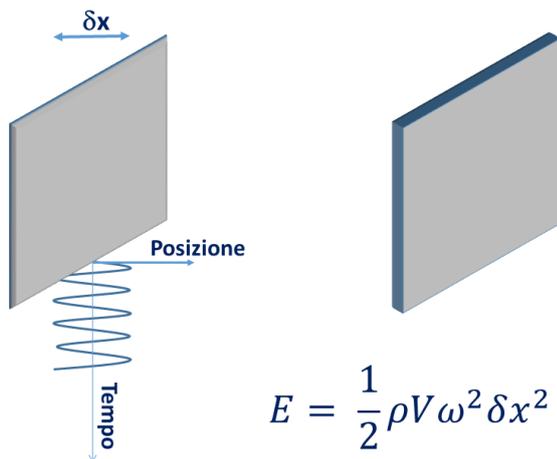


Figura 6: Grafico della *Work function* di un bulk in alluminio ottenuto con la *Kelvin probe*.



$$E = \frac{1}{2} \rho V \omega^2 \delta x^2$$

Figura 7: Nell'effetto *Casimir dinamico* lo spostamento periodico di uno dei due specchi risulta nella creazione di fotoni reali. L'effetto è proporzionale alla frequenza di oscillazione ω^2 , alla ampiezza dell'oscillazione δx^2 .

Effetto Casimir Dinamico

L'idea che uno specchio che si muova ad una velocità confrontabile a quella della luce possa convertire fotoni virtuali in fotoni reali effettivamente osservabili è alla base dell'effetto Casimir dinamico [5, 6]. Anche questa è una conseguenza del fatto che lo spazio vuoto non sia privo di struttura. Così come nell'effetto Casimir statico la forza deriva dalla differenza di pressione fra i modi possibili all'interno della cavità formata dalle due superfici, nell'effetto Casimir dinamico la struttura del vuoto varia nel tempo. Se il movimento è lento, lo stato fondamentale del campo elettromagnetico del vuoto si riadatta adiabaticamente, ma quando la velocità degli specchi è confrontabile con quella della luce il campo elettromagnetico non ha il tempo di adattarsi, e si ha la generazione dei fotoni reali [19]. Lo specchio deve dunque muoversi di un moto sinusoidale ad una frequenza molto alta, almeno di qualche GHz.

Il numero di fotoni atteso può venire calcolato, generalmente si distingue fra due situazioni:

1) Configurazione unidimensionale, dove uno specchio segue un moto armonico di frequenza Ω ed ampiezza a , per un tempo finito T . Il numero di fotoni generati tramite l'effetto Casimir dinamico risulta:

$$N_{ph} = \frac{\Omega T}{6\pi} \left(\frac{v}{c}\right)^2 \quad (18)$$

dove v è la velocità alla quale si muove lo specchio e c la velocità della luce.

2) In una cavità tridimensionale per una variazione $\epsilon = \frac{dL}{L_0}$ della lunghezza della cavità e $n = \frac{\Omega t}{2\pi}$ essendo il numero di cicli, il numero di fotoni generati è:

$$N_{ph} = \sinh^2(n\pi\epsilon). \quad (19)$$

Chiaramente non è possibile muovere fisicamente uno specchio a tali velocità. Varie tecniche sperimentali sono state proposte, ad esempio variando la riflettività di uno dei due specchi mediante illuminamento con laser pulsati [20]. L'utilizzo di questa tecnica in una cavità ottica risonante inoltre aumenta l'effetto [7]: in una cavità risonante il numero di fotoni raddoppia. Essendo Q il fattore di qualità della cavità il numero

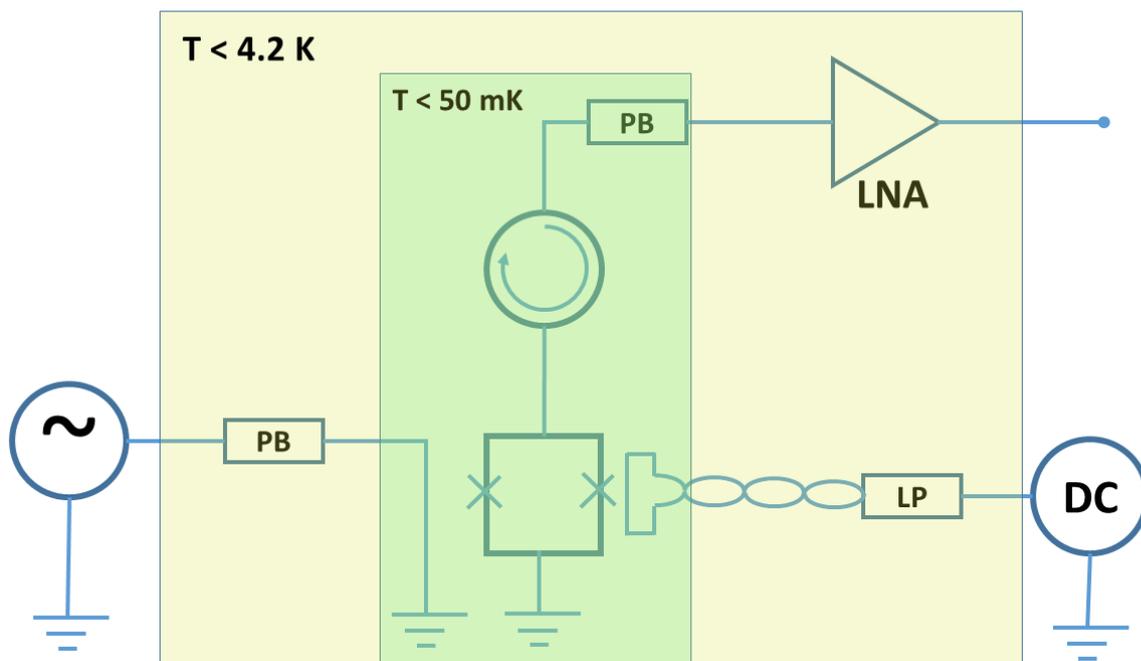


Figura 8: Lo SQUID, nella parte centrale in basso, è accoppiato ad un flusso DC (continuo) di campo magnetico detto di bias attraverso un filtro passa basso (a destra) e ad flusso variabile attraverso un filtro bassa-banda (da sinistra) il segnale che permette di rivelare l'effetto Casimir dinamico viene trasferito ad un amplificatore attraverso un circolatore. Tutto il sistema è a bassa temperatura: inferiore a 50 mK lo SQUID ed il circolatore, inferiore a 4.2 K l'amplificatore e i filtri (adattato da [19]).

di fotoni attesi per una cavità con $Q = 10^6 - 10^8$ è circa $N_{ph} \geq 10^2 - 10^3$.

Alla base di questa tecnica è l'idea di rendere trasparente la prima superficie di uno dei due specchi della cavità, in modo periodico. La cavità quindi cambia le proprie dimensioni alternando la posizione di uno dei due specchi, come si vede nelle Figure 7 e 9.

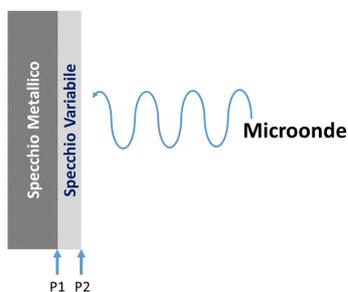


Figura 9: Rendendo periodicamente trasparente uno specchio usando illuminamento da un laser pulsato si sposta virtualmente la lunghezza della cavità, spostando la posizione dello specchio da quello che è reso trasparente ad un substrato posto al di sotto di esso.

Un'altra tecnica, che ha portato alla prima misura dell'effetto Casimir dinamico, consiste nell'utilizzo di SQUID (Superconducting Quantum Interference Device: dispositivo superconduttore a interferenza quantistica). Lo SQUID è un magnetometro. La lunghezza elettrica può essere modulata ad una frequenza superiore a 10 GHz, modulando l'induttanza dello SQUID, che può essere pensato come due giunzioni Josephson connesse in modo da formare un anello [19]. Ad alte frequenze lo SQUID si comporta come un'induttanza parametrica, cioè un'induttanza il cui valore L_J può essere variato applicando un flusso magnetico variabile Φ_{ext} attraverso l'anello. Lo SQUID viene quindi posizionato alla fine di una linea di trasmissione. La variazione di induttanza provocata dalla variazione di Φ_{ext} può essere pensata come ad una variazione della lunghezza (elettrica) del circuito, ottenendo un effetto simile alla variazione di posizione degli specchi. Il circuito utilizzato per questa misura si può vedere nella Figura 8. Lo SQUID è accoppiato ad un flusso DC (continuo) di campo magnetico detto di bias attraverso un filtro passa basso, mentre il flusso variabile Φ_{ext} viene

accoppiato allo SQUID attraverso un filtro passa-banda. Il segnale Casimir dinamico viene trasferito ad un amplificatore attraverso un circolatore. Lo SQUID richiede temperature criogeniche, che giovano comunque alla linea riducendone il rumore termico.

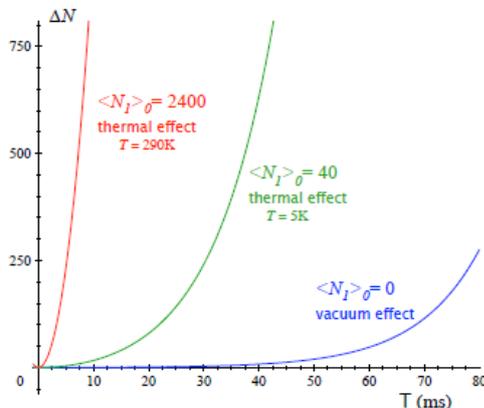


Figura 10: $\langle N_1 \rangle_0$ è il numero di fotoni termici calcolati alle diverse temperature. È nullo allo zero assoluto, ed aumenta con la temperatura [22].

È anche possibile fare una stima quantitativa dell'effetto atteso: se la variazione di induttanza avviene alla frequenza $\omega_d = 2\pi f$ allora i fotoni (reali) sono prodotti in coppie tali che le loro frequenze soddisfano alla relazione $\omega_+ + \omega_- = \omega_d$. Il flusso di fotoni atteso è lo stesso della configurazione a specchi mobili:

$$\Gamma = \left(\frac{\omega_d}{12\pi} \right) \left(\frac{v_e}{c_0} \right)^2, \quad (20)$$

dove $v_e = \delta l_e \omega_d$ e δl_e è la variazione di lunghezza elettrica del circuito. Sebbene questo sistema presenti diverse incertezze sperimentali, essa è considerata la prima misura dell'effetto Casimir dinamico.

Anche per l'effetto Casimir dinamico, come è intuibile, si deve considerare la presenza e l'influenza di fotoni reali dovuti alla temperatura non nulla alla quale viene condotto l'esperimento. La Fig. 10 ne mostra la simulazione numerica a diverse temperature: detto $\langle N_1 \rangle_0$ il numero di fotoni termici, il contributo a 0 K è nullo, mentre aumenta velocemente con la temperatura.

Conclusioni

L'effetto Casimir rappresenta una manifestazione macroscopica della meccanica quantistica, anche a questo si deve il fascino che esercitò sulla comunità dei fisici.

Ringraziamenti

L'autore ringrazia Caterina Braggio, Giovanni Carugno e Giuseppe Ruoso per lo scambio di informazioni che sono state utili nella stesura di questo articolo.



- [1] H. B. G. Casimir: *On the attraction between two perfectly conducting plates*, Proc. K. Ned. Akad. Wet., **51** (1948) 793.
- [2] H. B. G. Casimir and D. Polder: *The Influence of Retardation on the London-van der Waals Forces*, Phys. Rev. Lett., **73** (1948) 360.
- [3] E. J. W. Verwey et al: *Theory of the Stability of Lyophobic Colloids*. Elsevier, Amsterdam (1948).
- [4] D. Iannuzzi et al.: *Effect of hydrogen-switchable mirrors on the Casimir force*, PNAS, **101** (2004) 4019.
- [5] G. Moore: *Quantum theory of the electromagnetic field in a variable-length one-dymensional cavity*, J. Math. Phys., **11** (1970) 2679.
- [6] M. Kardar and R. Golestanian: *The "friction" of vacuum, and other fluctuation-induced forces*, Rev. Mod. Phys., **71** (1999) 1233.
- [7] V. V. Dodonov and A. B. Klimov: *Generation and detection of photons in a cavity with a resonantly oscillating boundary*, Phys. Rev. A, **53** (1996) 2664.
- [8] E. M. Lifshitz: *The Theory of Molecular Attractive Forces between Solids*, Sov. Phys. JETP, **2** (73) 1956.
- [9] P. H. G. M. Blokland and J. T. G. Overbeek: *van der Waals forces between objects covered with a chromium layer*, J. Chem. Soc. Faraday Trans., **I 74** (1978) 2637.
- [10] U. Mohideen and A. Roy: *Precision Measurement of the Casimir Force from 0.1 μm to 0.9 μm* , Phys. Rev. Lett., **81** (1998) 4549.
- [11] G. Bressi et al: *Measurement of the Casimir Force between Parallel Metallic Surfaces*, Phys. Rev. Lett., **88** (2002) 041804.
- [12] P. Antonini et al: *An experimental apparatus for measuring the Casimir effect at large distances*, J. Phys.: Conf. Ser., **161** (2009) 012006.
- [13] F. Capasso et al: *Quantum Mechanical Actuation of Microelectromechanical Systems by the Casimir Force*, Science, **291** (2001) 1941.
- [14] F. Chen et al: *New Features of the Thermal Casimir Force at Small Separations*, Phys. Rev. Lett., **90** (2003) 160404.

- [15] C. Genet et al: *Temperature dependence of the Casimir effect between metallic mirrors*, Phys. Rev. A, **62** (2000) 012110.
- [16] S. K. Lamoreaux: *Demonstration of the Casimir Force in the 0.6 to 6 μm Range*, Phys. Rev. Lett., **78** (1979) 5.
- [17] B. W. Harris et al: *Precision measurement of the Casimir force using gold surfaces*, Phys. Rev. A, **62** (2000) 052109.
- [18] H. B. Chan et al: *Quantum Mechanical Actuation of Microelectromechanical Systems by the Casimir Force*, Science, **291** (2001) 1971.
- [19] C. M. Wilson et al: *Observation of the dynamical Casimir effect in a superconducting circuit*, Nature, **479** (2011) 376.
- [20] C. Braggio et al: *Semiconductor microwave mirror for a measurement of the dynamical Casimir effect*, Rev. Sci. Instrum., **75** (2004) 4967.
- [21] C. Genet et al: *Temperature dependence of the Casimir effect between metallic mirrors*, Phys. Rev. A, **62** (2000) 012110.
- [22] C. Braggio: *Particle generation from the quantum vacuum: an experimental study to detect the Motion Induced Radiation in the laboratory*, Tesi Dottorato Università di Ferrara, (2007).



Piergiorgio Antonini: è un tecnologo dei Laboratori Nazionali di Legnaro (LNL) dell'INFN. Si è occupato di fisica nucleare, stabilizzazione di laser su cavità criogeniche, relatività ristretta, effetto Casimir ed ottica. Attualmente è fisico di macchina del ciclotrone recentemente installato presso i LNL.

Il vuoto come cristallo birifrangente: l'esperimento PVLAS.

Recent calculations of the changes in the absorption coefficient of hard gamma-rays due to the formation of electron-positron pairs have lent strong support to Dirac's picture of holes of negative energy. Still, the almost insurmountable difficulties which the infinite charge-density without field offers to our physical understanding make it desirable to seek further tests of the theory. Here purely radiation phenomena are of particular interest inasmuch as they might serve in an attempt to formulate observed effects as consequences of hitherto unknown properties of corrected electromagnetic equations. We are seeking, then, scattering properties of the "vacuum".

_____ O. Halpern, *Phys. Rev.* 44 855.2 (1933)

Guido Zavattini

Dipartimento di Fisica e Scienze della Terra - Università di Ferrara

Il concetto di vuoto ha subito un'evoluzione nel corso dei secoli partendo dall'idea di Aristotele che esso non potesse esistere, passando dal vuoto di Torricelli e di Maxwell fino al vuoto quantistico di oggi. Maxwell definiva il vuoto come una zona di spazio dalla quale fosse stato rimosso tutto il rimovibile. Mentre questa definizione potrebbe essere ancora valida, l'idea sottintesa che il vuoto sia un sistema statico, non lo è più. Al contrario, il vuoto è un mezzo dinamico per nulla in quiete: l'esistenza delle anti-particelle,

come previsto da Dirac, e del Principio di Indeterminazione di Heisenberg prevedono che quantisticamente il vuoto fluttua. Uno dei processi possibili è che il vuoto fluttui in coppie elettrone-positrone. In presenza di un campo esterno che polarizza queste coppie, il vuoto si comporta come un cristallo uniasico. L'esperimento PVLAS, installato a Ferrara presso la sezione INFN e il Dipartimento di Fisica e Scienze della Terra, sta cercando di osservare tale comportamento del vuoto.

Il primo a studiare empiricamente il vuoto fu Evangelista Torricelli. Egli spiegò il fatto che non si può *aspirare* l'acqua oltre circa 10 m di altezza intuendo che era la pressione atmosferica a definire l'altezza massima di una tale colonna d'acqua. Fu così che nacque il barometro di Torricelli: un tubo di vetro sigillato ad una estremità riempito di mercurio e tenuto con l'estremità aperta immersa in una bacinella piena di mercurio. L'altezza della colonna di mercurio nel tubo è una misura della pressione atmosferica. Poco tempo dopo Blaise Pascal, utilizzando un simile barometro, dimostrò che la pressione atmosferica diminuisce in funzione della quota sopra il livello del mare e che quindi l'atmosfera che circonda la terra è finita! Cosa c'è oltre l'atmosfera terrestre? Il *vuoto*¹.

Tre scoperte hanno rivoluzionato il concetto di vuoto nel XX-esimo secolo. La prima in ordine temporale risalente al 1905 è stata la famosa relazione di Einstein fra massa m ed energia \mathcal{E} : $\mathcal{E} = mc^2$. La seconda, nel 1927, fu il principio di indeterminazione di Heisenberg che mette in relazione la durata Δt della misura dell'energia di un sistema con la sua indeterminazione $\Delta \mathcal{E}$ tramite $\Delta \mathcal{E} \Delta t \geq \hbar/2$, dove \hbar è la costante di Planck ridotta. Infine ci fu la formulazione, un anno dopo, dell'equazione relativistica dell'elettrone (e in generale dei fermioni), che prevede stati di energia negativa, oggi noti come stati di antimateria. Questa previsione fu confermata nel 1932 con la scoperta del positrone, l'antielettrone. Con questi tre ingredienti è divenuto chiaro come il *vuoto* potesse fluttuare, per esempio, in coppie elettrone-positrone. Otto Halpern [1], nel 1933, fu il primo ad intuire che due fotoni potessero allora interagire, ovvero diffondere nel *vuoto* (vedi citazione sotto il titolo dell'articolo).

Oggi le moderne teorie quantistiche dei campi (l'Elettrodinamica quantistica, QED, la cromodinamica quantistica, QCD, etc.) definiscono il *vuoto* come lo stato di minima energia. Tale minimo di energia è un'energia media la quale, però, ha varianza non nulla dovuta proprio al Principio di Indeterminazione di Heisenberg. Diversi problemi nascono da questa visione del *vuoto*, in particolare la stima della sua densità di energia: questa risulta essere 10^{113} J/m^3 sulla base

di calcoli quantistici, in netto contrasto con il valore di 10^{-9} J/m^3 [2] stimato dall'espansione dell'Universo e dalla relatività generale².

La velocità della luce

A causa delle fluttuazioni del *vuoto* un fotone che si propaga liberamente in realtà si tramuta continuamente in coppie particella-antiparticella e altre combinazioni più complicate. Quest'idea è mostrata in Fig. 1, usando i moderni diagrammi di Feynman, con i processi mostrati in ordine di importanza. La linea ondulata spessa indica un fotone 'vestito', una linea ondulata sottile rappresenta un fotone 'nudo' mentre la linea chiusa indica un coppia fermione-antifermione (si ricordi che la propagazione dell'antifermione avviene nella direzione opposta a quella della freccia che lo indica).

La velocità della luce c è la velocità del fotone 'vestito': è il risultato delle interazioni di un fotone 'nudo' con le fluttuazioni del *vuoto*. Sperimentalmente la migliore misura di c è $c = 299792456.2 \pm 1.1 \text{ m/s}$ [3]. Vista la precisione con cui essa è stata misurata, dal 1983 c è stata *definita* essere $c \equiv 299792458 \text{ m/s}$ mentre il metro è diventato una grandezza derivata.

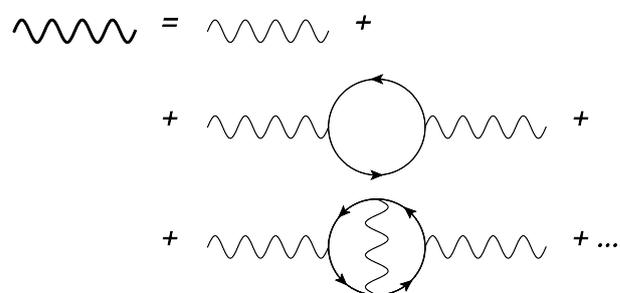


Figura 1: Rappresentazione di un fotone 'vestito' come l'interazione di un fotone 'nudo' con il vuoto che fluttua.

Sfruttando i tre ingredienti $\mathcal{E} = mc^2$, $\Delta \mathcal{E} \Delta t \geq \hbar/2$ e l'equazione relativistica dell'elettrone, nel 1935 due studenti di W. Heisenberg, H. Euler e B. Kockel [4], scrissero il primo lavoro in cui si calcolava l'interazione luce-luce al primo ordine nelle correzioni quantistiche nel limite di basse frequenze e campi piccoli rispetto ai campi elettrici E e magnetici B critici: $E_{\text{cr}} \ll m_e^2 c^3 / \hbar e =$

¹Una breve storia del vuoto è presentata nell'articolo di A. Calcatelli in questo numero di *Ithaca*.

²Si veda a tal proposito l'articolo di P. Ciafaloni in questo numero di *Ithaca*.

$1.3 \times 10^{18} \text{ V/m}$ e $B_{cr} \ll m_e^2 c^2 / \hbar e = 4.4 \times 10^9 \text{ T}$. Dovendo la Lagrangiana dell'elettromagnetismo essere invariante per trasformazioni relativistiche e per parità, la sua dipendenza dagli invarianti relativistici al quadrato $(E^2/c^2 - B^2)^2$ e $(\mathbf{E} \cdot \mathbf{B}/c)^2$ era l'unica possibile. Il calcolo di H. Euler e B. Kockel portò all'espressione della densità di Lagrangiana dell'elettromagnetismo in vuoto

$$\begin{aligned} \mathcal{L}^{(EK)} = & \frac{1}{2\mu_0} \left(\frac{E^2}{c^2} - B^2 \right) \\ + & \frac{A_e}{\mu_0} \left[\left(\frac{E^2}{c^2} - B^2 \right)^2 + 7 \left(\frac{\mathbf{E} \cdot \mathbf{B}}{c} \right)^2 \right] \\ + & \dots \end{aligned} \quad (1)$$

con il calcolo esplicito del parametro A_e

$$A_e = \frac{2}{45\mu_0} \frac{\alpha^2 \lambda_e^3}{m_e c^2} = 1.32 \times 10^{-24} \text{ T}^{-2}, \quad (2)$$

che definisce l'entità delle correzioni quantistiche, e dei valori 1 e 7, coefficienti che moltiplicano i suddetti invarianti relativistici al quadrato. La grandezza μ_0 è la permeabilità magnetica del vuoto. La quantità A_e risulta dipendere solo da costanti fondamentali quali la costante di struttura fine $\alpha = e^2/(4\pi\epsilon_0\hbar c)$, la lunghezza Compton ridotta dell'elettrone $\lambda_e = \hbar/(m_e c)$, la massa dell'elettrone m_e e la velocità della luce c in vuoto (in assenza di campi esterni). Essendo λ_e inversamente proporzionale alla massa dell'elettrone, A_e risulta essere inversamente proporzionale alla quarta potenza di m_e . Per questo motivo il contributo di fermioni carichi più pesanti, sconosciuti all'epoca di H. Euler e B. Kockel, è trascurabile.

Il primo termine della Lagrangiana (1) porta alle equazioni di Maxwell in vuoto (in assenza di correnti e/o cariche)

$$\begin{aligned} \nabla \cdot \mathbf{E} = 0; & \quad \nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \\ \nabla \cdot \mathbf{B} = 0; & \quad \nabla \times \mathbf{B} = \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} \end{aligned}$$

nelle quali sono implicitamente incorporate le equazioni delle onde elettromagnetiche. Il prodotto $\epsilon_0 \mu_0$ risulta essere legato alla velocità della luce c in vuoto:

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}. \quad (3)$$

Essendo le equazioni di Maxwell lineari nei campi \mathbf{E} e \mathbf{B} , vale il principio di sovrapposizione escludendo quindi l'interazione luce-luce.

Il secondo termine nell'equazione (1), quello fra parentesi quadrate e dipendente dalla quarta potenza nei campi, rappresenta il contributo al primo ordine delle correzioni quantistiche. Questo termine, rappresentato in Fig. 2, permette l'interazione luce-luce.

Nel 1936 H. Euler e W. Heisenberg [5] generalizzarono tale risultato includendo termini di ordine superiore e rendendolo valido a tutte le frequenze. Solo nel 1951, nell'ambito dell'Elettrodinamica Quantistica, J. Schwinger [6] dimostrò la correttezza dei risultati ricavati da H. Euler, B. Kockel e W. Heisenberg. Dal teorema ottico, l'interazione luce-luce Fig. 2, sinistra, e la velocità della luce in un campo elettrico o magnetico esterno, Fig. 2, destra, sono strettamente legati.

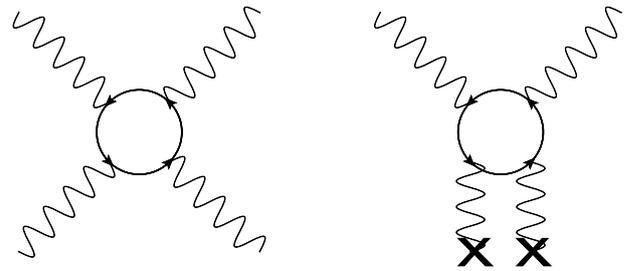


Figura 2: Sinistra: Rappresentazione dell'interazione luce-luce mediata da una coppia fermione-antifermione. Destra: Rappresentazione dell'interazione della luce con un campo classico esterno. Quest'ultimo diagramma porta ad una diminuzione della velocità della luce in un campo esterno.

Applicando un campo elettrico o magnetico esterno i contributi alla velocità della luce dovuti alle fluttuazioni del vuoto vengono modificati. L'effetto dominante è dovuto alla sostituzione del secondo termine di Fig. 1 con il diagramma di destra in Fig. 2. Di conseguenza la velocità della luce si riduce. Da un punto di vista microscopico il vuoto si polarizza esattamente come un mezzo materiale.

Partendo dalle equazioni costitutive

$$\mathbf{D} = \frac{\partial \mathcal{L}^{(EK)}}{\partial \mathbf{E}} \quad \text{e} \quad \mathbf{H} = - \frac{\partial \mathcal{L}^{(EK)}}{\partial \mathbf{B}} \quad (4)$$

si può determinare la relazione fra \mathbf{D} ed \mathbf{E} e quella fra \mathbf{B} ed \mathbf{H} . Considerando il primo termine della Lagrangiana (1), quello che porta alle equa-

zioni di Maxwell, si trova che $\mathbf{D} = \epsilon_0 \mathbf{E}$, $\mathbf{B} = \mu_0 \mathbf{H}$ e la velocità della luce risulta quella riportata nell'equazione (3).

Nel caso di un mezzo isotropo lineare ed omogeneo le relazioni fra \mathbf{D} ed \mathbf{E} e fra \mathbf{B} ed \mathbf{H} risultano lineari con $\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}$ e $\mathbf{B} = \mu_0 \mu_r \mathbf{H}$, dove ϵ_r e μ_r sono rispettivamente la costante dielettrica relativa e la permeabilità magnetica relativa. In questo caso la velocità della luce è

$$v = \frac{1}{\sqrt{\epsilon_0 \mu_0 \epsilon_r \mu_r}} = \frac{c}{n},$$

dove n ($n > 1$) è l'indice di rifrazione del mezzo. Chiaramente in *vuoto*, $n = 1$.

Vedremo nel prossimo paragrafo che considerando le correzioni quantistiche, la velocità della luce in *vuoto*, $v(\mathbf{B}_{\text{ext}})$ o $v(\mathbf{E}_{\text{ext}})$, dipende dalla presenza di un campo esterno magnetico \mathbf{B}_{ext} o elettrico \mathbf{E}_{ext} , e si riduce rispetto a c di una quantità proporzionale ad A_e :

$$n(\mathbf{B}_{\text{ext}}) - 1 = \frac{c}{v(\mathbf{B}_{\text{ext}})} - 1 \propto A_e B_{\text{ext}}^2$$

$$n(\mathbf{E}_{\text{ext}}) - 1 = \frac{c}{v(\mathbf{E}_{\text{ext}})} - 1 \propto A_e \frac{E_{\text{ext}}^2}{c^2}.$$

Poiché l'indice di rifrazione dipende dal campo esterno, il *vuoto* si comporta come un mezzo non lineare.

Notare che con campi realizzabili in laboratorio ($|\mathbf{E}_{\text{ext}}| \approx 100 \text{ MV/m}$, $|\mathbf{B}_{\text{ext}}| \approx 2 \text{ T}$) l'effetto di un campo magnetico è circa 40 volte maggiore di quello di un campo elettrico. Anche in presenza un campo magnetico piuttosto intenso di 2 T l'aumento di n risulta essere appena $\approx 5 \times 10^{-24}$.

Anisotropia del *vuoto*

Vediamo il caso più generale dell'effetto di un campo esterno sulla velocità della luce [7]: eseguendo le derivate (4) sulla Lagrangiana completa $\mathcal{L}^{(EK)}$, si trovano le seguenti espressioni per \mathbf{D} ed \mathbf{H} in funzione di \mathbf{E} e \mathbf{B} :

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \epsilon_0 A_e \left[4 \left(\frac{E^2}{c^2} - B^2 \right) \mathbf{E} + 14 (\mathbf{E} \cdot \mathbf{B}) \mathbf{B} \right] \quad (5)$$

$$\mu_0 \mathbf{H} = \mathbf{B} + A_e \left[4 \left(\frac{E^2}{c^2} - B^2 \right) \mathbf{B} - 14 \left(\frac{\mathbf{E}}{c} \cdot \mathbf{B} \right) \frac{\mathbf{E}}{c} \right].$$

Consideriamo il caso particolare di un fascio di luce che si propaga perpendicolarmente ad un campo magnetico esterno. Inoltre assumiamo che la polarizzazione della luce, definita dalla direzione del campo elettrico della luce \mathbf{E}_γ , sia lineare e che possa essere diretta parallelamente oppure perpendicolarmente al campo esterno. I campi \mathbf{E} e \mathbf{B} nelle equazioni (5) si riferiscono ai campi elettrici e magnetici *totali*. Nel caso in esame l'unico campo elettrico presente è quello della luce \mathbf{E}_γ mentre per il campo magnetico bisogna sostituire \mathbf{B} con $\mathbf{B}_\gamma + \mathbf{B}_{\text{ext}}$. Con alcuni passaggi, ricordando che per la luce $\left(\frac{E_\gamma^2}{c^2} - B_\gamma^2 \right) = 0$ e ponendo $|\mathbf{B}_\gamma| \ll |\mathbf{B}_{\text{ext}}|$, si ottengono le seguenti relazioni per la costante dielettrica relativa, la permeabilità magnetica relativa e l'indice di rifrazione:

- polarizzazione parallela (\parallel) al campo esterno

$$\begin{cases} \epsilon_{r,\parallel}^{(EK)} &= 1 + 10 A_e B_{\text{ext}}^2 \\ \mu_{r,\parallel}^{(EK)} &= 1 + 4 A_e B_{\text{ext}}^2 \\ n_{\parallel}^{(EK)} &= 1 + 7 A_e B_{\text{ext}}^2 \end{cases}$$

- polarizzazione perpendicolare (\perp) al campo esterno

$$\begin{cases} \epsilon_{r,\perp}^{(EK)} &= 1 - 4 A_e B_{\text{ext}}^2 \\ \mu_{r,\perp}^{(EK)} &= 1 + 12 A_e B_{\text{ext}}^2 \\ n_{\perp}^{(EK)} &= 1 + 4 A_e B_{\text{ext}}^2 \end{cases}$$

Come è già stato accennato precedentemente, l'indice di rifrazione del *vuoto* in presenza di un campo esterno è maggiore di 1 di una quantità proporzionale ad A_e . Inoltre risulta evidente che l'indice di rifrazione per luce polarizzata parallelamente al campo esterno è *diverso* da quello per luce polarizzata perpendicolarmente:

$$n_{\parallel} - n_{\perp} = \Delta n = 3 A_e B_{\text{ext}}^2. \quad (6)$$

Il *vuoto* in presenza di un campo esterno diventa birifrangente comportandosi come un cristallo uniassico.

Dato il piccolo valore di A_e , questa birifrangenza (e l'interazione luce-luce) a tutt'oggi non è stata ancora osservata direttamente. Recentemente c'è stata una prima osservazione indiretta della birifrangenza magnetica del *vuoto* [8] studiando la polarizzazione della luce emessa da una stella di neutroni. Inoltre l'esperimento ATLAS [9] al

CERN ha osservato l'interazione fotone-fotone ad alta energia.

Per completezza vogliamo aggiungere che l'indice di rifrazione può avere anche una parte immaginaria:

$$\hat{n} = n + i\kappa, \quad (7)$$

dove κ è il coefficiente di estinzione e descrive l'assorbimento della luce da parte del mezzo. Se κ è anisotropo il mezzo è detto essere dicroico. Il coefficiente di assorbimento lineare (in intensità) μ di un mezzo assorbente è legato a κ e alla lunghezza d'onda λ dalla relazione $\mu = 4\pi\frac{\kappa}{\lambda}$.

Ci si può allora chiedere se la Lagrangiana (1) prevede anche un dicroismo. I processi rappresentati in Fig. 3 contribuiscono al coefficiente di assorbimento, dal momento che l'energia del fotone primario si divide tra i due fotoni secondari, processo noto come 'photon splitting'. Dalla Lagrangiana (1) è stato dedotto [7, 10] che il caso più semplice mostrato in Fig. 3, sinistra, in cui un fotone interagisce con una sola linea di campo esterno, è proibito. Il processo dominante di 'photon splitting' è rappresentato da un diagramma a 6 vertici tipo quello mostrato in Fig. 3, destra, che da luogo ad un effetto, in pratica, non misurabile.

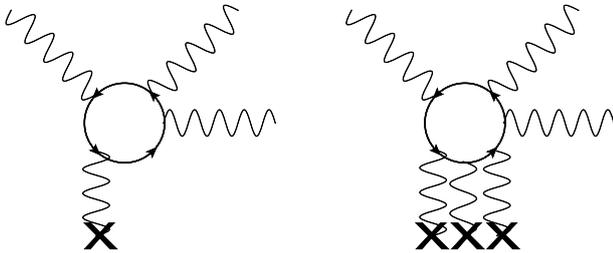


Figura 3: Sinistra: Photon splitting con quattro linee di campo, proibito nella QED. Destra: Diagramma dominante per il 'photon splitting' che porta ad un coefficiente di assorbimento estremamente piccolo.

Il risultato è che in presenza di un campo magnetico \mathbf{B}_{ext} i coefficienti di assorbimento lineari μ per diversi stati di polarizzazione dei fotoni iniziale e finali, nel limite delle basse energie, per i processi nelle equazioni (8) e (9), risultano [10]

$$\begin{aligned} & \mu [(\perp) \rightarrow (\parallel)' + (\parallel)''] = \\ & = \mu [(\parallel) \rightarrow (\perp)' + (\parallel)''] = \mu [(\parallel) \rightarrow (\parallel)' + (\perp)''] = \\ & = \frac{\alpha^3}{\lambda_e 60\pi^2} \left(\frac{B_{\text{ext}} \sin \vartheta}{B_{\text{cr}}} \right)^6 \left(\frac{\hbar\omega}{m_e c^2} \right)^5 \left(\frac{26}{315} \right)^2 = \\ & = 2 \times 10^{-84} \text{ m}^{-1} \end{aligned} \quad (8)$$

e

$$\begin{aligned} & \mu [(\perp) \rightarrow (\perp)' + (\perp)''] = \\ & = \frac{\alpha^3}{\lambda_e 60\pi^2} \left(\frac{B_{\text{ext}} \sin \vartheta}{B_{\text{cr}}} \right)^6 \left(\frac{\hbar\omega}{m_e c^2} \right)^5 \left(\frac{48}{315} \right)^2 = \\ & = 7 \times 10^{-84} \text{ m}^{-1}, \end{aligned} \quad (9)$$

mentre

$$\begin{aligned} & \mu [(\parallel) \rightarrow (\parallel)' + (\parallel)''] = \\ & = \mu [(\perp) \rightarrow (\parallel)' + (\perp)''] = \\ & = \mu [(\perp) \rightarrow (\perp)' + (\parallel)''] = 0. \end{aligned}$$

I valori numerici nelle equazioni (8) e (9) sono stati ottenuti considerando $\hbar\omega = 1.17 \text{ eV}$ ($\lambda = 1064 \text{ nm}$) e $B_{\text{ext}} = 2.5 \text{ T}$. Questi processi effettivamente descrivono un dicroismo, cioè un'attenuazione del fascio incidente che dipende dalla polarizzazione della luce incidente. Però i valori previsti sono talmente piccoli che corrispondono ad una lunghezza di assorbimento pari a $\sim 10^{57}$ volte le dimensioni dell'universo.

Concludiamo, quindi, che nell'elettrodinamica quantistica l'effetto dominante di un campo esterno è di rendere il vuoto birifrangente.

Misurare una birifrangenza

Come discusso nella sezione precedente, equazioni (2) e (6), il valore di birifrangenza che l'esperimento PVLAS (Polarizzazione del Vuoto con LASer) si propone di misurare è estremamente piccolo. Nel seguito cercheremo di discutere i vari aspetti dell'esperimento.

Per misurare una birifrangenza lineare indotta da un campo esterno (da distinguere da una birifrangenza circolare) bisogna far propagare un fascio di luce linearmente polarizzato attraverso il mezzo birifrangente. Se ϑ (con $\vartheta \neq 0$) è l'angolo fra la direzione della polarizzazione e la direzione del campo esterno, la luce diventa polarizzata ellitticamente. In altre parole il campo elettrico, anziché descrivere una retta (polarizzazione lineare), descrive un'ellisse. Questo concetto è mostrato in Fig. 4. Il rapporto fra l'asse minore a e l'asse maggiore b di tale ellisse è chiamata ellitticità (con $a \ll b$) ed è indicata con ψ . Come vedremo, in base al verso di percorrenza dell'ellisse, l'ellitticità può essere positiva oppure negativa. La relazione fra la birifrangenza Δn di un mezzo di lunghezza L e l'ellitticità

indotta è

$$\psi = \pm \frac{a}{b} = \pi \frac{\Delta n L}{\lambda} \sin 2\vartheta, \quad (10)$$

dove λ è la lunghezza d'onda della luce. Si noti che la birifrangenza è una proprietà del mezzo mentre l'ellitticità è l'effetto indotto dalla birifrangenza sulla luce.

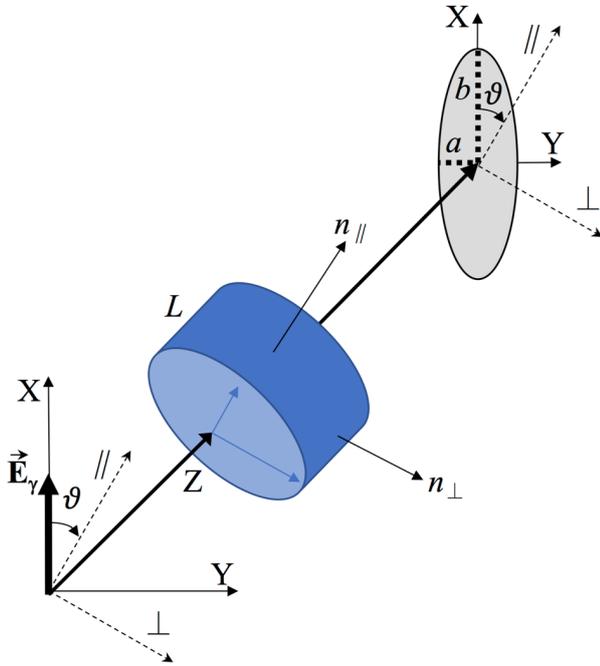


Figura 4: Il campo elettrico \mathbf{E}_γ di un fascio di luce polarizzato linearmente acquisisce una componente ortogonale a \mathbf{E}_γ sfasata di $\frac{\pi}{2}$. Il risultato è che la luce diventa polarizzata ellitticamente con l'ellitticità data da $\psi = \pm \frac{a}{b}$.

Ellitticità Vs. Rotazione

Per capire meglio l'origine dell'equazione (10) possiamo fare riferimento alla Fig. 4. Il campo elettrico \mathbf{E}_γ della luce incidente può essere proiettato lungo gli assi \parallel e \perp del sistema di riferimento definito dal campo esterno. Così si ottengono le due componenti del campo elettrico

$$\begin{aligned} E_{\parallel} &= E_\gamma \cos \vartheta \\ E_{\perp} &= -E_\gamma \sin \vartheta. \end{aligned}$$

Entrambe le componenti si propagano indipendentemente sperimentando indici di rifrazione diversi. Dopo un tratto di lunghezza L queste componenti saranno sfasate di una quantità ϕ

data da

$$\phi = 2\pi \frac{(n_{\parallel} - n_{\perp})L}{\lambda} = 2\pi \frac{\Delta n L}{\lambda}. \quad (11)$$

Le componenti del campo elettrico, a meno di una fase assoluta ininfluente, possono essere scritte come numeri complessi

$$\begin{aligned} E_{\parallel}(L) &= E_\gamma e^{i\phi/2} \cos \vartheta \\ E_{\perp}(L) &= -E_\gamma e^{-i\phi/2} \sin \vartheta. \end{aligned}$$

Nel sistema di riferimento (X,Y) del laboratorio le componenti $E_X(L)$ e $E_Y(L)$ risultano essere

$$\begin{aligned} E_X(L) &= E_\gamma \left(e^{i\phi/2} \cos^2 \vartheta + e^{-i\phi/2} \sin^2 \vartheta \right) \\ E_Y(L) &= E_\gamma \left(e^{i\phi/2} - e^{-i\phi/2} \right) \sin \vartheta \cos \vartheta. \end{aligned}$$

Assumendo che lo sfasamento sia $|\phi| \ll 1$ le componenti risultano, al prim'ordine in ϕ ,

$$\begin{aligned} E_X(L) &= E_\gamma \left(1 + i \frac{\phi}{2} \cos 2\vartheta \right) \\ E_Y(L) &= i E_\gamma \frac{\phi}{2} \sin 2\vartheta. \end{aligned}$$

Ricordando che $E_\gamma = E_0 e^{i\omega t}$, con $\omega = 2\pi c/\lambda$, queste sono le equazioni di un'ellisse nel tempo. La luce ha acquisito una componente $E_Y(L)$ del campo elettrico perpendicolare a \mathbf{E}_γ che, essendo moltiplicata per i , è sfasata di $\pi/2$ rispetto alla componente $E_X(L)$. Il rapporto delle due componenti risulta allora

$$\frac{E_Y(L)}{E_X(L)} = i \frac{\frac{\phi}{2} \sin 2\vartheta}{1 + i \frac{\phi}{2} \cos 2\vartheta} \approx i \frac{\phi}{2} \sin 2\vartheta. \quad (12)$$

Sostituendo l'equazione (11) nell'equazione (12) si ottiene l'equazione (10) dalla quale si vede che il valore massimo dell'ellitticità ψ (per $\sin 2\vartheta = 1$) è $\frac{\phi}{2}$. Se, al primo ordine, la componente $E_X(L)$ è reale (asse maggiore dell'ellisse), l'ellitticità è descritta da un numero immaginario puro.

Si noti che in base al segno di ϕ e all'angolo ϑ lo sfasamento della componente $E_Y(L)$ rispetto alla componente $E_X(L)$ può essere positivo oppure negativo: l'ellitticità ha quindi un segno.

Dal punto di vista sperimentale sono utili le seguenti osservazioni:

- essendo l'ellitticità legata direttamente al ritardo di fase acquisito lungo il tratto L fra

le componenti del campo elettrico secondo i due assi \parallel e \perp , è chiaro che tale ellitticità non dipende dal verso di propagazione lungo l'asse Z;

- essendo $\psi \propto L$, piccole ellitticità acquisite nell'attraversare mezzi diversi, si sommano algebricamente (tenendo conto anche del segno del prodotto $\Delta n \sin 2\vartheta$).

Questi fatti risultano importanti sperimentalmente per amplificare l'effetto di una birifrangenza e massimizzare l'ellitticità accumulata dalla luce.

Anche una rotazione di \mathbf{E}_γ genera una componente E_Y . In questo caso, però, le due componenti $E_X(L)$ e $E_Y(L)$ rimangono in fase. Questo può succedere quando un fascio di luce linearmente polarizzato attraversa un mezzo dicroico. Se $\Delta\kappa$ è il dicroismo del mezzo [vedi equazione (7)], riferito alle direzioni \parallel e \perp rispetto al campo esterno, la differenza di riduzione Δq delle componenti del campo elettrico è

$$\Delta q = q_{\parallel} - q_{\perp} = 2\pi \frac{\Delta\kappa L}{\lambda}. \quad (13)$$

Ripercorrendo gli stessi passaggi fatti per un mezzo birifrangente con la sostituzione $e^{i\phi/2} \rightarrow e^{-\Delta q/2}$, le due componenti $E_X(L)$ e $E_Y(L)$ diventano

$$\begin{aligned} E_X(L) &= E_\gamma (1 - \varphi \cos 2\vartheta) \\ E_Y(L) &= -E_\gamma \varphi \sin 2\vartheta, \end{aligned} \quad (14)$$

rimanendo in fase. Se la componente $E_X(L)$ è reale, lo è pure $E_Y(L)$. La rotazione φ risulta legata a Δq tramite $\varphi = \frac{\Delta q}{2}$. Anche per un mezzo dicroico che genera una rotazione della polarizzazione valgono le stesse considerazioni pratiche, elencate sopra, quando si misura un'ellitticità.

In conclusione, descrivendo il campo elettrico con numeri complessi, a un'ellitticità corrisponde una componente trasversa immaginaria mentre a una rotazione una componente reale. Vedremo in seguito che queste proprietà ci permetteranno di distinguere fra queste due grandezze durante una misura.

Il polarimetro PVLAS

Osservando le equazioni (6) e (10) si vede che per massimizzare l'ellitticità indotta su un fascio inizialmente polarizzato linearmente bisogna massimizzare la lunghezza della zona di campo magnetico L e l'intensità del campo magnetico B_{ext} , e minimizzare λ . Numericamente, per fissare le idee, con $B_{\text{ext}} \sim 2.5$ T, $L \sim 1$ m, $\lambda = 1064$ nm e $\sin 2\vartheta = 1$, l'ellitticità indotta per un passaggio in vuoto attraverso il campo magnetico è dell'ordine di

$$\psi \sim \pi \frac{3A_e B_{\text{ext}}^2 L}{\lambda} \sim 3 \times 10^{-17}. \quad (15)$$

Uno schema dell'ellissometro dell'esperimento PVLAS [11] è riportato in Fig. 5. Un polarizzatore all'ingresso del polarimetro polarizza la luce linearmente lungo la direzione X. Successivamente il fascio entra in una cavità Fabry-Perot che, come vedremo, permette di amplificare la lunghezza effettiva della zona soggetta al campo magnetico. L'effetto dovuto al campo magnetico è indicato con ξ e tiene conto sia della birifrangenza che del dicroismo: $\xi = i\phi - \Delta q = 2i\psi - 2\varphi$. Un modulatore di ellitticità poi permette di portare tutti i segnali fisici rilevanti lontani dalle frequenze basse dove varie sorgenti di rumore sono dominanti. Infine un analizzatore (polarizzatore posto a 90° rispetto al polarizzatore d'ingresso) seleziona la componente della luce polarizzata lungo la direzione Y. Le funzioni dei vari elementi sono discusse di seguito.

Il principio ottico

Lo schema ottico dell'esperimento, mostrato in Fig. 5, si basa essenzialmente sull'idea originale di E. Iacopini e E. Zavattini proposta nel 1979 [12].

Cominciamo con un solo polarizzatore, una zona birifrangente, ma non dicroica, e un analizzatore tralasciando quindi il Fabry-Perot e il modulatore. L'analizzatore ha due uscite che permettono di misurare sia il fascio estinto I_Y sia quello trasmesso I_X che sostanzialmente rappresenta l'intensità di luce incidente sull'analizzatore I_0 .

Dopo il polarizzatore d'ingresso la luce entra nella zona birifrangente dove acquisisce una certa ellitticità ψ . Per determinare ψ si può misurare l'intensità trasmessa $I_Y \propto |E_Y|^2$ dopo

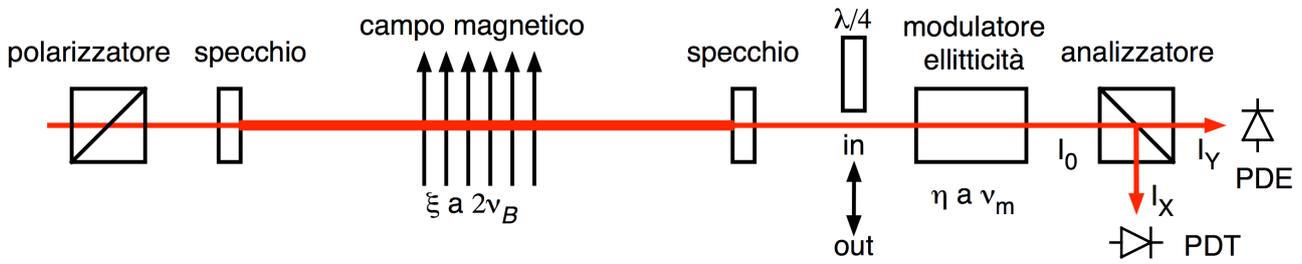


Figura 5: Schema di principio del polarimetro di PVLAS. Il parametro $\xi = 2i\psi - 2\varphi$ tiene conto sia dell'ellitticità indotta, sia della rotazione dovuta ad un dicroismo.

l'analizzatore:

$$I_Y = I_0 \psi^2.$$

L'intensità all'uscita risulta quadratica in ψ e quindi difficile da osservare visto che nel nostro caso abbiamo $\psi \ll 1$.

Per ovviare a ciò si può inserire un elemento con ellitticità nota η prima dell'analizzatore. L'intensità I_Y risulterà

$$I_Y = I_0 |i\psi + i\eta|^2 = I_0 [\psi^2 + 2\psi\eta + \eta^2],$$

lineare in ψ , con un notevole guadagno sulla capacità di misurare piccoli valori di ψ .

Esiste un problema sperimentale: ogni elemento ottico ha una sua piccola birifrangenza intrinseca che genera a sua volta una piccola ellitticità statica (più precisamente, quasi statica) β . Il problema è che non è possibile distinguere ψ da β . La soluzione a questo inconveniente è di rendere dipendente dal tempo il segnale di ellitticità: $\psi = \psi(t)$. Tipicamente $\psi(t)$ viene resa sinusoidale facendo, per esempio, ruotare la direzione del campo magnetico quindi rendendo dipendente dal tempo l'angolo ϑ nell'espressione (10): $\vartheta = \vartheta(t)$. Un'altra possibilità per modulare l'ellitticità indotta è modulare l'intensità del campo magnetico. L'intensità in uscita, considerando anche l'ellitticità spuria β , sarà a sua volta dipendente dal tempo: $I_Y = I_Y(t)$:

$$\begin{aligned} I_Y(t) &= I_0 |i\psi(t) + i\beta + i\eta|^2 = \\ &= I_0 [\eta^2 + 2\eta\psi(t) + 2\eta\beta + \dots]. \end{aligned}$$

Essendo $\psi \ll 1$ e $\beta \ll 1$ sono stati trascurati termini quadratici in queste grandezze. La dipendenza dal tempo dell'intensità $I_Y(t)$ permette di distinguere $\psi(t)$ da β .

Un altro potenziale problema riguarda la gene-

razione di piccole rotazioni $\varphi(t)$ che potrebbero anche dipendere dal tempo (per esempio effetti Faraday nei vari elementi ottici dovuti al campo magnetico della terra, campi di fuga dei magneti variabili nel tempo etc.). Come abbiamo discusso precedentemente tali rotazioni sono descritte da numeri reali mentre le ellitticità sono descritte da numeri immaginari. Sommando $\varphi(t)$ alle varie ellitticità si ottiene

$$\begin{aligned} I_Y(t) &= I_0 |\varphi(t) + i\psi(t) + i\beta + i\eta|^2 = \\ &= I_0 [\varphi(t)^2 + \psi(t)^2 + 2\psi(t)\eta + 2\beta\eta + \eta^2 + \dots]. \end{aligned}$$

Vista la piccolezza del valore di $\psi(t)$ che vorremmo misurare, il termine $\varphi(t)^2$ potrebbe essere dominante rispetto a $2\psi(t)\eta$. Inoltre se entrambi derivano dal campo magnetico variabile, potrebbero aver componenti di Fourier alla stessa frequenza rendendo la loro distinzione difficile. Si noti, però, che l'ellitticità η non moltiplica la rotazione $\varphi(t)$. Quindi rendendo variabile nel tempo anche η possiamo distinguere questi due termini in base alla loro frequenza:

$$\begin{aligned} I_Y(t) &= I_0 |\varphi(t) + i\psi(t) + i\beta + i\eta(t)|^2 = \\ &= I_0 [\varphi(t)^2 + \eta(t)^2 + 2\eta(t)\psi(t) + 2\eta(t)\beta + \dots]. \end{aligned}$$

Nello schema di Fig. 5 prima del modulatore di ellitticità può essere inserita una lamina a quarto d'onda. Se questa è allineata in modo da aver uno dei suoi assi parallelo alla direzione X, trasforma rotazioni in ellitticità e *vice-versa*. Con la lamina inserita l'apparato può eseguire misure di rotazione anziché di ellitticità. In questo caso l'espressione per $I_Y(t)$ risulta

$$\begin{aligned} I_Y(t) &= I_0 |\pm i\varphi(t) \mp \psi(t) \mp \beta + i\eta(t)|^2 = \\ &= I_0 [\varphi(t)^2 + \eta(t)^2 \pm 2\eta(t)\varphi(t) + \dots]. \end{aligned}$$

I segni \pm dipendono se la polarizzazione della

luce è allineata con l'asse veloce oppure l'asse lento della lamina. Come si vede, ora l'ellitticità del modulatore non moltiplica la $\psi(t)$ ma solo la $\varphi(t)$.

Ultima considerazione: due polarizzatori posti a 90° hanno sempre una piccola quantità di luce trasmessa per via di imperfezioni nei cristalli di cui sono fatti. Questa caratteristica di una coppia polarizzatore-analizzatore è detta estinzione σ^2 . Valori dell'ordine di $\sigma^2 \lesssim 10^{-8}$ possono essere ottenuti.

L'espressione finale per l'intensità trasmessa dal polarimetro nel caso di misure di ellitticità (ancora senza il Fabry-Perot) risulta allora

$$\begin{aligned} I_Y(t) &= I_0 \left[\sigma^2 + |i\psi(t) + i\beta + i\eta(t)|^2 \right] = \\ &= I_0 \left[\sigma^2 + \eta(t)^2 + 2\eta(t)\psi(t) + 2\eta(t)\beta + \dots \right]. \end{aligned} \quad (16)$$

Amplificazione del cammino ottico

Ci sono due tecniche principali per allungare il cammino ottico effettivo all'interno del campo magnetico: una cavità multi-pass e un interferometro di Fabry-Perot. Entrambe usano due specchi fra i quali è presente il campo magnetico.

Nel caso della cavità multi-pass il fascio entra tramite un foro in uno degli specchi e esegue un certo numero di riflessioni prima di uscire dallo stesso foro d'ingresso. Il fascio in questo modo colpisce gli specchi in punti diversi generando delle figure di Lissajous. Dato che ad ogni riflessione il fascio deve rimanere distinguibile dalla riflessione adiacente, il numero di riflessioni massimo è limitato dalle dimensioni del fascio laser di raggio r_0 e dal diametro degli specchi S . Valori tipici sono $r_0 \approx 1$ mm e $S \approx 10$ cm. L'ordine di grandezza del numero di riflessioni massimo risulta allora di circa $N \sim (S/r_0)^2 \sim 10^3$. Un grosso inconveniente di una tale cavità è la sua dimensione trasversa dell'ordine delle dimensioni degli specchi. L'inconveniente sta nel fatto che si possono ottenere campi magnetici dipolari più intensi su diametri piccoli. Questa tecnica fu usata nella prima proposta sperimentale per la misura della birifrangenza magnetica del vuoto ma fu poi abbandonata.

La seconda tecnica, oggi maggiormente utilizzata per allungare un cammino ottico, si basa sull'interferometro Fabry-Perot. Esso è compo-

sto da due specchi, posti ad una distanza D , con riflettività \mathcal{R} quasi uguale a 1 e trasmittività \mathcal{T} tale che $\mathcal{R} + \mathcal{T} = 1$. In pratica ci sono sempre delle perdite p nelle riflessioni ma finché queste sono tali che $p \lesssim \mathcal{T}$ le principali caratteristiche del Fabry-Perot sono dominate da \mathcal{R} e \mathcal{T} . Nel seguito trascureremo tali perdite.

Un fascio di luce di lunghezza d'onda λ che incide su una cavità Fabry-Perot è in risonanza se la distanza fra gli specchi è un numero semi-intero di lunghezze d'onda della luce. Infatti all'inizio del processo di carica della cavità lo specchio d'ingresso fa passare una piccola quantità di luce pari ad una frazione dell'intensità incidente \mathcal{T} . Essa viene riflessa quasi completamente dal secondo specchio e, quando ritorna al primo specchio e viene riflessa una seconda volta, si somma ad un'altra piccola quantità di luce entrante. Se la luce ha una lunghezza d'onda tale che il percorso d'andata e di ritorno corrisponde ad un numero intero di lunghezze d'onda questi due fasci sono in fase e subiscono un'interferenza costruttiva. Il fascio risultante si somma di nuovo con altra luce in ingresso dopo aver percorso la cavità in andata e ritorno. Se tutti questi fasci sono in fase c'è un accumulo di intensità finché la perdita per trasmissione \mathcal{T} attraverso il secondo specchio non uguaglia la luce in ingresso.

Per capire meglio l'amplificazione del cammino ottico dovuto ad una cavità Fabry-Perot possiamo scrivere il campo elettrico all'uscita della cavità a regime:

$$E_{\text{out}} = E_{\text{in}} \frac{\mathcal{T}}{1 - \mathcal{R}e^{i\delta}} \quad (17)$$

dove la fase $\delta = \frac{4\pi D}{\lambda} = \frac{4\pi D\nu}{c}$ rappresenta la fase totale accumulata dall'onda (in vuoto) in un'andata e ritorno fra gli specchi. La condizione di di risonanza si ottiene ponendo $\delta = 2m\pi$ dove m è l'ordine d'interferenza. In questo caso $E_{\text{out}} = E_{\text{in}}$ e l'intensità all'uscita è uguale a quella all'ingresso.

La separazione in frequenza fra due risonanze successive corrisponde al passaggio da un generico valore di m a $m+1$ ed è chiamata *free spectral range*. Essa è definita ponendo $\delta = 2\pi$:

$$\nu_{fsr} = \frac{c}{2D}.$$

Più alta è la riflettività degli specchi, più nu-

merosi sono i fasci che interferiscono e quindi più critica è la condizione di interferenza costruttiva. Possiamo scrivere la larghezza della risonanza in funzione di \mathcal{R} . L'intensità trasmessa $I_{\text{out}} \propto |E_{\text{out}}|^2$ ed è funzione della frequenza della luce incidente. La funzione che descrive l'intensità trasmessa si ricava dalla (17) ed è data dalla funzione di Airy:

$$I_{\text{out}}(\nu) = I_{\text{in}} \frac{1}{1 + \left(\frac{2\mathcal{F}}{\pi}\right)^2 \sin^2\left(\pi \frac{\nu}{\nu_{f_{sr}}}\right)},$$

dove I_{in} è l'intensità della luce incidente e \mathcal{F} è la *finesse* della cavità

$$\mathcal{F} = \frac{\pi\sqrt{\mathcal{R}}}{1 - \mathcal{R}}.$$

La larghezza totale della risonanza a metà altezza $\Delta\nu_c$ risulta direttamente legata alla finesse:

$$\frac{\nu_{f_{sr}}}{\Delta\nu_c} = \mathcal{F}.$$

Si noti che ad alte \mathcal{F} corrispondono risonanze strette.

Cerchiamo di capire come il Fabry Perot moltiplica l'effettivo cammino ottico fra i due specchi. Supponiamo di essere nella condizione di risonanza e di cambiare di una quantità 2ϵ il valore di δ : $\delta = 2m\pi + 2\epsilon$. Tale variazione di δ potrebbe essere dovuta, per esempio, ad un allungamento della cavità di una piccola quantità ΔD . Lo sfasamento subito dalla luce in assenza della cavità in una sola andata sarebbe

$$\epsilon = \frac{2\pi\Delta D}{\lambda}.$$

Se guardiamo lo sfasamento della luce in uscita dalla cavità rispetto a quella in ingresso troviamo

$$\begin{aligned} \tan \phi &= \frac{\Im[E_{\text{out}}]}{\Re[E_{\text{out}}]} = \\ &= \frac{\mathcal{R} \sin \delta}{1 - \mathcal{R} \cos \delta} \approx \frac{2\epsilon}{1 - \mathcal{R}} = \frac{2\mathcal{F}}{\pi} \epsilon. \end{aligned}$$

Lo sfasamento senza cavità, ϵ , è amplificato di un fattore $N = \frac{2\mathcal{F}}{\pi}$ corrispondente ad un numero effettivo di passaggi fra gli specchi.

Tenendo quindi conto di questo fattore di guadagno nella lunghezza effettiva del campo magnetico l'ellitticità *totale* indotta fra gli specchi

della cavità sarà

$$\Psi = \frac{2\mathcal{F}}{\pi} \psi.$$

Quindi nell'espressione (16) la quantità $\psi(t)$ va sostituita con $\Psi(t) = \frac{2\mathcal{F}}{\pi} \psi(t)$:

$$\begin{aligned} I_Y(t) &= I_0 \left[\sigma^2 + \left| i \frac{2\mathcal{F}}{\pi} \psi(t) + i\beta + i\eta(t) \right|^2 \right] = \\ &= I_0 \left[\sigma^2 + \eta(t)^2 + 2 \frac{2\mathcal{F}}{\pi} \psi(t)\eta(t) + 2\eta(t)\beta + \dots \right]. \end{aligned} \quad (18)$$

Con questa tecnica si possono ottenere fattori di amplificazione $N \approx 5 \times 10^5$ di gran lunga superiore a quelli delle cavità multipass. Inoltre le dimensioni del fascio rimangono dell'ordine del fascio in ingresso.

Estrazione del segnale

Nell'esperimento PVLAS il modulatore genera un'ellitticità sinusoidale $\eta(t) = \eta_0 \sin(2\pi\nu_m t + \theta_m)$ mentre l'ellitticità totale indotta dal campo magnetico rotante è $\Psi(t) = \frac{2\mathcal{F}}{\pi} \psi_0 \sin 2\vartheta(t)$, con $\vartheta(t) = 2\pi\nu_B t + \theta_B$. Le intensità $I_Y(t)$ e $I_X(t)$ sono raccolte ciascuna da un fotodiode che le converte in corrente $i_Y(t)$ e $i_X(t)$ con un'efficienza q . Come si deduce dall'equazione (18) la corrente $i_Y(t)$ presenta una serie di componenti di Fourier con relative ampiezze e fasi riportate in Tabella 1, mentre la corrente $i_X(t)$ è sostanzialmente costante nel tempo: $i_X(t) = i_X$. Tenendo conto della piccola perdita dovuta all'analizzatore, i_X permette la misura indiretta di I_0 altrimenti inaccessibile.

Il termine $\eta^2(t)$ genera un segnale $i_{2\nu_m}$ alla frequenza $2\nu_m$ e serve come normalizzazione per determinare η_0 . Il doppio prodotto $2\eta(t)\psi(t)$ genera due segnali $i_{\nu_m \pm 2\nu_B}$ rispettivamente alle frequenze $(\nu_m + 2\nu_B)$ e $(\nu_m - 2\nu_B)$. Da queste componenti di Fourier e dal valore di i_X si può determinare Ψ_0 :

$$\Psi_0 = \frac{i_{\nu_m + 2\nu_B} + i_{\nu_m - 2\nu_B}}{2\sqrt{2}i_X i_{2\nu_m}}.$$

Sperimentalmente risulta più conveniente demodulare il segnale $i_Y(t)$ alla frequenza del modulatore ν_m in modo da riportare il segnale di ellitticità a bassa frequenza. I due segnali alle frequenze $(\nu_m \pm 2\nu_B)$ sono unite in un unico segnale

Tabella 1: Componenti di Fourier dei segnali in corrente dopo l'analizzatore.

Frequenza	Componente di Fourier	Intensità/ I_0	Fase
DC	$i_Y(0)$	$\sigma^2 + \beta^2 + \eta_0^2/2$	–
ν_m	i_{ν_m}	$2\beta\eta_0$	θ_m
$\nu_m \pm 2\nu_B$	$i_{\nu_m \pm 2\nu_B}$	$\eta_0 \frac{2F}{\pi} \psi_0$	$\theta_m \pm 2\theta_B$
$2\nu_m$	$i_{2\nu_m}$	$\eta_0^2/2$	$2\theta_m$

$i_{2\nu_B}$ alla frequenza $2\nu_B$. Per farlo si usa un amplificatore lock-in. L'espressione per l'ellitticità diventa allora

$$\Psi_{2\nu_B} = \frac{i_{2\nu_B}}{2\sqrt{2}i_X i_{2\nu_m}} = \frac{\eta_0 i_{2\nu_B}}{4 i_{2\nu_m}}$$

Se $i_{2\nu_B}$ è un segnale sinusoidale puro, tale è anche $\Psi_{2\nu_B}$, che può essere estratto dal rumore incoerente presente integrando nel tempo.

Rumore

La questione fondamentale ora è: qual'è la più piccola ellitticità osservabile e in quanto tempo? Dopo la demodulazione della corrente i_Y la sua densità spettrale di rumore r_{rms} , $r_{rms}(\nu)$, funzione della frequenza ν , genera una densità spettrale di *picco* di rumore in ellitticità, $s_\psi(\nu)$, corrispondente a

$$s_\psi(\nu) = \frac{\eta_0 r_{rms}(\nu)}{2 i_{\nu_m}} = \frac{r_{rms}(\nu)}{i_X \eta_0} \quad (19)$$

dove abbiamo sostituito i_0 con i_X . Poiché l'indeterminazione Δi_{rms} di una corrente si determina dalla densità spettrale r_{rms} tramite

$$\Delta i_{rms} = \frac{r_{rms}}{\sqrt{t}},$$

con t il tempo di integrazione, anche l'indeterminazione di un'ellitticità diminuisce con $1/\sqrt{t}$.

Riportiamo qui di seguito le più note sorgenti di rumore che generalmente riguardano esperimenti di ottica come PVLAS.

Il rumore fondamentale, noto come shot-noise, sotto il quale non si può scendere è dovuto alla natura corpuscolare della corrente elettrica. La densità spettrale in corrente $r_{rms}^{(shot)}$ di una corrente continua i_{DC} è data da

$$r_{rms}^{(shot)} = \sqrt{2e i_{DC}} = \sqrt{2e i_0 \left(\frac{\eta_0^2}{2} + \sigma^2 \right)}$$

dove e è la carica elettrica dell'elettrone e i_{DC} è la corrente continua generata dall'intensità I_Y sul fotodiodo PDE (vedi Fig. 5). Inoltre lo spettro in frequenza di questo rumore è piatto.

Dalla (19) si ricava che la densità spettrale di sensibilità limite in ellitticità, $s_{(shot)}$, è

$$s_{(shot)} = \sqrt{\frac{2e}{i_X} \left(\frac{\eta_0^2/2 + \sigma^2}{\eta_0^2} \right)}.$$

Nel caso in cui $\eta_0^2 \gg \sigma^2$ la sensibilità limite si riduce a

$$s_{(shot)} = \sqrt{\frac{e}{i_X}}$$

e dipende solo dall'intensità di corrente i_X . Con un'intensità $I_X = 2.5$ mW e un'efficienza $q = 0.7$ A/W la sensibilità risulta $s_{(shot)} = 10^{-8} \frac{1}{\sqrt{\text{Hz}}}$.

Altri contributi al rumore in corrente vengono dal rumore Johnson

$$r_{rms}^{(Johnson)} = \sqrt{\frac{4k_B T}{G}},$$

dove k_B è la costante di Boltzmann, T è la temperatura assoluta e G è il guadagno in V/W del sensore di luce, e dalla corrente di buio i_{Dark} del fotodiodo PDE

$$r_{rms}^{(Dark)} = \sqrt{2e i_{Dark}}$$

ciascuno contribuendo alla densità spettrale di rumore di ellitticità:

$$s_{(Johnson)} = \sqrt{\frac{4k_B T}{G}} \frac{1}{i_X \eta_0}$$

$$s_{(Dark)} = \frac{\sqrt{2e i_{Dark}}}{i_X \eta_0}.$$

Infine c'è il contributo dovuto alla densità spettrale di rumore in intensità $s_{I_0}(\nu)$ di I_0 espresso come $s_{I_0}(\nu)/I_0 = RIN(\nu)$ il cui rumore r_{rms} in

corrente è quindi

$$r_{\text{rms}}^{(\text{RIN})}(\nu) = \text{RIN}(\nu) i_X$$

e la sensibilità

$$s_{(\text{RIN})} = \text{RIN}(\nu_m) \frac{\sqrt{(\sigma^2 + \eta_0^2/2)^2 + (\eta_0^2/2)^2}}{\eta_0}$$

dove in questa espressione si è tenuto conto sia del contributo della DC sia di quello della frequenza $2\nu_m$.

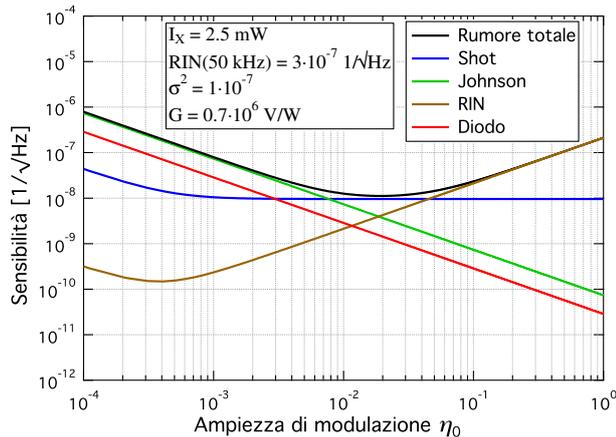


Figura 6: Contributo delle varie sorgenti di rumore note.

Tutti questi rumori sono funzione dell'ampiezza di modulazione η_0 e sono mostrati in Fig. 6. Il rumore totale è la somma in quadratura di tutti i contributi:

$$s_{(\text{tot})} = \sqrt{s_{(\text{shot})}^2 + s_{(\text{Johnson})}^2 + s_{(\text{Dark})}^2 + s_{(\text{RIN})}^2}$$

Come si vede esiste un'ampiezza di modulazione di $\eta_0 \approx 2 \times 10^{-2}$ per la quale il rumore totale previsto ha un minimo che corrisponde al limite shot-noise.

Per concludere, data la densità spettrale di rumore in ellitticità $s_{(\text{tot})}$, il tempo necessario per misurare con rapporto segnale su rumore unitario un valore di ellitticità $\Psi(t)$, con andamento sinusoidale puro di frequenza ν , è

$$t = \left(\frac{s_{(\text{tot})}}{\Psi_\nu} \right)^2$$

I magneti

Come si deduce dalle equazioni (6) e (10), l'ellitticità ψ indotta da un campo magnetico in vuoto è proporzionale a $B_{\text{ext}}^2 \sin 2\vartheta$. Per massimizza-

re ψ è necessario generare un campo magnetico dipolare il più intenso possibile compatibile con altre due restrizioni sperimentali importanti: il campo magnetico deve essere variabile nel tempo con una frequenza la più alta possibile (decine di Hz) e il duty cycle deve essere il massimo possibile per permettere tempi di integrazione lunghi. Una modulazione di ψ nel tempo può essere ottenuta sia variando l'intensità del campo, $B_{\text{ext}} = B_{\text{ext}}(t)$, sia modulando l'angolo fra il campo esterno e la direzione della polarizzazione della luce, $\vartheta = \vartheta(t)$.

Si potrebbe pensare in un primo momento all'utilizzo di un magnete superconduttore che in effetti può generare campi magnetici molto intensi. Sorgono, però, alcune limitazioni. Se si vuole un campo di qualche tesla su una lunghezza di qualche metro e diametro di qualche centimetro, serve una corrente di qualche migliaio di ampere. Inoltre l'induttanza di un tale magnete è dell'ordine di 1 H. L'energia immagazzinata risulta allora di qualche MJ. Modulare una tale quantità di energia a qualche decina di hertz richiede una potenza enorme. Inoltre i magneti superconduttori hanno un limite intrinseco sulla rapidità con cui possono variare il campo magnetico. Per evitare transizioni da superconduttore a conduttore normale (quench) solo modulazioni in corrente fino a qualche millihertz possono essere applicate.

Per ovviare a questi problemi, in una versione precedente dell'esperimento PVLAS [13], si è utilizzato un magnete superconduttore le cui bobine potevano essere messe in corto circuito dopo averle caricate di corrente e l'alimentazione quindi scollegata. L'asse del magnete era posto in verticale e la modulazione dell'ellitticità era ottenuta ruotando l'intero sistema magnete e criostato riempito di elio liquido per un peso totale di 6 tonnellate. In questo modo era possibile modulare l'angolo $\vartheta(t)$ nel tempo fino ad una frequenza $\nu_B = 0.5 \text{ Hz}$ con un campo fino a 5.5 T. La limitata fornitura di elio liquido e la disponibilità di un solo magnete limitavano la ricerca di segnali spuri e riducevano il duty cycle a qualche settimana di presa dati all'anno. L'idea del magnete superconduttore fu allora abbandonata.

La disponibilità di magneti permanenti ad alto campo ha permesso una soluzione alternativa. Oggi è possibile acquistare magneti dipolari con

campi di 2.5 T su diametri di qualche centimetro. Nonostante il campo non sia intenso quanto quello generato da un magnete superconduttore, un magnete permanente può essere ruotato a diverse decine di hertz essendo costituito da un cilindro di ferro. Inoltre il campo magnetico è sempre acceso e ha costi di esercizio nulli.

Nella configurazione di Halbach mostrata in Fig. 7 il magnete è realizzato ponendo una serie di spicchi di magneti permanenti su una circonferenza. Ciascun elemento è magnetizzato con un preciso orientamento 2ζ rispetto alla direzione del campo dipolare desiderato. In questo modo il campo magnetico risulta dipolare e uniforme all'interno del foro centrale. Per un magnete di lunghezza $L \gtrsim R_{\text{ext}}$ l'intensità del campo è

$$B \simeq B_{\text{rem}} \ln \left(\frac{R_{\text{ext}}}{R_{\text{int}}} \right), \quad (20)$$

dove $B_{\text{rem}} = 1.1 - 1.3 \text{ T}$ per magneti di Neodimio-Ferro-Boro ($\text{Nd}_2\text{Fe}_{14}\text{B}$) e R_{ext} e R_{int} sono il raggio esterno ed interno del magnete. Nel caso mostrato in Fig. 7, destra, che rappresenta una sezione di uno dei due magneti dell'esperimento PVLAS, la circonferenza è suddivisa in 16 settori con un rapporto $\frac{R_{\text{ext}}}{R_{\text{int}}} = 10$. Il campo magnetico all'interno del foro risulta 2.5 T su una lunghezza totale di 1.64 m e $\int B^2 dL = 10.25 \text{ T}^2\text{m}$. Inoltre i magneti possono ruotare fino a 23 Hz corrispondente ad una frequenza del segnale di ellitticità di 46 Hz.

Nell'esperimento PVLAS sono stati installati due magneti identici anziché uno più grande per ottimizzare la ricerca di segnali spuri in presenza di campo. Infatti facendo ruotare i due magneti insieme ma con i campi magnetici ortogonali fra loro, il segnale fisico di ellitticità generato da una birifrangenza magnetica è nullo. Quindi l'osservazione di un segnale con i magneti ortogonali fra loro indica la presenza di sistematici.

Allineando le direzioni dei campi magnetici si realizza la configurazione analoga ad un solo magnete di lunghezza doppia.

Condizioni sperimentali dell'esperimento PVLAS

Una fotografia dell'intero apparato e uno schema del banco ottico sono mostrati in Fig. 8. Il laser utilizzato dall'esperimento PVLAS è un

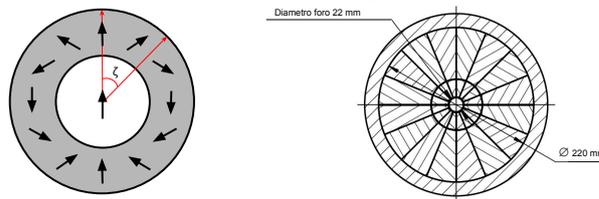


Figura 7: Sinistra: Concetto della configurazione Halbach per un magnete dipolare. La direzione della magnetizzazione di ciascun settore deve aver un angolo 2ζ rispetto alla direzione del campo desiderato. Destra: sezione di uno dei magneti di PVLAS.

Nd:YAG che emette nell'infrarosso ad una lunghezza d'onda di 1064 nm. Uno dei grandi vantaggi di questo tipo di laser è la possibilità di controllare la frequenza di emissione entro qualche decina di gigahertz. Questo è un elemento fondamentale per chi vuole utilizzare una cavità Fabry-Perot soprattutto se la finesse della cavità è molto elevata e di conseguenza la larghezza della risonanza è stretta come in questo caso ($\mathcal{F} \approx 7 \times 10^5$, $\Delta\nu_c \approx 70 \text{ Hz}$). Infatti, assumendo una lunghezza d'onda fissa, è sufficiente spostare uno specchio di una quantità $\sim \lambda/\mathcal{F} \approx 10^{-12} \text{ m}$ (circa 1/100 della dimensione di un atomo) perché la cavità esca dalla condizione di risonanza.

Risulta chiaro come sia necessario controllare la frequenza del laser con grande sensibilità e con una banda passante sufficiente per compensare vibrazioni meccaniche dell'ordine dei kilohertz.

Il laser è inserito all'interno di un circuito di retroazione che lo mantiene agganciato in frequenza alla cavità.

Per quanto riguarda la cavità Fabry-Perot gli specchi utilizzati hanno una riflettività $\mathcal{R} = 99.99953\%$ con perdite $p \approx 2 \text{ ppm}$. La finesse misurata risulta $\mathcal{F} \approx 7 \times 10^5$ corrispondente ad un guadagno in lunghezza effettiva di $N = \frac{2\mathcal{F}}{\pi} \approx 4.5 \times 10^5$. Con una lunghezza totale di campo magnetico $L = 1.64 \text{ m}$ la lunghezza effettiva del campo magnetico attraversato dalla luce risulta $\approx 700 \text{ km}$. Nel caso dell'esperimento PVLAS la cavità Fabry-Perot è lunga $D = 3.3 \text{ m}$ corrispondente a $\nu_{f_{sr}} = 45.4 \text{ MHz}$. Dopo lo spegnimento del laser, l'intensità della luce in uscita dalla cavità decade esponenzialmente con una

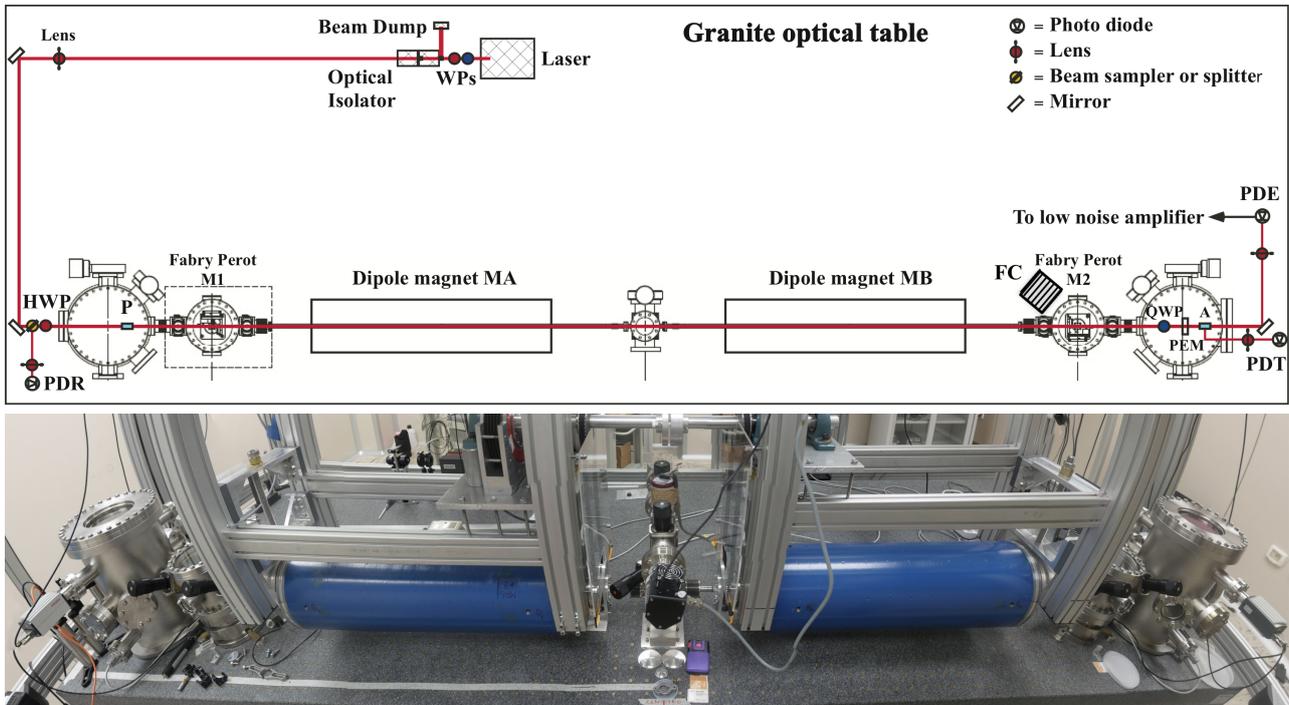


Figura 8: Pannello superiore: Schema del banco ottico di PVLAS. Pannello inferiore: Fotografia dell'apparato.

costante di tempo

$$\tau = \frac{\mathcal{F}}{2\pi\nu_{fsr}} = 2.4 \text{ ms.}$$

Si noti che questo significa che la larghezza di una risonanza $\Delta\nu_c$ è di soli 67 Hz da confrontare con la frequenza della luce ($\lambda = 1064 \text{ nm}$) di $2.8 \times 10^{14} \text{ Hz}$.

Il modulatore di ellitticità scelto per l'esperimento è di tipo fotoelastico risonante che lavora ad una frequenza $\nu_m = 50 \text{ kHz}$. I magneti ruotano normalmente ad una frequenza $\nu_B = 10 \text{ Hz}$.

Con questi parametri sperimentali l'ellitticità prevista dovuta alla birifrangenza magnetica del vuoto è

$$\psi_{\text{QED}} = \frac{2\mathcal{F}}{\lambda} 3A_e \int B_{\text{ext}}^2 dL = 5 \times 10^{-11} \quad (21)$$

Il rumore calcolato con $I_X = 2.5 \text{ mW}$, rappresentato in Fig. 6, mostra un valore minimo corrispondente allo shot noise intorno a $s_{\text{tot}} = 10^{-8} \frac{1}{\sqrt{\text{Hz}}}$ ottenuto con modulazioni dell'ordine di $\eta_0 \simeq 2 \times 10^{-2}$. In principio il tempo necessario per misurare l'effetto della birifrangenza magnetica del vuoto con rapporto segnale su rumore

unitario è

$$t = \left(\frac{s_{\text{(tot)}}}{\psi_{\text{QED}}} \right)^2 = 11 \text{ ore} \quad (22)$$

Purtroppo sperimentalmente risulta che l'effettiva sensibilità è $s_{\text{(tot)}}^{\text{(PVLAS)}} = 4 \times 10^{-7} \frac{1}{\sqrt{\text{Hz}}}$ ad una frequenza del segnale di 10 Hz. Si sospetta che questo rumore in eccesso sia originato dagli specchi della cavità ed è attualmente oggetto di studio.

Calibrazioni

La calibrazione dell'apparato è un aspetto fondamentale per verificare il corretto funzionamento del sistema. Nell'esperimento PVLAS le calibrazioni vengono fatte inserendo all'interno della cavità gas puri a bassa pressione. Questi gas sono soggetti a due effetti magneto-ottici importanti: l'effetto Faraday e l'effetto Cotton-Mouton. L'effetto Faraday consiste in una birifrangenza circolare proporzionale a \mathbf{B}_{ext} e genera una rotazione del piano di polarizzazione su un fascio che si propaga parallelamente al campo magnetico. L'effetto Cotton-Mouton invece consiste in una birifrangenza lineare proporzionale a B_{ext}^2 che genera un'ellitticità su un fascio iniziale.

Tabella 2: Costanti Cotton-Mouton per alcuni gas importanti e le pressioni parziali, $3A_e/\Delta n_u$, che generano una birifrangenza magnetica uguale a quella del vuoto. La misura riportata per H_2 è stata fatta a $\lambda = 633 \text{ nm}$ mentre tutte le altre a $\lambda = 1064 \text{ nm}$.

Specie	$\Delta n_u \text{ (T}^{-2}\text{atm}^{-1}\text{)}$	$3A_e/\Delta n_u \text{ (atm)}$
H_2	$(8.82 \pm 0.25) \times 10^{-15}$	4.5×10^{-10}
He	$(2.20 \pm 0.14) \times 10^{-16}$	1.8×10^{-8}
H_2O	$(6.67 \pm 0.21) \times 10^{-15}$	5.9×10^{-10}
N_2	$(-2.66 \pm 0.42) \times 10^{-13}$	1.5×10^{-11}
O_2	$(-2.29 \pm 0.08) \times 10^{-12}$	1.7×10^{-12}
Ar	$(7.5 \pm 0.5) \times 10^{-15}$	5.3×10^{-10}

mente polarizzato linearmente che si propaga perpendicolarmente al campo magnetico.

L'effetto Cotton-Mouton è quello usato nell'esperimento PVLAS per le calibrazioni. La birifrangenza indotta su un gas da un campo magnetico esterno è caratterizzata dalla birifrangenza unitaria, Δn_u , che rappresenta il valore della birifrangenza per $B_{\text{ext}} = 1 \text{ T}$ e pressione atmosferica. La birifrangenza di un gas è quindi data da

$$\Delta n = \Delta n_u P B_{\text{ext}}^2.$$

con P la pressione in atmosfere e B_{ext} in tesla. Nella Tabella 2 sono riportati i valori di Δn_u per alcuni gas importanti.

Ogni misura con gas fornisce una calibrazione sia in ampiezza che in fase. Infatti l'ellitticità $\Psi(t)$ indotta da un campo magnetico rotante è massima quando la direzione della polarizzazione è a 45° rispetto alla direzione del campo magnetico. Quindi nella trasformata di Fourier di $\Psi(t)$ la fase del segnale a $2\nu_B$ determina la condizione di ellitticità massima e quindi l'asse 'fisico' di una misura. Un'eventuale segnale di ellitticità in vuoto deve cadere su tale asse.

Le calibrazioni si eseguono misurando l'ellitticità e la relativa fase indotta in funzione della pressione del gas scelto. Prima di ogni presa dati si esegue una calibrazione tipicamente con gas Argon. Tra i vari gas che si possono usare, quello che evidenzia maggiormente la sensibilità dell'apparato è l'elio poiché questo gas ha la costante Cotton-Mouton più piccola. In Fig. 9 è riportato un grafico dell'ampiezza della birifrangenza dell'elio in funzione della pressione mentre in Fig. 10 è mostrato l'asse 'fisi-

co'. Il punto a pressione più bassa corrisponde a $\Delta n^{(\text{He})}/B_{\text{ext}}^2 = (8.6 \pm 1.0) \times 10^{-21}$ con $P = 32 \mu \text{ atm}$. La misura di questo singolo punto è durata circa 4 ore dimostrando la stabilità dell'apparato e il suo ottimo funzionamento.

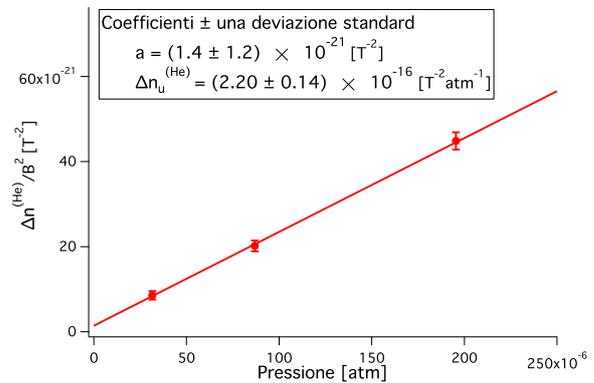


Figura 9: Calibrazione in ampiezza dell'apparato PVLAS con gas elio a bassa pressione.

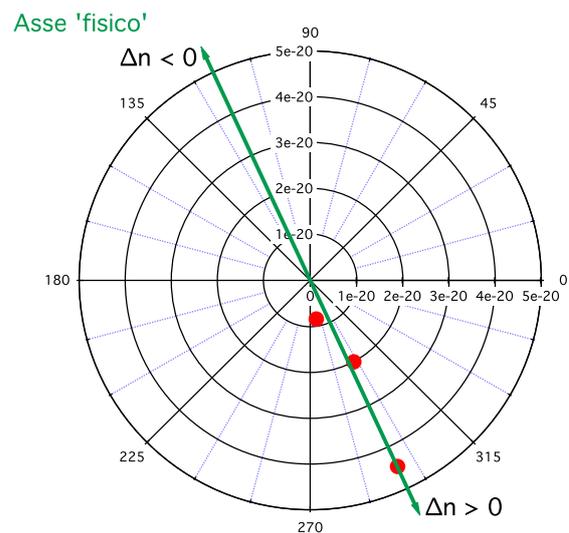


Figura 10: Grafico polare della componente della trasformata di Fourier a $2\nu_B$ che permette la calibrazione in fase dell'apparato PVLAS. Il gas utilizzato è elio a bassa pressione. In verde è rappresentato l'asse 'fisico'.

Poiché l'esperimento PVLAS si propone di misurare la birifrangenza magnetica del vuoto le costanti Cotton-Mouton Δn_u impongono un valore massimo per la pressione parziale, $3A_e/\Delta n_u$, di ciascuna specie di gas nel tubo da vuoto che attraversa i magneti corrispondente ai valori che generano una birifrangenza uguale a quella del vuoto. Questi valori sono riportati in Tabella 2. Per questo motivo il sistema di pompaggio da

vuoto è anche stato dotato di misuratori di gas residuo che permettono di tenere sotto controllo ciascuna specie di gas.

Misure in vuoto

Molti dettagli del montaggio dell'apparato sono stati tralasciati nella descrizione dell'esperimento PVLAS. Spesso, e questo è il caso anche per l'esperimento PVLAS, questi dettagli risultano fondamentali per eliminare effetti sistematici permettendo di sfruttare al meglio un apparato. L'implementazione dei magneti permanenti ha reso possibile una serie di studi sull'apparato che hanno permesso di portare gli errori sistematici al di sotto dell'errore statistico su tempi di integrazione di 23 giorni.

Nella Fig. 11 sono riportati i risultati sia del presente esperimento, denominato PVLAS-FE, sia di tentativi passati. Gli errori riportati corrispondono a una deviazione standard e come si vede tutti gli esperimenti passati hanno dato valori di $|\Delta n|/B_{\text{ext}}^2$ incompatibili con il valore previsto, indicando la presenza di contributi di effetti sistematici.

La collaborazione BFRT (Brookhaven-Fermilab-Rochester-Trieste) [14] fece il primo vero tentativo di misurare la birifrangenza magnetica del vuoto seguendo lo schema ottico di E. Iacopini e E. Zavattini del 1979 [12]. Si basava su due magneti superconduttori lunghi ciascuno 4.4 m capaci di generare un campo di 4 T. Il campo magnetico era modulato a 32 mHz fra un valore massimo di 3.87 T e un valore minimo di 2.63 T. Per aumentare il cammino ottico all'interno del magnete si utilizzava una cavità multi-pass ottenendo un numero massimo di passaggi di $N = 578$. Il laser utilizzato emetteva una lunghezza d'onda $\lambda = 514.5$ nm.

Anche la versione successiva, PVLAS-LNL (presso i Laboratori Nazionali di Legnaro dell'INFN) [13], utilizzava un magnete superconduttore lungo 1 m ma, per aumentare la frequenza di modulazione del segnale, il magnete e il criostato ruotavano intorno ad un asse verticale con frequenza 0.3 Hz, corrispondente ad una frequenza del segnale di 0.6 Hz. In questo caso il cammino ottico della luce nel campo magnetico era amplificato da una cavità Fabry-Perot. L'utilizzo di un Fabry-Perot è stato possibile grazie al-

l'evoluzione dei laser che permettevano agganci in frequenza. Questo ha permesso di aumentare il numero di passaggi effettivi nel campo magnetico a qualche decina di migliaia e di migliorare il risultato finale di un fattore circa 20 rispetto al valore di BFRT. La limitata disponibilità di elio liquido non ha permesso lo studio dettagliato dei sistematici limitando i tempi di integrazione a $t \approx 6 \times 10^4$ s.

Una svolta importante dal punto di vista sperimentale è stata l'introduzione di magneti permanenti capaci di generare campi magnetici di circa 2.5 T su un diametro di 2 cm. Questi magneti, molto più compatti, hanno permesso di progettare apparati 'table-top'. Un primo prototipo, PVLAS-Test [15], aveva un campo magnetico generato da due magneti identici ciascuno lungo 20 cm. Per la prima volta è stata sperimentata la cancellazione di un segnale fisico mantenendo i campi magnetici orientati a 90° fra loro durante la rotazione per studiare effetti sistematici. Misure con i magneti in rotazione fino a 3 Hz sono state eseguite con orientamenti sia a 0° che a 90° . Anche la finesse della cavità Fabry-Perot era aumentata a $\mathcal{F} = 2.4 \times 10^5$. La presenza di picchi sistematici con i magneti orientati a 90° ha limitato il tempo di integrazione a $\simeq 10^4$ s. Nonostante ciò questo prototipo 'table-top' ha permesso un miglioramento rispetto a PVLAS-LNL di circa un fattore 2.

L'esperimento BMV (Birifringence Magnétique du Vide, presso l'Università di Tolosa in Francia) [16], in corso di sviluppo, adotta uno schema ottico molto simile a quello di PVLAS ma, invece di utilizzare magneti a campo fisso, rotanti, utilizza magneti impulsati ad alto campo non superconduttori. Intensità di campo fino a 14 T sono stati ottenuti su tempi dell'ordine di 1 ms. La finesse della cavità è $\mathcal{F} = 4.5 \times 10^5$ e la prima versione ha dato un risultato simile a quello di PVLAS-Test.

L'esperimento PVLAS nella sua versione attuale e discusso nelle sezioni precedenti, è installato a Ferrara presso la sezione INFN e Dipartimento di Fisica e Scienze della Terra. Esso si basa sul prototipo 'table-top' di PVLAS-Test e implementa diverse migliorie. I risultati presentati nella Fig. 11, indicati come PVLAS-FERRARA, hanno permesso di mettere il miglior limite sulla birifrangenza magnetica del vuoto e sull'interazione

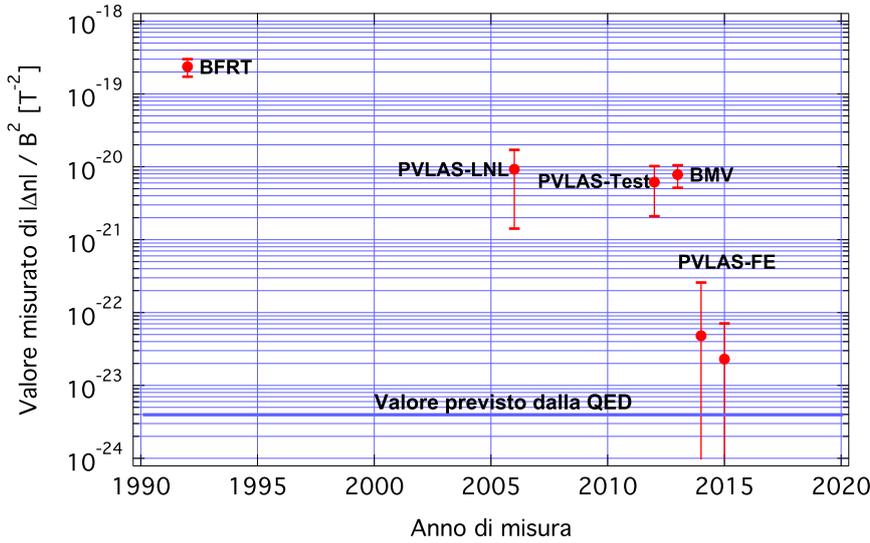


Figura 11: Misure della birifrangenza magnetica del vuoto negli anni riportate come valore di $|\Delta n|/B_{ext}^2$. La misura di PVLAS-FE del 2015 include quella del 2014.

fotone-fotone a bassa energia [11]

$$\Delta n^{(PVLAS)} = (-1.5 \pm 3.0) \times 10^{-22} @ 2.5 \text{ T.} \quad (23)$$

Assumendo un andamento di questa birifrangenza con B_{ext}^2 , come previsto dalla Lagrangiana (1), si ottiene

$$\frac{\Delta n^{(PVLAS)}}{B_{ext}^2} = (-2.4 \pm 4.8) \times 10^{-23} \text{ T}^{-2},$$

compatibile sia con zero che con il valore previsto

$$\frac{\Delta n^{(EK)}}{B_{ext}^2} = 3A_e = 3.96 \times 10^{-24} \text{ T}^{-2},$$

entro una deviazione standard. Il risultato globale di PVLAS è stato ottenuto con un tempo totale di integrazione di $t = 2 \times 10^6 \text{ s}$ grazie all'approfondito studio degli effetti sistematici. Purtroppo guadagnare il fattore mancante di circa un fattore 12 per misurare la birifrangenza magnetica del vuoto necessiterebbe un aumento del tempo di integrazione t di un fattore $12^2 = 144$ corrispondente a 9 anni, un tempo irrealistico.

Con la lamina quarto d'onda inserita prima del modulatore è stato possibile anche mettere un limite anche sul dicroismo magnetico del vuoto

$$\Delta \kappa^{(PVLAS)} = (-1.6 \pm 3.5) \times 10^{-22} @ 2.5 \text{ T.} \quad (24)$$

La comprensione del rumore intrinseco dovuto presumibilmente agli specchi della cavità insieme ad un aumento ulteriore della frequenza del segnale di ellitticità sono due possibili metodi per migliorare la misura e osservare diret-

tamente questo fenomeno quantistico del vuoto con l'apparato PVLAS.

Fisica nuova

In realtà l'apparato PVLAS è sensibile a qualunque processo fisico che permette l'accoppiamento di due fotoni. Fluttuazioni del vuoto in particelle ipotetiche (MilliCharge Particles, MCP) leggere di massa m_e e carica q_e (sia scalari che fermioni) sono un esempio. Il processo potrebbe essere identico a quello mostrato in Fig. 2 dove al posto di A_e c'è A_e con

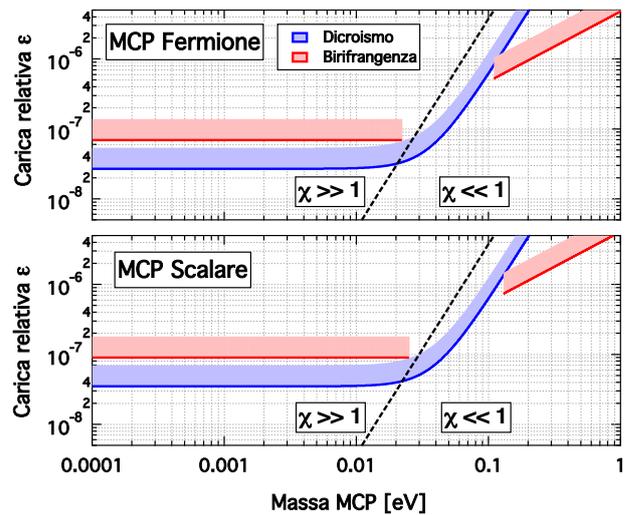


Figura 12: Grafici di esclusione per le MCP al 95% c.l. derivati dai limiti in dicroismo (24) e birifrangenza (23). Le zone escluse sono sopra le curve. La curva rossa della birifrangenza è spezzata in quanto per $\chi = 1$ c'è un cambio di segno.

$$A_\epsilon = \frac{2}{45\mu_0} \frac{\epsilon^4 \alpha^2 \chi_\epsilon^3}{m_\epsilon c^2}$$

e $\epsilon = q_\epsilon/e$. Se la massa di tali particelle fosse minore dell'energia del fotone incidente si potrebbe anche aver anche un processo di creazione di particelle reali che implicherebbe anche un dicroismo magnetico del vuoto. I limiti posti dall'esperimento PVLAS, sia in dicroismo che in birifrangenza, mettono dei limiti su ϵ in funzione di m_ϵ mostrati in Fig. 12 dove il parametro χ

$$\chi \equiv \frac{3}{2} \frac{\hbar\omega}{m_\epsilon c^2} \frac{\epsilon e B_{\text{ext}} \hbar}{m_\epsilon^2 c^2}.$$

distingue i casi di massa piccola oppure massa grande. Notiamo esplicitamente che questi limiti sulla carica elettrica si applicano anche ai neutrini.



- [1] O. Halpern: *Scattering Processes Produced by Electrons in Negative Energy States*, Phys. Rev., **44** (1933) 855.2.
- [2] S. Weinberg: *The Cosmological Constant Problem*, Rev. Mod. Phys., **61** (1989) 1.
- [3] K. M. Evanson *et al.*: *Speed of Light from Direct Frequency and Wavelength Measurements of the Methane-Stabilized Laser*, Phys. Rev. Lett., **29** (1972) 1346.
- [4] H. Euler, B. Kockel: *Über die Streuung von Licht an Licht nach der Diracschen Theorie (The scattering of light by light in the Dirac theory)*, Naturwiss., **23** (1935) 246.³
- [5] W. Heisenberg, H. Euler: *Folgerungen aus der Diracschen Theorie des Positrons (Consequences of Dirac's Theory of Positrons)*, Zeit. f. Phys., **98** (1936) 714.⁴
- [6] J. Schwinger: *On gauge invariance and vacuum polarization*, Phys. Rev., **82** (1951) 664.
- [7] S. L. Adler: *Photon Splitting and Photon Dispersion in a Strong Magnetic Field*, Ann. of Phys., **67** (1971) 599.
- [8] R. P. Mignani *et al.*: *Evidence for vacuum birefringence from the first optical-polarimetry measurement of the isolated neutron star RX J1856.5–3754*, Mon. Notices Royal Astron. Soc., **465** (2017) 492.
- [9] ATLAS collab.: *Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC*, Nature Physics, **13** (2017) 852.
- [10] R. J. Stoneham: *Photon splitting in the magnetised vacuum*, J. Phys. A: Math. Gen., **12** (1978) 2187.

³Una traduzione è scaricabile da http://www.neoclassical-physics.info/uploads/3/0/6/5/3065888/euler_-_scattering_of_light_by_light.pdf

⁴Una traduzione è scaricabile da [arXiv:physics/0605038](https://arxiv.org/abs/0605038)

- [11] F. Della Valle *et al.* (PVLAS collab.): *The PVLAS experiment measuring vacuum magnetic birefringence and dichroism with a birefringent Fabry-Perot cavity*, Eur. Phys. J. C, **76:24** (2016) 1.
- [12] E. Iacopini e E. Zavattini: *Experimental method to detect the vacuum birefringence induced by a magnetic field*, Phys.Lett.B., **85** (1979) 151.
- [13] M. Bregant *et al.* (PVLAS collab.): *Limits on low energy photon-photon scattering from an experiment on magnetic vacuum birefringence*, Phys.Rev. D, **78** (2008) 032006.
- [14] R. Cameron *et al.*: *Search for nearly massless, weakly coupled particles by optical techniques*, Phys.Rev. D, **47** (1993) 3707.
- [15] F. Della Valle *et al.* (PVLAS collab.): *Measurements of vacuum magnetic birefringence using permanent dipole magnets: the PVLAS experiment*, New J. Phys., **15** (2013) 053026.
- [16] A. Cadène *et al.* (BMV collab.): *Vacuum magnetic linear birefringence using pulsed fields: status of the BMV experiment*, Eur. Phys. J. D, **68:16** (2014) 1.



Guido Zavattini: Si è laureato nel 1989 presso l'Università di Pisa e dal 2009 è Professore Associato presso l'Università di Ferrara. È Responsabile dell'esperimento PVLAS e si occupa di misure di precisione con tecniche ottiche. Ha lavorato anche nel campo della fisica applicata alla medicina nucleare e nel campo dell'astronomia X.

La ricerca di nuova fisica nel vuoto quantistico

Se credete di aver capito la teoria dei quanti, vuol dire che non l'avete capita.

Richard Feynman

Marco Incagli Istituto Nazionale di Fisica Nucleare - INFN Pisa

La Meccanica Quantistica ha rivoluzionato la fisica imponendo idee e modi di pensare nuovi ed apparentemente in contraddizione con la realtà quotidiana. Il concetto di *vuoto quantistico* è senz'altro fra quelli che richiedono una modifica radicale del nostro approccio mentale: dal vuoto come *Nulla* al vuoto come *Tutto*. All'interno di questo *Tutto* si crea energia, sotto forma di onde elettromagnetiche, e si creano particelle che modificano le proprietà macroscopiche della materia. Attraverso misure di precisione, questi effetti sono stati isolati in laboratorio e ci consentono di esplorare non solo la materia nota, ma anche quella ancora sconosciuta.

Nascita della Meccanica Quantistica

Intorno al 1900 si sviluppò una teoria atomica nella quale l'atomo veniva descritto come un sistema solare in miniatura con al centro un nucleo pesante di carica positiva circondato da elettroni

leggeri¹ che gli ruotano intorno. Tuttavia i tentativi di spiegare il moto degli elettroni con le leggi dell'elettromagnetismo note al tempo furono un completo fallimento e molte predizioni risultarono sbagliate. La natura corpuscolare della luce, l'effetto fotoelettrico, la stessa stabilità atomica non potevano essere spiegate senza introdurre regole e concetti che sfuggivano totalmente al buon senso. Queste regole, sviluppate nei primi trenta anni del secolo scorso, confluirono nella Meccanica Quantistica.

L'esempio più noto di queste regole che "sfuggono al buon senso" è dato dal *Principio di indeterminazione*, o di *Heisenberg*: non è possibile determinare contemporaneamente velocità e posizione di una particella. Se Δp_x è l'incertezza con la quale è nota la velocità (più precisamente, l'impulso) di una particella e Δx è l'incertezza con la quale è nota la posizione, allora risulta:

$$\Delta p_x \Delta x \geq h/4\pi \quad (1)$$

essendo h la costante di Planck. Determinando la posizione di una particella con precisione sempre maggiore ($\Delta x \rightarrow 0$) si perde ogni informazione

¹La massa dell'elettrone è circa 2000 volte minore rispetto a quella del protone.



Figura 1: Il complesso di acceleratori presso il laboratorio Fermilab di Chicago

sulla sua velocità ($\Delta p_x \rightarrow \infty$). Questo in netto contrasto con il concetto deterministico classico. Tuttavia il valore sperimentale della costante h è:

$$h = 6.6 \times 10^{-35} \text{ J} \cdot \text{s}$$

cioè così “piccolo” che in tutti i fenomeni macroscopici questo effetto è completamente invisibile.

Una versione del principio di indeterminazione ancora più strana rispetto al senso comune e dalle conseguenze rilevanti per la nostra storia è quella che collega l’incertezza sulla misura della *energia* di un sistema (ΔE) con l’incertezza sulla variabile temporale Δt . Anche in questo caso vale la relazione:

$$\Delta E \Delta t \geq h/4\pi \quad (2)$$

Questa formulazione viene spesso descritta nel modo seguente: in un sistema, o in una trasformazione, l’energia può variare, e quindi non essere *conservata*, purchè ciò avvenga per un periodo Δt sufficientemente breve in modo da rispettare l’equazione 2. La relazione appena scritta, come vedremo in seguito, è alla base della struttura

quantistica del vuoto e della possibilità di creare, o estrarre, da esso coppie di particelle cariche², le quali poi si annichilano e scompaiono nuovamente nel vuoto. Infatti la fluttuazione in energia ΔE si può manifestare sotto forma di una coppia particella-antiparticella di massa totale $m = \Delta E/c^2$.

Queste particelle e antiparticelle prodotte nel vuoto quantistico si annichilano e scompaiono immediatamente, e per questo vengono dette *particelle virtuali*. Esse possono essere estratte dal vuoto, e diventare particelle reali, fornendo energia al sistema ad esempio sotto forma di onde elettromagnetiche oppure sotto forma di energia gravitazionale, come nelle vicinanze di un buco nero (radiazione di Hawking).

Anche se normalmente non rilevabili, queste particelle virtuali modificano la struttura elettromagnetica del sistema dando luogo ad effetti di schermaggio o di amplificazione della carica elettrica o magnetica del sistema stesso che possono essere misurati.

²La carica elettrica totale di un sistema deve comunque conservarsi, per cui non è possibile estrarre una singola particella carica.

Uno dei primi esempi sperimentali dell'effetto di tali particelle virtuali è stata la misura del *momento magnetico anomalo* di una particelle particolare chiamata *muone* e spesso identificata dalla lettera greca μ . Il termine *anomalo* si riferisce proprio agli effetti quantistici appena accennati, effetti non presenti nella trattazione classica.

Tale misura [1], effettuata per la prima volta alla fine degli anni '50 del secolo scorso, è talmente importante da essere stata ripetuta con precisione sempre maggiore prima al CERN (Ginevra) [2] e poi al laboratorio di Brookhaven (New York) [3]. Una versione moderna di questa misura sarà effettuata presso il laboratorio Fermilab (Chicago), mostrato in Fig. 1, e l'apparato sperimentale che sarà utilizzato è attualmente in fase di costruzione.

Prima di addentrarci nella descrizione del *momento magnetico anomalo* del muone e dell'esperimento attualmente in fase di realizzazione, è necessario spendere alcune parole sulla struttura fondamentale della materia secondo la teoria attuale nota con il nome Modello Standard.

Il Modello Standard delle interazioni fondamentali

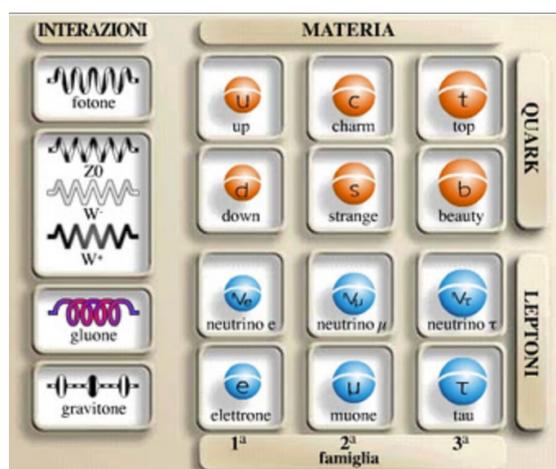


Figura 2: Il Modello Standard delle interazioni fondamentali. A sinistra le "interazioni", cioè le particelle mediatrici delle forze, a destra i mattoni che costituiscono la materia. In questa tabella non è incluso il bosone di Higgs recentemente scoperto al Large Hadron Collider (LHC) del CERN, Ginevra.

Secondo il Modello Standard delle interazioni fondamentali, le particelle che costituiscono i mattoni fondamentali della materia che osserviamo si dividono in due categorie, quark e leptoni (Fig. 2). I quark si legano insieme e formano strutture complesse, chiamate *adroni*³, fra le quali troviamo il protone ed il neutrone, mentre l'elettrone fa parte della categoria dei *leptoni*⁴.

La differenza fondamentale fra quark e leptoni è che i primi posseggono, oltre alla carica elettrica, anche una carica analoga detta *carica di colore* che permette loro di legarsi in strutture compatte che sono alla base della stabilità dei nuclei. La forza di legame dovuta a questa carica è detta, con fantasia decisamente limitata, *forza forte* in quanto più forte della repulsione elettrostatica che tende a far allontanare cariche dello stesso segno. Questa *carica di colore* non esiste nei leptoni, motivo per cui non si osservano strutture nucleari composte da elettroni.

Quark e leptoni sono inoltre strutturati in *famiglie*. Alla prima famiglia appartengono il doppietto di quark (u, d) (up e down), aventi carica elettrica, rispetto al protone, $+2/3$ e $-1/3$, rispettivamente, ed il doppietto di leptoni (ν_e, e) (neutrino elettronico, elettrone), di carica elettrica 0 e -1 . Al di là dei nomi, più o meno "fantasiosi", tutta la materia è composta da particelle di questa prima famiglia. In particolare, i protoni ed i neutroni sono formati da combinazioni ben specifiche dei due quark: il protone è composto dalla combinazione uud , mentre il neutrone da quella udd . Si noti che, sommando le cariche frazionarie dei quark, si ottengono i valori $+1$ per la carica elettrica del protone e 0 per quella del neutrone.

Tuttavia per mezzo di misure effettuate sia su radiazione cosmica sia in laboratori sotterranei, una seconda, e poi una terza, famiglia sono state osservate nel corso degli ultimi ottant'anni. Le tre famiglie hanno la stessa struttura e gli stessi numeri quantici, come se la natura ripetesse se stessa. I componenti delle diverse famiglie differiscono per la loro massa che, per quanto riguarda le particelle cariche, varia dall'elettrone, 2000 volte più leggero del protone, al quark *top*⁵,

³Dal termine greco *adròs* = intenso, forte

⁴Dal termine greco *leptòs* = leggero

⁵Osservato per la prima volta proprio al laboratorio Fermilab nel 1994.

che ha una massa di poco inferiore a quella di un nucleo di oro, circa $4 \cdot 10^5$ volte maggiore di quella dell'elettrone! Al momento attuale non abbiamo nessuna spiegazione sul perché esistano 3 famiglie né sul perché ci sia questa grande dispersione in massa.

Il primo membro della seconda famiglia ad essere scoperto è stato il *muone*, o leptone μ , osservato nel 1936 da Anderson e Neddermeyer [4] studiando la radiazione cosmica. Al momento dell'osservazione, il leptone μ sembrava non avere nessun ruolo nelle interazioni nucleari tanto che il fisico americano Isaac Rabi, premio Nobel nel 1944 per i suoi studi pionieristici sulla risonanza magnetica nucleare, commentò la scoperta con le parole: "Who ordered that?".

Nel 1975 un terzo elettrone pesante, chiamato leptone τ , fu scoperto all'acceleratore lineare di Stanford (*Stanford Linear Accelerator Center - SLAC*) nei pressi di San Francisco [5].

I tre leptoni carichi, elettrone, muone e τ , sono molto simili tra loro per quanto riguarda le interazioni con il mondo esterno, ma hanno masse molto diverse. Nelle unità usate nel campo delle particelle elementari, gli *elettronvolt*⁶ (eV), esse valgono: $m_e = 0.511$ MeV, $m_\mu = 105.7$ MeV, $m_\tau = 1777$ MeV, dove la M maiuscola indica mega, cioè 10^6 . Come riferimento, nelle stesse unità la massa del protone vale $m_p \simeq 938$ MeV. La grande differenza di massa fa sì che la *vita media*⁷ di queste particelle sia drasticamente diversa. Gli elettroni sono particelle stabili, e quindi hanno vita media infinita, mentre i muoni hanno una vita media di alcuni microsecondi ($t_\mu = 2.2 \mu s$); dopo questo tempo, mediamente, un muone decade (cioè "scompare trasformandosi") in un elettrone e due neutrini. Il leptone τ ha una vita molto più breve e decade dopo soli 10^{-13} secondi!

L'elettrone, essendo stabile, è storicamente il leptone più studiato, a partire dagli esperimenti pionieristici di Rutherford e Thomson. Tuttavia anche le proprietà del muone sono state studiate

⁶L'elettronvolt è una misura di energia pari a $\sim 1.6 \cdot 10^{-19}$ J.

Quando viene riferita alla massa, si sottintende che questa energia debba essere divisa per la velocità della luce al quadrato: $1 \text{ MeV} \rightarrow 1 \text{ MeV} / c^2 = 1.8 \cdot 10^{-30}$ kg.

⁷Il concetto di *vita media* è molto simile a quello di *tempo di decadimento* di una sostanza radioattiva: non si può determinare la vita di una singola particella, ma solamente la sua distribuzione statistica o, il che è equivalente, la probabilità di *decadimento* in funzione del tempo.

accuratamente, risultando rilevanti prima nelle verifiche della teoria chiamata Elettrodinamica Quantistica (*Quantum ElectroDynamics - QED*) e poi nella ricerca di Nuova Fisica al di là del Modello Standard.

Dalla Meccanica all'Elettrodinamica Quantistica

Le particelle fondamentali brevemente descritte nel paragrafo precedente sono caratterizzate, oltre che dalla loro massa e dalla carica elettrica, anche da altri *numeri quantici* fondamentali, il più noto dei quali è lo *spin*. Talvolta questa grandezza viene descritta come una rotazione della particella attorno al proprio asse, rotazione che genera un *momento magnetico*, facendo così un parallelo con il momento magnetico terrestre derivante dalla rotazione della Terra intorno al proprio asse; in effetti *spin* significa proprio "rotazione". Tuttavia questa descrizione è puramente illustrativa: gli elettroni sono particelle puntiformi, quindi non hanno un asse intorno al quale ruotare, e lo spin è realmente una proprietà intrinseca della particella stessa, così come lo è la carica elettrica.

Secondo la teoria quantistica relativistica sviluppata da Dirac alla fine degli anni '20, allo spin è associato un momento magnetico, simile a quello posseduto dall'ago di una bussola, che per i leptoni ha intensità:

$$\mu = \frac{g}{2} \cdot \frac{e}{m} S \quad (3)$$

dove e , m ed S rappresentano, rispettivamente, carica, massa e spin della particella, mentre g è una costante di proporzionalità che, nella teoria di Dirac [6], vale esattamente 2.

Intorno alla fine degli anni '40 Kush e Foley [7] osservarono⁸ che il valore di g era diverso dal valore previsto per una quantità pari a circa l'uno per mille. In particolare, introducendo il *momento magnetico anomalo*

$$a = \frac{g}{2} - 1 = \frac{g - 2}{2} \quad (4)$$

⁸La misura fu fatta studiando la *struttura iperfine* degli spettri atomici in presenza di un campo magnetico esterno; la *struttura iperfine* è dovuta alla interazione dello spin degli elettroni con il nucleo.

risultò

$$a = 0.00118 \pm 0.00003 \quad (5)$$

Vedremo in seguito che ogni particella ha il proprio valore del *momento magnetico anomalo*, per cui parleremo di a_e per l'elettrone, a_μ per il muone e così via.

Negli stessi anni si stava sviluppando una teoria che tentava di unificare la Meccanica Quantistica con l'Elettromagnetismo, inglobando ed ampliando le formulazioni di Dirac, Schrödinger, Pauli ed altri. Questa teoria venne chiamata *Elettrodinamica Quantistica* (QED); fra i principali autori vanno menzionati i fisici americani Richard Feynman e Julian Schwinger e il fisico giapponese San-Itiro Tomonaga, premi Nobel nel 1965. In particolare, nel 1948 Schwinger [8] calcolò per primo il valore di a ottenendo:

$$a = 0.00116 \quad (6)$$

Un successo spettacolare per la QED!

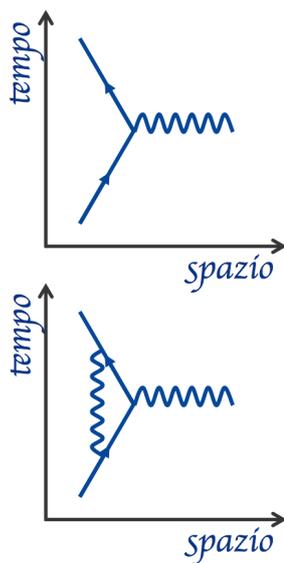


Figura 3: Diagrammi di Feynman. Alto: diagramma di base rappresentante il muone che si propaga nello spazio-tempo deflesso da un fotone del campo magnetico (onda elettro-magnetica). Basso: un fotone virtuale viene emesso e riassorbito, modificando il valore di base di circa l'uno per mille.

La QED

Uno dei contributi più importanti del fisico americano Richard Feynman allo sviluppo della

QED, è stata l'invenzione di un sistema di calcolo dei processi quantistici attraverso un sistema di diagrammi utilizzati normalmente anche ai giorni nostri chiamati *diagrammi di Feynman*⁹. L'interazione di un elettrone o di un muone con il campo elettro-magnetico può essere rappresentata usando questi diagrammi. Useremo qui la versione originale descritta da Feynman, anche se nella versione moderna ci sono piccole differenze. Inoltre utilizzeremo, come particella carica di riferimento, il muone anziché l'elettrone.

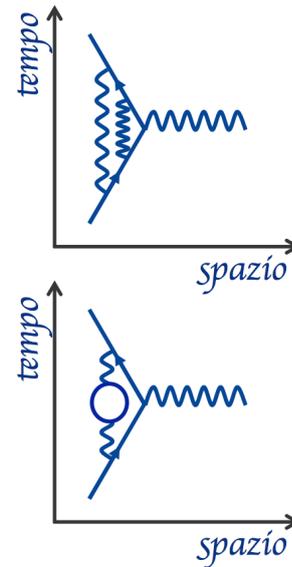


Figura 4: Diagrammi di ordine superiore. Alto: emissione e riassorbimento di due fotoni virtuali. Basso: diagramma (o loop) di polarizzazione del vuoto, nel quale una coppia elettrone-antielettrone è creata e riassorbita.

Il diagramma di base è mostrato in Fig. 3, alto. I due assi rappresentano, rispettivamente, il tempo (asse verticale) e lo spazio (asse orizzontale). Si può pensare allo spazio come ad un semplice spazio unidimensionale; le conclusioni sono le stesse passando allo spazio tridimensionale, mentre il disegno sul foglio risulterebbe alquanto più complicato!

Il muone che si muove da un punto all'altro dello spazio-tempo è rappresentato da una freccia; il campo magnetico interagisce con il muone attraverso un fotone, rappresentato da una

⁹Molto materiale di questo capitolo deriva dalle lezioni di Feynman raccolte nel libro "QED: la strana teoria della luce e della materia" [9]. Il libro non richiede alcuna conoscenza avanzata di matematica e fisica, ma sicuramente un certo impegno nella lettura.

linea ondulata. A seguito di questa interazione, il muone cambia traiettoria ed è rappresentato dalla freccia uscente. Questa rappresentazione è molto più profonda di una semplice illustrazione grafica ed in essa si condensano alcuni concetti tipici della fisica moderna:

- l'onda elettromagnetica e la particella fotone sono la stessa cosa (dualismo onda-particella), per cui la linea ondulata rappresenta allo stesso tempo sia il campo magnetico che il fotone interagente con il muone;
- il concetto di "interazione a distanza", tipico della teoria di campo di fine '800, viene abbandonato e sostituito dallo scambio di un *mediatore di forza*. Nel caso di interazione elettromagnetica il *mediatore di forza* è, appunto, il fotone (=onda elettromagnetica).

Non possiamo qui approfondire ulteriormente il concetto, ma lo spostamento di paradigma è evidente: fra due particelle si esercita una forza se, e solo se, viene scambiato un *mediatore di forza* cioè, di fatto, un'altra particella!

Ma torniamo ai muoni in campo magnetico.

Il diagramma descritto in Fig. 3, alto, corrisponde all'interazione di base, quella descritta anche dalla teoria di Dirac, che prevede il valore $g = 2$. Tuttavia, facendo esperimenti più precisi ed andando quindi a guardare con una lente di ingrandimento quello che succede nel vertice dell'interazione, si osservano contributi come quello descritto nel diagramma di Fig. 3, basso. In questo caso il muone che si propaga emette un fotone il quale è poi riassorbito dal muone stesso. Il fotone non viene rivelato, per cui è chiamato *virtuale*, ma l'energia del muone cambia modificando la sua interazione con il campo magnetico, e quindi il valore di g . Questo effetto è proprio quello misurato nel 1948: $a = 0.00118$.

Si noti quello che è successo: il muone ha *spontaneamente* cambiato la sua energia emettendo un fotone; questa violazione della conservazione dell'energia è durata per il tempo necessario ad interagire con il campo magnetico, ma senza violare il principio di Heisenberg.

Ma non è finita qui: si può andare ancora più in dettaglio ed immaginare effetti ancora più elaborati come quelli descritti in Fig. 4: in alto si

ha l'emissione ed il riassorbimento di due fotoni virtuali, mentre in basso si ha il diagramma di *polarizzazione del vuoto*. In quest'ultimo, una coppia elettrone-antielettrone, o positrone, e^+e^- viene creata dal vuoto, le due particelle poi si annichilano producendo un fotone il quale viene riassorbito. In realtà non è necessario che la coppia particella-antiparticella sia quella formata da elettrone e positrone, tuttavia è intuitivo capire che più la particella è leggera, maggiore è la sua probabilità di venire estratta dal vuoto.

Questi sono effetti tipicamente quantistici, e non hanno nessun equivalente in fisica classica.

Esiste però un ostacolo ulteriore da superare per comprendere appieno il funzionamento della teoria quantistica. Nelle figure precedenti 3 e 4 sono rappresentate quattro possibili interazioni della particella carica con il campo magnetico. Ma se viene isolato un singolo muone è possibile capire quale di queste interazioni sarà effettivamente realizzata? Quel particolare muone interagirà direttamente con il campo, emetterà uno o due fotoni virtuali, oppure creerà una coppia particella-antiparticella polarizzando il vuoto? La risposta è che il muone subirà contemporaneamente *tutte* queste interazioni, così come tutte le altre che è possibile immaginare!

Questo concetto non è facile da accettare ed è decisamente contrario alla nostra intuizione, tuttavia le predizioni basate su questo modello sono state tutte verificate sperimentalmente.

Possiamo considerare una particella che si propaga come un'onda; in effetti il dualismo onda-particella che la prima Meccanica Quantistica aveva affrontato trattando la luce si ripropone anche per le particelle massive come l'elettrone, il muone, ecc. Per questo le particelle sono anche definite *onde di materia* ed il loro stato descritto da una *funzione d'onda*.

Un'onda che attraversa un ostacolo dotato di più fenditure non *sceglie* da quale di esse passare: le attraversa tutte ricombinandosi successivamente dando luogo a fenomeni di interferenza. In maniera analoga, la funzione d'onda che descrive una particella non *sceglie* quale interazione fare: è *come se* un pezzetto di funzione d'onda seguisse un percorso, un altro un percorso diverso. Ognuna di queste "strade", cioè ognuna di queste interazioni, è caratterizzata da una *ampiezza di probabilità*, tecnicamente rappresentata

da un numero complesso; tutte queste ampiezze vanno combinate (sommate) fra di loro, ed è esattamente questo modo di combinare i diversi percorsi che dà luogo ai fenomeni di interferenza osservati.

La meccanica quantistica è decisamente distante dalla nostra esperienza quotidiana!

Lo sviluppo di questa teoria ha inizialmente incontrato alcune difficoltà a causa della proliferazione dei diagrammi possibili, come ad esempio quelli rappresentati in Fig. 4. Alcuni dettagli aggiuntivi sulla strada seguita per superare questo problema sono accennati nel riquadro *Sviluppo perturbativo*.

Limitandoci al caso specifico del momento magnetico anomalo del muone, il calcolo teorico e la misura sperimentale sono state spinte fino a precisioni dell'ordine della parte per milione (ppm = 10^{-6}). I valori più recenti ¹⁰ sono:

$$a_{\mu}^{\text{teo}} = 0.00116591790(65)$$

$$a_{\mu}^{\text{exp}} = 0.00116592080(63) ,$$

dove il numero fra parentesi indica l'errore sulle ultime due cifre. La precisione, sia teorica che sperimentale, è impressionante!

Spesso questi numeri vengono scritti in unità 10^{-11} :

$$a_{\mu}^{\text{teo}} = 116591790 \pm 65$$

$$a_{\mu}^{\text{exp}} = 116592080 \pm 63 .$$

Si può notare una differenza importante ma ancora non così significativa da far parlare di scoperta:

$$\Delta a_{\mu} = a_{\mu}^{\text{exp}} - a_{\mu}^{\text{teo}} = 290 \pm 90 \quad (7)$$

sempre in unità 10^{-11} .

Perché è necessaria una precisione così spinta nella determinazione sia teorica che sperimentale di a_{μ} ? Cosa è possibile imparare, oltre al fatto che esiste uno strumento matematico che funziona correttamente?

La risposta è che queste precisioni sono utili perché all'interno del vuoto quantistico si tro-

va tutto, potenzialmente anche nuove forme di materia! Un esempio di nuova fisica che si può estrarre dal vuoto e di cui, fino ad ora, abbiamo avuto una prova sperimentale solamente dallo studio del moto delle galassie è la Materia Oscura.

Il vuoto e la Materia Oscura

La teoria brevemente accennata nel paragrafo precedente descrive le interazioni elettromagnetiche dei fotoni con elettroni e muoni (in realtà anche con i leptoni τ).

Abbiamo già accennato brevemente all'esistenza dei quark: essi sono caratterizzati dalle cosiddette *interazioni forti* i cui mediatori si chiamano *gluoni* (dall'inglese *glue = colla*).

Infine dobbiamo citare i neutrini che sono sensibili ad un altro tipo di interazione, che si chiama, (anche in questo caso senza grossa fantasia) *interazione debole*, mediata dalle particelle W^{\pm} e Z_0 scoperte da Rubbia nel 1983 [11]. Tale interazione svolge una parte attiva nei processi di fusione nucleare che avvengono all'interno delle stelle e molti decadimenti radioattivi sono dovuti ad essa.

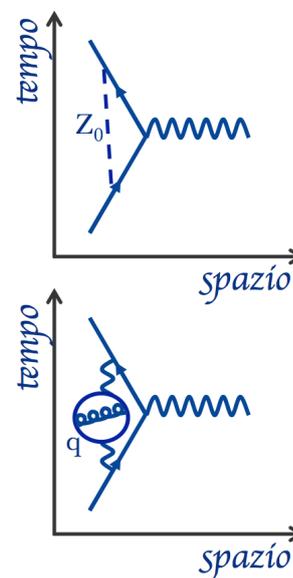


Figura 5: Diagrammi di Feynman che includono uno scambio di un bosone vettore, lo Z_0 nel diagramma in alto, ed un loop di quark con un gluone al centro, nel diagramma in basso.

Così come esistono diagrammi di interazioni nei quali sono coinvolti elettroni e fotoni, se

¹⁰I valori riportati qui ed in seguito sono presi dal seguente articolo (per addetti ai lavori) [10].

Sviluppo perturbativo

Lo sviluppo della QED ha inizialmente comportato alcune difficoltà dal punto di vista teorico. È facile, infatti, immaginare diagrammi sempre più complessi che si possono ottenere continuando ad aggiungere linee di fotoni virtuali e diagrammi di polarizzazione del vuoto ai diagrammi rappresentati nelle figure 3, 4 o 5. In questo modo si possono disegnare infinite “fenditure” attraverso le quali far passare la funzione d’onda che rappresenta la particella. In effetti i primi calcoli teorici ottenevano, per il momento magnetico anomalo del muone, un valore infinito, il che è sbagliato sperimentalmente!

Feynman, Schwinger e Tomonaga risolsero il problema ordinando i diagrammi in base al numero di vertici presenti. Escludendo l’interazione con il fotone del campo magnetico, comune a tutti, il diagramma rappresentato in Fig. 3, alto, ha zero vertici (ordine zero), il diagramma di Fig. 3, basso, ha una coppia di vertici (ordine uno), mentre i due diagrammi di Fig. 4 hanno due coppie di vertici (ordine due). Per ogni coppia di vertici^a il diagramma è soppresso di un fattore pari a circa 1000. Per cui si possono disegnare una serie infinita di percorsi, il cui peso relativo va rapidamente a zero. Questo accorgimento matematico, basato su serie infinite convergenti, è noto come *sviluppo perturbativo* ed ha permesso di ottenere predizioni teoriche molto precise.

^aIn questi diagrammi “chiusi” i vertici sono sempre a coppie.

ne possono immaginare altri che coinvolgono i quark ed i bosoni mediatori delle interazioni (Fig. 5). Le particelle coinvolte in questi diagrammi hanno massa maggiore, talvolta molto maggiore, di quella dell’elettrone, per cui la probabilità, o meglio l’ampiezza di probabilità, di trovarle nel vuoto diminuisce. Infatti l’energia ΔE necessaria a produrre una coppia di particelle all’interno del vuoto quantistico aumenta con la massa della particella stessa, per cui la *vita media* di questo stato deve diminuire per rispettare il vincolo dato dall’equazione di Heisenberg Eq. (2).

Il loro contributo ad a_μ è inferiore rispetto a quello dovuto alla QED. Ad esempio in unità 10^{-11} il contributo dei quark, o *contributo adronico*, vale:

$$a_\mu^{HAD} = 6903 \pm 53 \quad (8)$$

mentre quello dei bosoni vettori, o contributo *elettrodebole (electroweak)*, vale:

$$a_\mu^{EW} = 153 \pm 2 \quad (9)$$

Grazie alle misure sperimentali molto precise, questi contributi sono stati individuati e, entro l’errore sperimentale, confermati.

Ma ci può essere ancora qualcosa in più! Studiando i moti delle galassie esistono prove speri-

mentali dell’esistenza nell’universo di un tipo di materia formata da una o più particelle che non fanno parte di nessuna delle categorie note ed elencate in precedenza. Questa materia, di cui non conosciamo la composizione, forma l’85% della massa dell’universo ed è chiamata *Materia Oscura (Dark Matter)*.

Esiste la probabilità intrigante che la differenza fra i valori di a_μ determinati sperimentalmente e teoricamente sia dovuta proprio a questo tipo di nuova particella. Infatti una tale particella potrebbe essere “estratta dal vuoto” e contribuire al valore sperimentale di a_μ . Naturalmente questo dipende dalla massa e dal modo con cui questa particella interagisce con le altre, ma è possibile immaginare un contributo ad a_μ dovuto alla materia oscura: a_μ^{DM} . Ecco perché è importante misurare sperimentalmente, e determinare teoricamente, con grande accuratezza questa grandezza: la risposta ad un problema legato ad effetti di gravitazione universale si può trovare all’interno del vuoto quantistico!

L’esperimento *g-2* al Fermilab

La fisica richiede la capacità di comprendere i fenomeni e costruire modelli matematici che portino alla formulazioni di predizioni teoriche. L’al-

tro aspetto, altrettanto importante, è la capacità di progettare e costruire esperimenti e di saper interpretare le misure fatte in termini di parametri fondamentali della teoria stessa.

Nell'esperimento $g - 2$ ingegno, capacità e perseveranza hanno permesso di raggiungere precisioni che ci consentono di esplorare l'eventuale contributo della materia oscura al vuoto quantistico.

L'esperimento ha avuto molte fasi nel tempo. Nella sua versione moderna, esso è stato realizzato una prima volta al CERN negli anni '70, poi al Brookhaven National Laboratory di New York, negli anni '90, e adesso una nuova versione è in costruzione presso il laboratorio Fermilab di Chicago. Fermilab è il più grande laboratorio di fisica delle alte energie esistente negli Stati Uniti, equivalente al CERN in Europa. Fondato nel 1967, esso è stato dedicato, così come l'Università di Chicago, alla memoria di Enrico Fermi. Il fisico italiano emigrò nel 1938 proprio a Chicago, subito dopo la consegna del premio Nobel e a seguito della promulgazione in Italia delle leggi razziali. La moglie di Fermi, Laura Capon, di origine ebraica, fu invitata alla inaugurazione del Fermilab nel 1967.

L'esperimento si basa sul moto in campo magnetico uniforme di muoni, cioè di particelle cariche dotate di momento magnetico.

Come è noto dai corsi delle scuole superiori, una particella carica in campo magnetico uniforme con velocità iniziale perpendicolare al campo si muove di moto circolare uniforme. In formule, la particella è sottoposta alla forza di Lorentz

$$\vec{F} = q\vec{v} \times \vec{B} \quad , \quad (10)$$

la circonferenza risultante ha raggio

$$r = \frac{mv}{qB} \quad , \quad (11)$$

e viene percorsa con velocità angolare:

$$\omega_c = \frac{qB}{m} \quad . \quad (12)$$

Si noti che nella trattazione classica la frequenza di rotazione non dipende dall'impulso (cioè dalla velocità) della particella. Questa relazione viene però modificata in condizioni di moto

relativistico dal fattore

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

essendo c la velocità della luce, come segue:

$$\omega_c = \frac{qB}{m\gamma} = \frac{eB}{m\gamma} \quad (13)$$

avendo inserito la carica elementare e per il muone.

Cioè aumentando l'impulso, e quindi γ , la velocità angolare diminuisce. È come se la massa al denominatore venisse accresciuta di un fattore γ , aumentando così l'inerzia del sistema.

Oltre a ciò, il momento magnetico della particella interagisce con il campo magnetico. Si può dimostrare che la particella si comporta come un piccolo giroscopio: il momento magnetico percorre un moto di precessione intorno al campo simile a quello che l'asse di una trottola inclinata percorre intorno alla verticale.

In formule, il campo causa un momento torcente:

$$\vec{M} = \vec{\mu} \times \vec{B} \quad (14)$$

a seguito del quale l'asse magnetico acquista una velocità angolare di precessione è:

$$\omega_s = \frac{g}{2} \cdot \frac{e}{m} \cdot B \quad (15)$$

Anche in questo caso va aggiunta una correzione relativistica all'espressione classica che modifica il risultato precedente come segue:

$$\omega_s = \frac{geB}{2m} + (1 - \gamma) \frac{geB}{m\gamma} = \frac{eB}{m\gamma} + a_\mu \cdot \frac{eB}{m} \quad (16)$$

In definitiva, il vettore spin ruota "un pò più velocemente" del vettore impulso per una quantità proporzionale ad a_μ :

$$\omega_a = \omega_s - \omega_c = a_\mu \cdot \frac{eB}{m} \quad (17)$$

Questo effetto, rappresentato in Fig. 6, è puramente quantistico: se non esistessero i diagrammi virtuali i due vettori procederebbero appaiati.

Invertendo l'equazione (17), si ottiene:

$$a_\mu = \frac{m\omega_a}{eB} \quad (18)$$

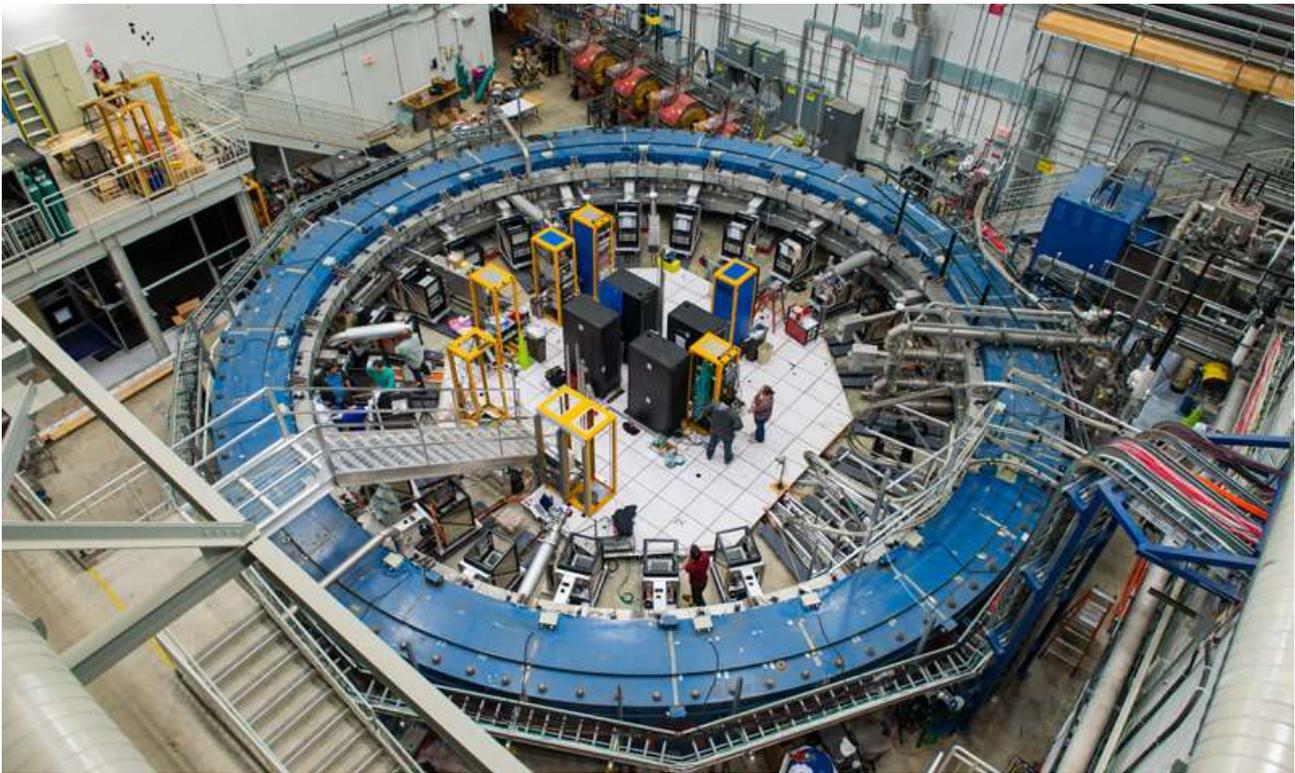


Figura 7: L'anello di accumulazione presso il laboratorio Fermilab di Chicago

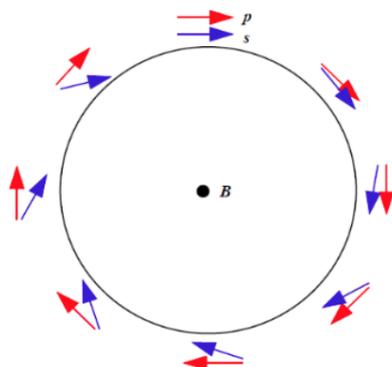


Figura 6: Il vettore impulso è tangente alla traiettoria, mentre lo spin ruota più rapidamente puntando talvolta verso l'esterno, talvolta verso l'interno della stessa.

Sperimentalmente occorre quindi misurare con estrema precisione la velocità angolare relativa ω_a , oltre al campo magnetico \vec{B} .

Va aggiunto che i muoni sono particelle instabili, con una vita media di alcuni microsecondi (a parte effetti relativistici che vedremo in seguito), per cui questa misura non si può fare su un singolo muone, ma costruendo un fascio di molte particelle che vengono studiate statisticamente.

Un esperimento ideale può essere pensato nel modo seguente:

1. produrre un numero sufficiente di muoni aventi spin parallelo o, come si dice in gergo, *polarizzati*;
2. far muovere questi muoni in un *anello di accumulazione* immerso in un campo magnetico uniforme forzandoli a rimanere confinati all'interno dell'anello;
3. misurare accuratamente il campo magnetico, con le sue inevitabili variazioni;
4. misurare, in funzione del tempo, la direzione dello spin rispetto alla direzione di volo, cioè l'angolo fra i vettori \vec{s} e \vec{p} .

Tutte queste condizioni sono state realizzate nella serie di esperimenti già ricordati in precedenza. In Fig. 7 si può vedere l'anello di accumulazione attualmente installato al Fermilab di Chicago nel quale i muoni sono confinati percorrendolo con velocità prossima a quella della luce¹¹. I muoni sono immessi nell'anello dal sistema di acceleratori del laboratorio, entrando

¹¹Per la precisione, i muoni hanno un impulso $p = 3.1$ GeV/c.

dalla linea di fascio che si può vedere al centro ed in alto in Fig. 7. Un campo magnetico di 1.5 T li forza a percorrere l'anello all'interno del quale è fatto il vuoto¹² a circa 10^{-7} atmosfere. A causa dell'espansione relativistica, la vita media dei muoni è di $64 \mu\text{s}$, invece dei $2.2 \mu\text{s}$ di un muone a riposo e l'anello viene percorso alcune centinaia di volte: un po' come se un essere umano vivesse 3000 anni!

Ma come si riescono a produrre muoni polarizzati e come si misura la direzione dello spin in funzione del tempo? Per questo ci aiutano la natura e le proprietà particolari dei muoni.

Senza entrare troppo in tecnicismi, il muone è prodotto in un processo a due stadi nel quale un fascio di protoni mandati contro un cilindro di Tungsteno magnetizzato usato come bersaglio producono alcune particelle intermedie, chiamate mesoni π , che decadono in un centesimo di microsecondo in muoni. Questo decadimento ha la proprietà di produrre un muone avente lo spin orientato di preferenza lungo il proprio impulso. Quindi, selezionando muoni che hanno una specifica direzione di volo, e quindi una specifica direzione dell'impulso, gli spin risultano paralleli.

Anche le proprietà di decadimento del muone sono particolari: esso decade in un elettrone e due neutrini, ma l'elettrone è emesso di preferenza nella direzione opposta allo spin.

Quindi andando a contare gli elettroni¹³ emessi in una certa direzione, ad esempio verso l'interno dell'anello, si ha un'informazione della direzione dello spin in quell'istante.

Un esempio di tale conteggio in funzione del tempo, riportato nell'articolo [3], è rappresentato in Fig. 8.

In questa figura ogni banda è la continuazione di quella superiore; esse sono separate solamente per renderle più visibili. Congiungendo le bande consecutivamente, si ottiene una sola funzione distribuita su un tempo di 600 microsecondi, circa 10 volte la vita media della particella.

La distribuzione misurata rappresenta la sovrapposizione di due processi fisici. I muoni inse-

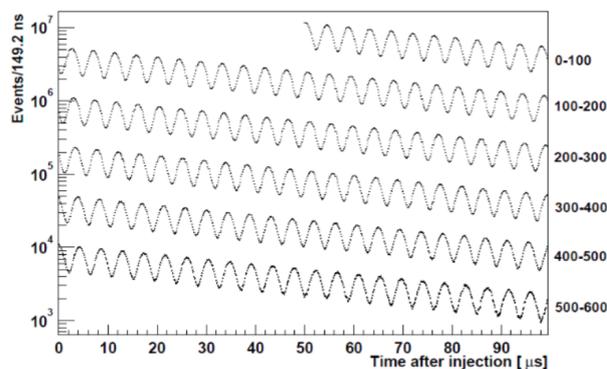


Figura 8: Numero di positroni, provenienti dal decadimento dei muoni del fascio, in funzione del tempo, misurato a partire dalla loro iniezione nell'anello.

riti nell'anello di accumulazione decadono in un elettrone e due neutrini, questi ultimi invisibili al rivelatore, con il tipico andamento esponenziale seguito dai processi radioattivi. Per questo il conteggio di elettroni osservati diminuisce con lo scorrere del tempo. Oltre a ciò si osserva l'effetto che ci interessa: il numero di elettroni emessi è massimo quando il vettore spin punta verso l'esterno (si ricordi che gli elettroni sono emessi "all'indietro", cioè in direzione opposta allo spin) ed è minimo quando il vettore spin punta verso l'interno della circonferenza. Quindi si osserva una oscillazione di pulsazione uguale proprio alla velocità angolare relativa ω_a descritta in Eq. (17). In definitiva, una tale oscillazione è proprio la rappresentazione visiva della complessa struttura del vuoto quantistico!

L'esperimento appena descritto raccoglierà dati nel triennio 2018-2020 con lo scopo di ridurre l'errore sperimentale su a_μ (vedi Eq.(7)) di un fattore 4 e di verificare se la discrepanza fra teoria ed esperimento in Eq.(7) risulti confermata dalla nuova misura con una maggiore significatività. Se così fosse, avremmo la prova che il Modello Standard non è completo e che nel vuoto quantistico si nascondono nuove forme di materia.

Conclusioni

Lo studio dei fenomeni microscopici ci ha condotti ad una descrizione della realtà molto diversa, ed in apparente contrasto, con quella che verifichiamo con i nostri sensi. Nel corso del-

¹²Le particelle devono muoversi nel vuoto o la traiettoria sarebbe continuamente deviata dagli urti microscopici

¹³In realtà l'esperimento usa muoni di carica positiva, per cui le particelle prodotte sono elettroni di carica positiva, o positroni, e questi vengono emessi nella direzione verso la quale punta lo spin.

l'ultimo secolo questa descrizione si è perfezionata dalla Meccanica Quantistica all'Elettrodinamica Quantistica fino a quello che oggi chiamiamo il Modello Standard delle Interazioni Fondamentali.

La nuova concezione di *vuoto* è forse la caratteristica che più si allontana dalla nostra esperienza: all'interno del vuoto quantistico si possono trovare particelle, luce e forze ancora sconosciute. E secondo alcune teorie, con fluttuazioni opportune, anche se altamente improbabili, si può trovare anche un intero universo!



- [1] R. L. Garwin, L. Lederman, M. Weinrich, Phys. Rev. 105 (1957) 1415
- [2] J. Bailey, *et al*, Nuovo Cimento A9 (1972) 369
- [3] G. Bennet, *et al*, Phys. Rev. D73(2006) 072003
- [4] C. Anderson, S. Neddermeyer, Phys. Rev. 51 (1937) 884
- [5] M. Perl, *et al*, Phys. Rev. Lett. 35 (1975) 22
- [6] P. Dirac, Proc. Roy. Soc. A117 (1928) 610
- [7] P. Kusch, H. M. Foley, Phys. Rev. 73 (1948) 421
- [8] J. Schwinger, Phys. Rev. 73 (1948) 416
- [9] R. Feynman, *QED: la strana teoria della luce*, ed. Adelphi (2010)
- [10] F. Jegerlhener, A. Nyffler, *The Muon g-2*, arXiv:0902.3360v1
- [11] Carlo Rubbia, *Nobel Lecture 1984*, <https://www.nobelprize.org>
- [12] si veda la pagina web dell'esperimento: <http://muon-g-2.fnal.gov>



Marco Incagli: È Primo Ricercatore presso la sezione di Pisa dell'Istituto Nazionale di Fisica Nucleare. Ha svolto attività di docenza nei corsi di *Interazioni radiazione-materia*, *Particelle Elementari*, *Astroparticelle* e *Fisica 1* al Corso di Studi in Fisica dell'Università di Pisa. Ha lavorato per quindici anni nel campo della ricerca di Materia Oscura nello Spazio partecipando all'esperimento internazionale AMS02, installato sulla Stazione Spaziale Internazionale, essendo il responsabile di una parte dell'apparato (calorimetro elettromagnetico). Recentemente si è dedicato alla ricerca di nuova fisica in laboratorio nell'ambito dell'esperimento g-2 in costruzione al Fermilab (Chicago).

All'origine della massa: la QCD e il suo vuoto

Luca Girlanda

Dipartimento di Matematica e Fisica "E. De Giorgi", Università del Salento e INFN sezione di Lecce

La Cromodinamica Quantistica (QCD) è la moderna teoria delle interazioni forti tra i costituenti fondamentali della materia, il cui stato fondamentale, o stato di minima energia, si chiama tecnicamente vuoto. Le sue complesse proprietà determinano per più del 95% l'intera massa visibile dell'Universo. In questo contributo si illustrano le ragioni teoriche e fenomenologiche che hanno portato alla formulazione della QCD, evidenziando in particolare il meccanismo di generazione della massa.

Campi e particelle

Tra le maggiori acquisizioni della fisica del '900, riassuntiva, a giudizio di Feynman, di tutta la conoscenza scientifica, vi è la concezione atomistica della materia [1], che tutto cioè sia riconducibile a particelle elementari in reciproca interazione. Lungi dal costituire una posizione puramente filosofica, questa può invece ritenersi una vera e propria conquista sperimentale. Se infatti alla fine dell'800 si poteva ancora dibattere sulla realtà delle molecole, il dibattito si chiuse con la determinazione sperimentale del numero di Avogadro, che equivale a pesare le molecole stesse, una impresa che valse al francese Perrin il premio Nobel nel 1926, e che vide un significativo contributo nella tesi di dottorato di Einstein "Una nuova determinazione delle dimensioni molecolari"

[2].

Anche un concetto fisico antitetico quale quello di campo assume, nel quadro della teoria quantistica, un carattere discreto, una sorta di collezione di particelle localizzate. Benchè sia vero anche il viceversa, che cioè le particelle materiali siano descritte nella meccanica quantistica da funzioni estese, le funzioni d'onda.

Così come la meccanica quantistica postula l'esistenza di una granularità ultima, che Dirac eleva addirittura a requisito filosofico [3], la relatività conduce invece naturalmente alla prevalenza del concetto di campo. Essa infatti assume l'esistenza di un limite invalicabile alla velocità di propagazione delle interazioni dato dalla velocità della luce c . Queste non hanno il carattere di azione a distanza, come invece ipotizzato, seppur a malincuore, da Newton: per calcolare la forza agente su una particella bisogna conoscere la posizione delle altre particelle non ad uno stesso istante, ma ad istanti precedenti diversi, a seconda delle rispettive distanze. Risulta molto più conveniente introdurre la nozione di campo, che riassume l'effetto congiunto di tutte le altre particelle ed è definito in modo da fornire, ad ogni istante e in ogni punto, la forza agente su una particella che si trova in quel punto a quell'istante. Del resto la nozione di campo elettromagnetico aveva già con Maxwell e Faraday assunto connotati di vero e proprio ente fisico. I campi, che riempiono lo spazio e il tempo, diventano quindi elementi primari della realtà fisica.

La sintesi più naturale della meccanica quantistica e della relatività speciale è quindi la teoria quantistica dei campi. Nel processo di quantizzazione il campo diventa una combinazione di eccitazioni elementari quantizzate, i fotoni nel caso del campo elettromagnetico, così come la configurazione di una corda vibrante può essere vista come sovrapposizione di componenti con diverse lunghezze d'onda. Reciprocamente, anche le particelle elementari possono essere considerate come eccitazioni di altrettanti campi quantistici. Tutti gli elettroni dell'universo ad esempio, possono essere considerati come eccitazioni di un unico campo elettronico che pervade lo spazio e il tempo.

Vi è un'importante conseguenza dei principi della meccanica quantistica: così come il principio di indeterminazione di Heisenberg vieta di realizzare una situazione in cui una particella sia ferma in un punto, così una situazione di "vuoto" in cui il campo sia nullo e resti tale non è pensabile quantisticamente. Vi sono invece continuamente delle "fluttuazioni quantistiche" attorno a una configurazione di campo nullo. La configurazione di assenza di particelle è solo apparente. Continuamente emergono particelle cosiddette virtuali, insieme alle corrispondenti antiparticelle, in modo da conservare numeri quantici come la carica elettrica. Queste coppie però si riannichilano dopo brevissimo tempo, da cui deriva il termine "virtuali". Il loro apparire viola la conservazione dell'energia, perché richiede una certa energia ΔE , ma solo per un tempo $\Delta t \sim \hbar/\Delta E$ brevissimo, in accordo appunto col principio di Heisenberg¹. E proprio queste fluttuazioni conferiscono al "vuoto" (che tecnicamente indica lo stato di minima energia) caratteristiche ben precise e misurabili.

In tal modo il vuoto quantistico si allontana moltissimo dalla visione classica di "palcoscenico degli eventi", ma assume invece un ruolo da primo attore. Dalle sue proprietà dipendono quelle di tutti gli altri stati descritti da una data

¹David Griffiths [4] avverte i suoi lettori di tenere stretto il portafogli quando un fisico chiama in causa il principio di indeterminazione di Heisenberg. È vero che la descrizione dei fenomeni quantistici presuppone un consistente formalismo matematico, ma è altrettanto vero che anche queste definizioni sono fondate su presupposti classici, almeno nell'ortodossa interpretazione di Copenhagen [5].

teoria quantistica dei campi. Per la Cromodinamica Quantistica (QCD) portano in particolare all'emergere della massa dei costituenti del nucleo, rendendo in tal modo conto della gran parte della massa (visibile) dell'universo e di noi stessi. Il vuoto della QCD è dunque un mezzo dinamico altamente complesso, che modella il mondo con la sua attività. Ma per illustrare queste caratteristiche è necessario prima descrivere la moderna teoria delle interazioni forti [6, 7, 8].

Preistoria della QCD

La storia della ricerca dell'interazione forte, responsabile dell'esistenza e stabilità dei nuclei atomici, nasce con la scoperta del nucleo e dei suoi costituenti, protoni e neutroni, all'inizio degli anni '30 del secolo scorso. Il carattere nascosto di questa forza dipende dal suo piccolissimo raggio d'azione limitato alle dimensioni dei nuclei atomici, dell'ordine di 10^{-15} m, unità di lunghezza intitolata a Fermi. Inizialmente il legame tra questi costituenti veniva imputato allo scambio di particelle mediatrici, nello stesso senso in cui i fotoni sono i mediatori della forza elettromagnetica tra particelle cariche. Il raggio d'azione della forza indotta dallo scambio delle particelle mediatrici è tanto minore quanto maggiore è la loro massa m . Infatti l'emissione delle particelle mediatrici viola la conservazione dell'energia di una quantità $\Delta E \sim mc^2$. Come già visto ciò è consentito solo nei limiti previsti dal principio di Heisenberg, cioè per un tempo $\Delta t \sim \hbar/\Delta E \sim \hbar/(mc^2)$, dopodiché dovrà aversi il riassorbimento da parte dell'altra particella in interazione. In questo tempo l'interazione, per i limiti imposti dalla relatività non potrà essersi estesa a distanze maggiori di $r \sim c\Delta t \sim \hbar/(mc)$. Se l'interazione elettromagnetica, dovuta allo scambio dei fotoni, si estende a lungo raggio, in virtù del fatto che questi hanno massa nulla, così l'interazione forte dovrà essere mediata da particelle mediatrici di massa $m \sim \hbar/(rc) \sim 200 \text{ MeV}/c^2$. Fu il fisico giapponese Yukawa a proporre per primo questa idea nel 1934 e le particelle mediatrici furono poi effettivamente scoperte nel 1947 e chiamate mesoni π o pioni, con una massa pari a circa $140 \text{ MeV}/c^2$.

I tentativi però di descrivere l'interazione forte sul modello dell'interazione elettromagnetica, che nella sua versione quantistica mieteva i pri-

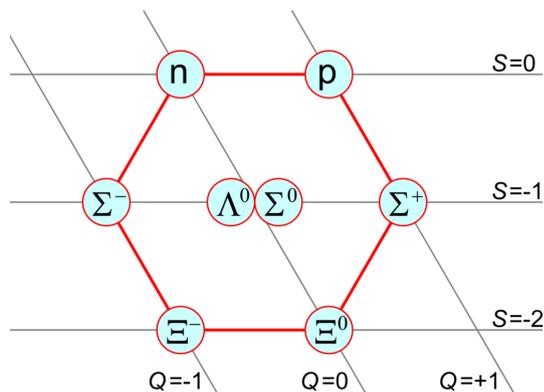


Figura 1: L'ottetto barionico. Le particelle più leggere con spin $1/2$ ordinate secondo i numeri quantici di stranezza S e di carica elettrica Q . Con p e n si indicano protone e neutrone rispettivamente.

mi formidabili successi, furono però infruttuosi. Tanto più che nel corso degli anni successivi vi fu un fiorire di scoperte di nuovi "adroni", particelle cioè soggette alle interazioni forti. Si dovette dunque abbandonare l'idea che queste particelle fossero tutte "elementari". Piuttosto, venne alla luce una struttura, l'organizzazione di questi adroni in gruppi o "multipletti" di massa simile. Questo condusse all'idea che gli adroni fossero a loro volta costituiti da entità più fondamentali. Così come, infatti, le regolarità nelle proprietà degli elementi chimici, riassunte dalla tabella periodica di Mendeleev, sono spiegate dall'esistenza di medesimi costituenti subatomici combinati diversamente, anche per gli adroni vale lo stesso argomento. In questo caso le proprietà a cui si fa riferimento non sono naturalmente quelle chimiche, ma quelle che costituiscono la carta di identità delle particelle, ossia la massa e lo spin. Chi avanzò e portò avanti questa idea fu soprattutto Gell-Mann, che introdusse il termine "quarks" per queste entità più fondamentali, insieme a Ne'eman e a Zweig, che invece li chiamava "assi".

Ricorrevano multipletti di adroni costituiti da 8 membri e questo suggeriva che il numero di costituenti fondamentali fosse pari a 3 ("Three quarks for Muster Mark" è l'esclamazione che apre il capitolo 4 dell'ultima onirica opera di Joyce, "Finnegans Wake"), nell'ambito di una descrizione matematica che Gell-Mann battezzò "l'ottupla via", parafrasando la via buddista al nirvana. Si interpretò allora questi ottupletti come costituiti da due di queste entità fonda-

tali equivalenti, ciascuna presente in 3 diverse specie, successivamente chiamati "sapori". L'equivalenza indica la possibilità di mescolare tra loro i sapori senza cambiare la dinamica dei sistemi di cui fanno parte. Da qui la possibilità di 9 diverse combinazioni, otto delle quali si trasformano tra di loro a seguito del mescolamento dei sapori, mentre una rimane invariata, analogamente a quanto avviene per il prodotto scalare tra due vettori sotto rotazioni. Matematicamente il gruppo di simmetria è $SU(3)$ e l'ottetto è quella che si chiama una delle sue rappresentazioni irriducibili. Si hanno anche combinazioni di 3 quarks che possono dar luogo a multipletti formati da 8 o anche 10 membri, un'altra rappresentazione irriducibile dello stesso gruppo di simmetria. Queste combinazioni si presentano con spin semi-intero, corrispondenti a particelle che si chiamano "fermioni", contrariamente a quelle formate da due quarks (più precisamente da quark e antiquark). Da questo si deduce che i quarks siano essi stessi fermioni. Laddove sembravano mancare particelle per completare i multipletti, queste venivano sistematicamente scoperte, come nel caso della particella Ω , la più pesante del decupletto più leggero.

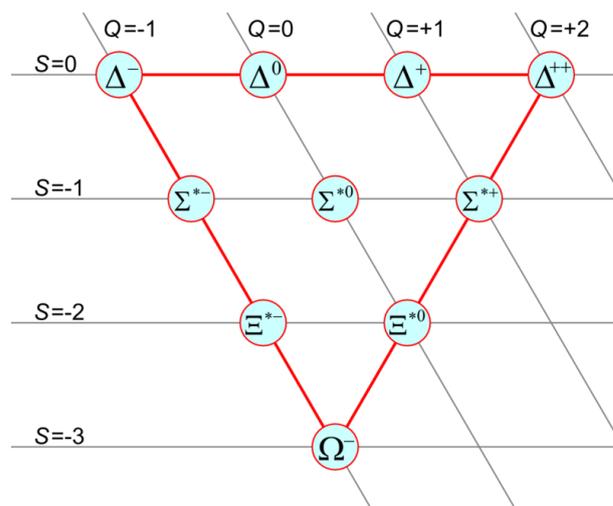


Figura 2: Il decupletto barionico. Le particelle più leggere con spin $3/2$ ordinate secondo i numeri quantici di stranezza S e di ipercarica Q .

Il modello a quarks guadagnava dunque negli anni '60 enorme credibilità, tuttavia la sua maggiore difficoltà era il fatto che tra le tante particelle scoperte, non vi era nessun quark! Tanto più che una loro individuazione sarebbe stata facilitata dal fatto che la loro carica elettrica era

prevista dal modello essere frazionaria rispetto alla carica dell'elettrone. Vi era chi considerava i quarks dei meri costrutti matematici, utili solo per descrivere le proprietà di simmetria degli adroni, e proponeva di trattare questi ultimi come tutti elementari, democraticamente. Vi era anche chi invece cercava di costruire modelli di interazione tra i quarks, ipotizzandoli confinati in una sacca ("bag models") o come estremità di stringhe le quali, spezzandosi, dessero origine a nuove coppie di quark-antiquark, similmente a come avviene per i poli di un magnete. Ma il meccanismo del confinamento e la natura di queste stringhe non erano affatto compresi.

Un'altra difficoltà era l'esistenza di adroni che il modello a quarks descriveva come costituiti da 3 quarks identici, tutti nello stesso stato quantistico e con spin allineato nella stessa direzione; e questo è vietato dal principio di esclusione di Pauli, cui i quarks, in quanto fermioni, dovrebbero invece obbedire. Per rimediare a queste difficoltà venne introdotto un nuovo numero quantico, che Fritsch e Gell-Mann (un vero linguista!) chiamarono "colore". Servono 3 colori per distinguere 3 quarks altrimenti identici, come i colori fondamentali rosso, verde e blu che colorano i pixel degli schermi LCD. Inoltre l'ipotesi del confinamento poteva essere riassunta nella richiesta che gli adroni siano bianchi, come risultante dalla accensione simultanea dei 3 colori in un pixel, oppure dalla presenza simultanea di un colore e del suo complementare (o "anticolore"), come accade nei mesoni, costituiti da una coppia di quark e antiquark.

Matematicamente si tratta di un altro gruppo $SU(3)$, detto $SU(3)$ di colore, rispetto al quale gli adroni sono supposti costituire la rappresentazione di singoletto, cioè quella invariante. I tre quarks di Muster Mark sono quindi in realtà 3×3 . Le tre varietà che mescolandosi danno origine all'ottupla via vennero chiamate sapori, in assonanza col numero quantico di colore appena introdotto. Davvero provvidenziale, il numero quantico di colore, per salvare la dotta citazione di Gell-Mann, o forse la profetica filastrocca di Joyce, in quanto di sapori se ne scoprirono altri, a partire dagli anni '70 fino alla scoperta del sesto sapore, il "top" nel 1994 al Fermilab. Naturalmente si parla sempre di manifestazioni alquanto indirette di questi nuovi sapori, dato

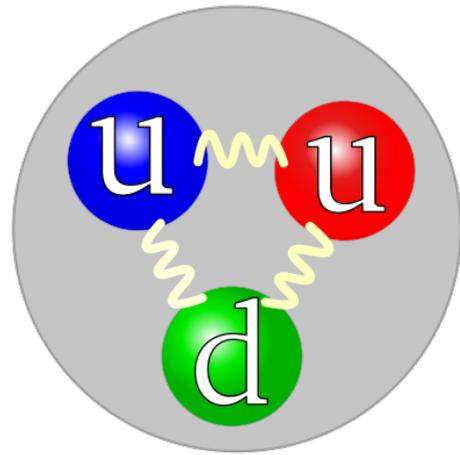


Figura 3: Nel modello a quark ingenuo i barioni, come il protone qui raffigurato, sono composti da tre quark.

che i singoli quarks restano invisibili, per l'ipotesi del confinamento.

Negli stessi anni, nuove indirette evidenze in favore di una sottostruttura degli adroni venivano dagli esperimenti di scattering profondamente inelastico di elettroni da nuclei, condotti a SLAC, l'acceleratore lineare di Stanford. Con questo termine si indicano processi di diffusione in cui parte dell'energia cinetica iniziale viene spesa per creare nuove particelle. Nel caso elastico l'energia con la quale l'elettrone emerge è strettamente correlata con l'angolo di deflessione. Nel caso inelastico invece queste due variabili cinematiche sono scorrelate. Tuttavia risultava da questi esperimenti la dipendenza da un'unica combinazione di tali variabili, un fenomeno chiamato "scaling di Bjorken" dal nome del fisico che l'ha proposto, sulla base dell'ipotesi che gli adroni fossero composti da costituenti puntiformi. A dare compiutezza all'idea fu il modello a "partoni" proposto da Feynman.

Secondo questo modello la diffusione sull'adrone avviene di fatto mediante un processo d'urto elastico che coinvolge l'elettrone e uno solo dei costituenti puntiformi dell'adrone alla volta, costituenti che vennero chiamati appunto partoni. Come tutti in tutti i fenomeni quantistici gli esiti di un esperimento hanno carattere probabilistico. Tali esiti sono descrivibili dall'ipotesi

che i partoni abbiano una certa distribuzione di massa nell'adrone. Nemmeno il loro numero è determinato, l'unica richiesta è che la somma delle loro energie coincida con quella dell'adrone in esame. Dato che in un processo d'urto elastico l'energia dell'elettrone diffuso ad un fissato angolo dipende dalla massa del centro diffusore, la distribuzione in energia dell'elettrone fornisce informazioni sulla distribuzione di massa dei partoni nell'adrone. Il maggiore successo del modello a partoni consiste nel fatto che le distribuzioni partoniche così ottenute sono distintive di ogni adrone e possono essere utilizzate in altri tipi di esperimenti.

L'identificazione dei partoni di Feynman con i quarks di Gell-Mann era tuttavia alquanto difficile, non foss'altro che per il fatto che, come già sottolineato, il numero di partoni non è fissato, e che inoltre alcuni di questi partoni risultano elettricamente neutri, in contrasto con il modello a quarks. Inoltre i partoni, per ipotesi, sono tra loro indipendenti e non-interagenti, al contrario dei quarks, che sono invece fortemente legati. I diversi modelli che abbiamo discusso catturano ciascuno un aspetto dell'interazione forte, un'interazione così forte da tenere confinati i costituenti elementari, ma che alle piccolissime distanze, esplorate negli esperimenti a SLAC, diventa evanescente, così da lasciar apparire partoni non-interagenti. Un tale comportamento apparentemente contraddittorio è proprio di una classe di teorie, dette teorie di gauge non abeliane, introdotte nel 1954 da Yang e Mills. Come scoperto nel 1972-73 da 't Hooft e da Gross, Wilczek e Politzer, tali teorie manifestano proprio questa proprietà, che venne chiamata della "libertà asintotica". Negli stessi anni la Cromodinamica Quantistica venne formulata, ad opera di Gell-Mann, Fritzsche e Leutwyler, interpretando il gruppo $SU(3)$ di colore come un gruppo di gauge, chiamando così in causa, oltre ai quarks, anche 8 gluoni, mediatori dell'interazione tra i quarks.

Il principio di gauge e la QCD

L'esempio più semplice di teoria di gauge è dato dall'elettrodinamica, che nella versione quantistica prende il nome di QED. Il tratto caratteristico di queste teorie risiede nel fatto che una richiesta

di simmetria della teoria porta necessariamente all'introduzione di interazioni con ben precise proprietà. Nel caso della QED questa simmetria consiste nel richiedere che per il campo che descrive gli elettroni, che è un campo complesso, sia possibile scegliere arbitrariamente la fase in ogni punto dello spazio-tempo, senza modificare le equazioni della teoria. È come se richiedessimo che i calciatori di un biliardino possano avere orientazioni diverse tra loro sulle stecche, e tuttavia continuare a costituire una squadra, cooperando nei passaggi per segnare un gol. Affinché questo sia possibile è necessario tener conto delle diverse orientazioni dei calciatori. La prescrizione che codifica le diverse orientazioni tra un punto e un altro vicino è espressa matematicamente dal campo di gauge, che viene anche chiamato, per questo motivo, campo di "connessione". Tornando alla QED, il campo di gauge è quello che descrive i fotoni. Le caratteristiche dei fotoni, il fatto di essere particelle di massa nulla e spin 1, sono tutte conseguenze dell'iniziale richiesta di invarianza rispetto a ridefinizioni della fase indipendenti in ogni punto dello spazio-tempo. È in questo senso che il principio di gauge "genera" l'interazione elettromagnetica. Altra conseguenza dell'invarianza è una legge di conservazione, che corrisponde in questo caso alla conservazione della carica elettrica.

Il termine invarianza di gauge si potrebbe tradurre invarianza per "ricalibrazione", e fu coniato da Weyl nel suo tentativo di unificare gravità ed elettromagnetismo. L'invarianza cui Weyl faceva riferimento era infatti la libertà di ridefinire in ogni punto dello spazio-tempo la distanza tra due eventi vicini. La reinterpretazione dell'idea di Weyl nel senso suddetto di invarianza per ridefinizione della fase quantistica fu invece dovuta a London.

La QCD può considerarsi una generalizzazione della QED. Nell'analogia del biliardino, è come se permettessimo ai giocatori non solo di cambiare indipendentemente orientazione rispetto all'asse della stecca, ma anche di ruotare attorno ad altri assi. Dato che la composizione di due generiche rotazioni dipende dall'ordine in cui si effettuano, il corrispondente gruppo di trasformazioni si dice "non-abeliano" e, corrispondentemente, la struttura della connessione di gauge è più complicata. La trasformazione

in questione, per la QCD, è quella che mescola i colori: il campo dei quarks può cambiare colore indipendentemente in ogni punto dello spazio-tempo, senza modificare le equazioni della teoria, purchè si introduca un campo di gauge con ben precise proprietà, l'aver cioè le proprie eccitazioni quantistiche, i gluoni, massa nulla e spin 1, analogamente a quanto avviene per il fotone. Così come la simmetria della QED conduce alla conservazione della carica elettrica, in QCD si ha la conservazione del numero quantico "colore". I gluoni trasportano il colore da un punto all'altro dello spazio-tempo, cambiando quello dei quarks nei punti corrispondenti. Così, ad esempio, un quark rosso può trasformarsi in uno verde emettendo un gluone che trasporta numeri quantici (rosso-antiverde). Vi sono quindi 8 gluoni diversamente colorati, pari alle 9 possibili combinazioni di colore-anticolore tranne una, corrispondente alla combinazione "bianca", in cui i tre colori e i tre anticolori sono presenti in ugual modo.

Una caratteristica distintiva della QCD rispetto alla QED, che dipende dal carattere non-abeliano della simmetria di gauge, è il fatto che i gluoni sono carichi, possiedono cioè definite caratteristiche di colore, contrariamente ai fotoni, che sono elettricamente neutri. Questo porta alla possibilità che essi interagiscano direttamente tra di loro. Ad esempio un gluone (rosso-antiverde) può trasformarsi in un gluone (rosso-antiblu) emettendo un gluone (blu-antiverde). Un processo del genere, in cui un fotone si trasforma in due, non è invece possibile in QED. È proprio questa caratteristica che conduce alle radicali e spettacolari differenze esistenti tra le due teorie.

La più sorprendente e fondamentale è il cosiddetto fenomeno dell'antiscreening, o antischermatura. Si può comprendere che la continua, seppur effimera, esistenza di coppie virtuali particella-antiparticella porti ad un effetto di schermatura di una carica elementare nel vuoto, e questo è infatti ciò che accade in QED, analogamente al fenomeno della polarizzazione di un mezzo materiale. Si parla in questo contesto di polarizzazione del vuoto: una carica positiva attrarrà le particelle virtuali cariche negativamente e respingerà quelle cariche positivamente, cosicché la carica effettiva diminuisce con la distanza. In QCD, oltre a questo effetto di schermatura di

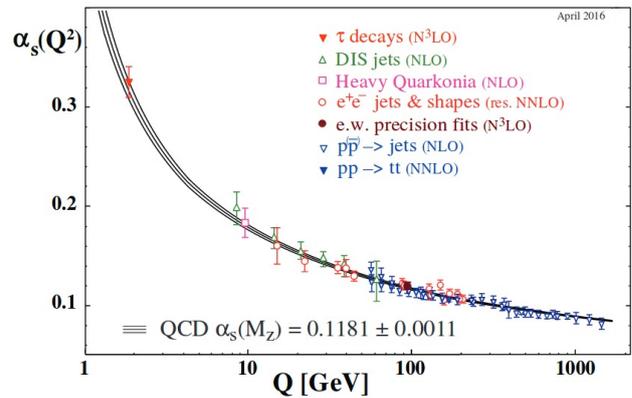


Figura 4: Variazione della carica effettiva di colore α_s con la scala di energia. All'aumentare di Q si esplorano distanze sempre più piccole, alle quali l'interazione forte diminuisce. Viceversa, per piccoli Q l'interazione cresce, portando la teoria nel cosiddetto regime "non-perturbativo", in cui $\alpha_s \sim 1$. I dati sperimentali seguono perfettamente l'andamento previsto dalla libertà asintotica. Tratta dalla Ref. [9].

una singola carica di colore da parte delle coppie virtuali quark-antiquark, si ha anche un effetto contrario, dovuto al contributo dei gluoni e alle loro mutue interazioni, un effetto assente in QED per quanto visto.

Il fenomeno dell'antiscreening porta al fatto che la carica di colore sia amplificata dalle fluttuazioni del vuoto. La nuvola di particelle virtuali che circonda una carica di colore ne potenzia l'effetto, in misura tanto maggiore quanto più ce ne si allontana. Viceversa, indagando sempre più all'interno di questa nuvola, mediante esperimenti di alta energia, l'effetto della carica di colore diminuisce, fino ad essere irrilevante. Nella Fig. 4 è illustrato l'andamento della carica effettiva di colore con la scala d'energia Q , così come previsto dalle equazioni della QCD e confrontata con una pluralità di risultati sperimentali [9].

È questo il meccanismo che tiene insieme la fenomenologia del modello a quarks e del modello a partoni. Inoltre lo stesso meccanismo rende plausibile l'ipotesi del confinamento: un singolo quark colorato crea una perturbazione nella configurazione dei campi circostanti che si estende sempre più intensamente con la distanza. Ciò porterebbe ad un costo energetico enorme, ed è questo costo ad impedire l'esistenza di cariche di colore isolate. Un modo per cancellare queste perturbazioni è di considerare insieme al quark

il corrispondente antiquark nello stesso punto. Ma localizzare con questa precisione una particella comporta, per il principio di Heisenberg, una grande dispersione dei valori del suo impulso e quindi dell'energia cinetica. Il risparmio sull'energia della nuvola di particelle virtuali si paga quindi in termini di energia cinetica delle sorgenti di colore. Da questo bilancio scaturisce la massa dell'adrone, e come corollario la giustificazione dell'ipotesi del confinamento.

Una dimostrazione matematicamente rigorosa dell'origine del confinamento è un problema ancora aperto, dopo quasi mezzo secolo di ricerche, e costituisce uno dei Problemi del Millennio, che il Clay Mathematics Institute di Cambridge premia con 1 milione di dollari. Tuttavia esistono oggi solide ragioni per ritenere il confinamento una conseguenza diretta delle equazioni della QCD, basate su evidenze numeriche provenienti da simulazioni su supercalcolatori. Quest'ultimo è l'unico approccio effettivamente percorribile, in quanto la dinamica è non-perturbativa, contrariamente al caso della QED. Infatti, alla scala a cui avviene la generazione della massa, le interazioni sono forti. Tanto forti appunto da confinare i costituenti fondamentali e dar vita ad altri gradi di libertà compositi.

Tutto da un numero

L'aspetto particolarmente affascinante delle teorie di gauge, e della QCD in particolare, è il fatto che, fissato il tipo di trasformazione dei campi elementari sotto il gruppo di simmetria, la loro forma è essenzialmente determinata. Tutto quanto è fissato in termini di un'unica costante adimensionale, g e delle masse m_f dei diversi sapori di quarks. Vi è in realtà un secondo parametro adimensionale, il cosiddetto angolo θ_{QCD} , che è legato a un altro aspetto della struttura del vuoto, di tipo topologico, cui però non si farà altro cenno. Detto parametro peraltro, per ragioni tuttora inspiegate (e questo costituisce il cosiddetto problema di *CP*-forte) risulta fenomenologicamente essere praticamente nullo. Non ci sono altri parametri, o almeno sono così piccoli da essere irrilevanti, per ragioni, queste sì, comprese. Gli $N_f = 6$ sapori di quarks hanno masse molto diverse tra loro: i tre più leggeri, detti up, down e strange, sono molto più leggeri dei tre

più pesanti, detti charm, bottom e top. Questa situazione è molto vicina a quello che Leutwyler chiama il "paradiso del teorico" [10], cioè una situazione in cui i 3 quarks leggeri siano di massa nulla, mentre quelli pesanti siano di massa infinita, in modo da non giocare più alcun ruolo.

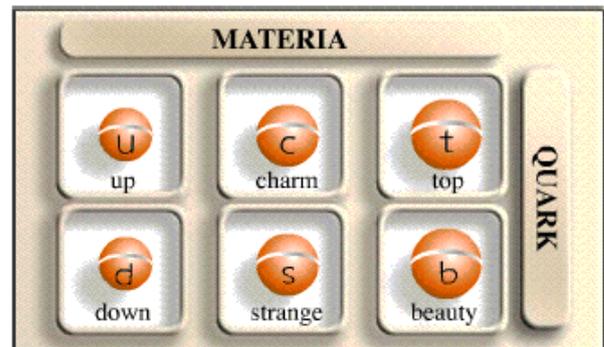


Figura 5: I sei quark.

In questo scenario resta un unico parametro, g da cui la teoria dipende. Un numero puro che stabilisce da solo la massa degli adroni, le sezioni d'urto delle reazioni nucleari che avvengono nelle stelle e che hanno prodotto i nuclei degli elementi pesanti da cui il nostro pianeta si è formato ed in esso la vita... Il motto Pitagorico "tutto è numero" trova qui una letterale conferma. Naturalmente riconoscere nella Lagrangiana della QCD, un'espressione che non prende più di una riga, tutte queste meraviglie è un compito piuttosto arduo, ma si può dire che essa riassume tutte queste informazioni, in un formato di compressione particolarmente estremo.

Nel paradiso del teorico, con solo 3 sapori di quarks di massa nulla, appare un'ulteriore simmetria, la cosiddetta simmetria chirale: avendo i quarks tutti identica massa nulla, si possono scambiare le "etichette" di sapore senza modificare le equazioni della teoria. Questa non è altro che la già discussa ottupla via di Gell-Mann. Inoltre il mescolamento dei sapori può avvenire indipendentemente per i quarks di elicità destra (il cui spin è allineato alla direzione dell'impulso) o sinistra (spin anti-allineato), da cui il nome di simmetria chirale. La stessa dinamica del vuoto che conduce al confinamento produce una rottura spontanea della simmetria chirale, termine con il quale si indica il fatto che lo stato fondamentale non rispetta tale simmetria. È un fenomeno analogo a quanto avviene nei sistemi di spin fer-

romagnetici, le cui interazioni sono invarianti per rotazioni globali del sistema ma il cui stato fondamentale esibisce una magnetizzazione spontanea, vale a dire con tutti gli spin che puntano in una stessa direzione, con il che l'originale invarianza per rotazioni è perduta. L'analogo della magnetizzazione spontanea nel caso della simmetria chirale della QCD è il condensato di quark-antiquark. Non si tratta delle effimere coppie di particelle virtuali, bensì di materia e antimateria persistente, che abita il vuoto e permea tutto lo spazio per sempre, come una specie di etere. Possiamo pensare che il vuoto, in assenza di condensato chirale, sia instabile, come un liquido al di sotto della temperatura di fusione pronto a solidificare alla minima perturbazione; la materializzazione di quark e antiquark comporta un guadagno di energia e avviene una transizione di fase. In questa materializzazione le elicità destre e sinistre dei quark vengono accoppiate in modo da rompere la simmetria iniziale sotto mescolamenti indipendenti delle diverse elicità. Le eccitazioni elementari di questo aggregato di quark e antiquark che pervade il vuoto sono interpretabili come particelle di massa nulla, chiamati bosoni di Nambu-Goldstone, che possono essere identificati con l'ottetto dei mesoni più leggeri.

Il condensato chirale è solo il più semplice di una infinità di parametri di questo genere, che segnalano la condensazione di combinazioni più complesse dei campi dei quarks e dei gluoni, e che costituiscono delle precise caratteristiche del vuoto della QCD. Dipendono anch'essi dall'unico parametro g , ma un loro calcolo in teoria delle perturbazioni (lo sviluppo in potenze di g) darebbe un risultato nullo ad ogni ordine. Si tratta infatti di proprietà prettamente non-perturbative, la cui determinazione sperimentale, tutt'altro che diretta, si basa principalmente sullo studio delle interazioni tra i bosoni di Nambu-Goldstone.

Una teoria senza nessun parametro dimensionale, che non ha quindi bisogno di appoggiarsi su unità di misura esterne che facciano da campione, riesce quindi a generare la complessità di cui siamo parte e a render conto della stragrande maggioranza della massa visibile dell'universo, tutto da un solo numero, g , senza bisogno del meccanismo di Higgs, che produce solo piccole correzioni rispetto a questo scenario. Correzioni

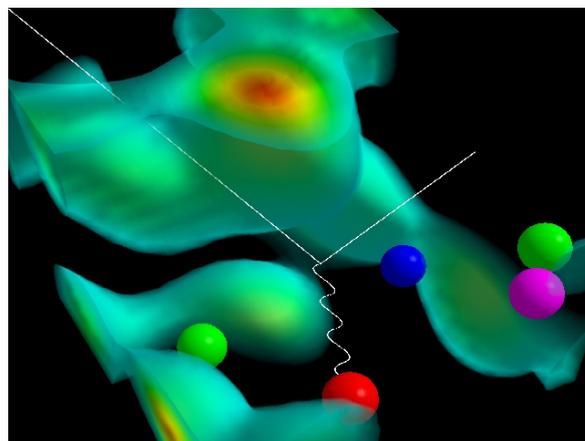


Figura 6: Una configurazione di gauge simulata in un reticolo del volume di $2.4 \times 2.4 \times 3.6$ fm, abbastanza da contenere un protone e la sua nuvola di pioni. Tratta dal sito di Derek B. Leinweber [11].

che sono calcolabili e calcolate in uno schema perturbativo.

In attesa di nuove prospettive teoriche sul problema, lo strumento più potente per decifrare il contenuto della Lagrangiana della QCD viene da simulazioni ai supercalcolatori. Questa tecnica è basata sulla formulazione di Feynman della teoria quantistica, per la quale ad esempio una particella che si muova da un punto A a un punto B segue in effetti *tutti* i possibili percorsi nello spazio-tempo, ciascuno con associato un peso statistico. Per simulare campi quantistici va fatta una media statistica su un formidabile numero di configurazioni, ciascuna generata con il peso statistico determinato dalla teoria.

Nella Ref. [11] sono contenute delle bellissime animazioni che illustrano la dinamica dei campi di gauge così generati, ottenute discretizzando lo spazio tempo e limitandolo ad un ipercubo, un reticolo delle dimensioni di pochi fermi.

In Fig. 6 è mostrata ad esempio la densità di energia di una di queste configurazioni; i quark rosso verde e blu, che costituiscono un protone sono confinati da queste configurazioni; appare anche una coppia di quark-antiquark (verde-antiverde) che circondano il protone come una nuvola. Il potere predittivo della QCD su reticolo è cresciuto enormemente con lo sviluppo di sofisticati algoritmi e con l'aumento delle risorse computazionali. Nella Fig. 7 sono mostrate le predizioni per le masse degli adroni più leggeri confrontate con quelle sperimentali. Un solo da-

un esempio.

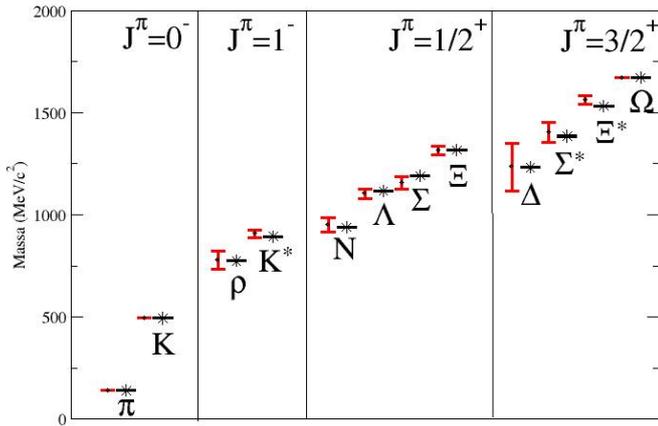


Figura 7: Predizioni di QCD sul reticolo per le masse degli adroni più leggeri (punti rossi), confrontate con i valori sperimentali (in nero). Le barre d'errore comprendono sia quello sistematico che quello statistico, inerente alla procedura numerica. Vi sono solo 3 input sperimentali: la massa del barione Ω , per fissare g , e la massa dei pioni π e dei kaoni K per fissare il valore delle masse dei quarks. I numeri quantici J^π si riferiscono allo spin e alla parità. I dati sono tratti dalla Ref. [12].

to sperimentale è utilizzato per fissare la costante g , o se si vuole per stabilire una scala assoluta. I restanti input del calcolo (dati dalla massa dei pioni π e dei kaoni K) servono solo a fissare le masse dei quarks più leggeri, up, down e strange.

Non ce ne sarebbe bisogno se ci accontentassimo dei risultati nella situazione ideale di quarks di massa nulla. In quel caso, pioni e kaoni emergerebbero come veri bosoni di Nambu-Goldstone, particelle quindi di massa nulla, mentre la massa degli altri adroni sarebbe solo leggermente modificata.

Forse il punto più significativo della figura è quello che rappresenta i nucleoni N , col quale termine si indicano protoni e neutroni, i costituenti del nucleo, che contribuiscono in maniera dominante alla massa visibile dell'universo. La loro esistenza simulata rende conto perfettamente delle loro caratteristiche reali, a tal punto da sfumare il confine tra realtà virtuale e realtà oggettiva. Un confine che potrebbe essere addirittura abbattuto dall'avvento di calcolatori quantistici, del quale in fondo l'intero Universo costituisce

- [1] R.P. Feynman, R.B. Leighton, M.L. Sands: *The Feynman Lectures on Physics*. Addison-Wesley, New York (1963).
- [2] A. Pais: *Subtle is the Lord* Oxford University Press, Oxford (1982).
- [3] P.A.M. Dirac: *I principi della meccanica quantistica*. Bollati Boringhieri, Torino (1976).
- [4] D. Griffiths: *Introduction to elementary particles*. Wiley-Vch, Torino (2004).
- [5] A. Pais: *Niels Bohr's times*. Clarendon Press, Oxford (1991).
- [6] A.S. Kronfeld, C. Quigg: *Resource Letter: Quantum Chromodynamics*, Am. J. Phys., **78** (2010) 1081.
- [7] F. Wilczek: *The lightness of being*. Basic Books, ? (2008).
- [8] Y. Nambu: *Quarks*. World Scientific, Singapore (1985).
- [9] C. Patrignani et al. (Particle Data Group): *The Review of Particle Physics*, Chin. Phys., **C40** (2016) 100001.
- [10] H. Leutwyler: *Principles of Chiral Perturbation Theory*, Proc. Int. Sch. Phys. Fermi, **130** (1996) 1.
- [11] <http://www.physics.adelaide.edu.au/theory/staff/leinweber/VisualQCD/Nobel/index.html>
- [12] S. Durr et al.: *Ab-Initio Determination of Light Hadron Masses*, Science, **322** (2008) 1224.



Luca Girlanda: si è laureato in Fisica presso l'Università di Pisa nel 1997. Dopo aver conseguito il Master in Fisica Teorica e il PhD presso l'Università di Paris-Sud XI, ha lavorato con contratti post-dottorali presso l'Università Autonoma di Barcelona, l'Università di Padova, l'ECT* di Trento, l'INFN di Pisa e l'Università di Pisa. Dal 2010 è ricercatore in Fisica Teorica presso l'Università del Salento. I suoi interessi di ricerca riguardano l'interazione nucleare e la simmetria chirale della QCD.

La più grande cantonata di Einstein

Paolo Ciafaloni INFN & Dipartimento "Ennio De Giorgi" - Lecce

Nel 1917 Einstein introduce nelle equazioni della relatività generale, che descrivono le interazioni di gravità, un termine denominato costante cosmologica. Dopo la scoperta dell'espansione dell'Universo ad opera di Hubble nel 1929, Einstein stesso ritratta la propria modifica etichettandola come il proprio errore più madornale. Per 70 anni il termine introdotto da Einstein viene ignorato dalla comunità scientifica, ma nel 1998 la costante cosmologica risorge in maniera drammatica grazie ad una serie di osservazioni su supernovae distanti 6 miliardi di anni luce da noi. Le attuali osservazioni astrofisiche indicano che la dinamica dell'Universo è dominata dalla moderna reincarnazione della costante cosmologia, denominata energia oscura, che compone il 70% dell'energia totale. L'energia oscura agisce come una sorta di antigravità a distanze cosmologiche. Le teorie delle interazioni fondamentali indicano valori della densità di energia oscura 120 ordini di grandezza superiori a quelli osservati: alcuni considerano questa come la peggior predizione nella storia della fisica.

Prologo

Su grandi scale di distanza, la dinamica dell'Universo è dominata dalle interazioni gravitazionali. In effetti, le interazioni deboli e forti sono confinate all'interno dei nuclei, quindi su scale di distanze microscopiche. Quanto all'interazione elettromagnetica, cariche positive e negative tendono a compensarsi, e l'interazione elettromagnetica è totalmente trascurabile già su distanze dell'ordine della distanza terra-luna. La teoria della gravità migliore che abbiamo è la teoria della relatività generale di Einstein, che ad oggi è compatibile con tutte le verifiche sperimentali. In questa teoria, un "oggetto" produce una curvatura nello spazio-tempo, la quale a sua volta produce un effetto fisico, determinando il moto di un altro "oggetto". Per "oggetto" si intende "qualunque forma di energia", quindi non solo la massa* ma anche, ad esempio, l'energia connessa con un campo elettromagnetico (cioè fotoni nella teoria quantistica dei campi).

Non è facile intuire cosa significhi "curvatura dello spazio-tempo", anche perché è ovviamente impossibile avere una immagine mentale di un oggetto a 4 dimensioni. Possiamo capire però che cosa sia la curvatura in uno spazio bidimensionale quale è, sostanzialmente, quello in cui viviamo sulla superficie della Terra. Fintanto che ci muoviamo su distanze piccole, tale superficie è

*La massa è una forma di energia, secondo la celebrata formula $E = mc^2$.



Figura 1: *La terra è curva! Nell'immagine possiamo vedere la curvatura globale determinata dal raggio della Terra e la curvatura locale prodotta dalle montagne. Per analogia l'Universo è caratterizzato da una curvatura globale e da curvature locali collegate ad esempio alle galassie.*

sostanzialmente piatta e ci possiamo muovere di moto rettilineo uniforme. Quando ci muoviamo su distanze grandi, cioè paragonabili al raggio della Terra, il moto rettilineo è impossibile e la curvatura ci obbliga a deviare da esso. In questa analogia possiamo comprendere come la curvatura faccia deviare gli oggetti dal moto rettilineo uniforme. A un livello di dettaglio maggiore, esistono sulla Terra delle piccole regioni con curvatura quali le montagne, che anch'esse influenzano il nostro moto. Per analogia possiamo dire che l'Universo è caratterizzato da curvature "locali", determinate ad esempio da una galassia, e da un raggio di curvatura "globale" che determina la deviazione dal moto rettilineo anche di oggetti lontani da qualsiasi curvatura locale. Sono curvature locali anche quelle prodotte dal nostro sole ad esempio, che determinano il moto dei pianeti. Possiamo distinguere la curvatura spaziale, che è quella discussa fino ad ora, da quella temporale, che determina una accelerazione come vedremo più in dettaglio nel prossimo paragrafo.

La gravità è un'interazione attrattiva, e agisce su qualunque tipo di energia o materia. Questo implica che se osserviamo un insieme di oggetti in una data configurazione iniziale, nel tempo le distanze fra gli oggetti tendono a diminuire e alla fine il sistema collassa su se stesso. Un universo statico è quindi instabile a causa delle interazioni gravitazionali, e non può esistere. Nel 1917 Einstein, nel tentativo di risolvere questa difficoltà, introduce un termine aggiuntivo nelle equazioni che descrivono la gravità. Tale termine, che

è caratterizzato da quella che oggi chiamiamo costante cosmologica, agisce come una sorta di antigravità, equilibrando su grandi distanze l'interazione attrattiva di gravità e stabilizzando l'Universo statico. Naturalmente la costante deve essere sufficientemente piccola da produrre effetti trascurabili su distanze dell'ordine del sistema solare o inferiori, ma sufficientemente grande da ottenere l'effetto voluto.

La legge di Hubble

Saltiamo di 12 anni e arriviamo al 1929. L'astronomo E. Hubble osserva una alterazione sistematica delle righe spettrali nella luce proveniente dalle galassie. Le righe spettrali sono una sorta di firma digitale degli elementi presenti in natura. In effetti un dato elemento, ad esempio l'idrogeno, non emette luce a tutte le lunghezze d'onda, ma solo a determinate lunghezze d'onda raffigurate in Fig. 2. Nella figura osserviamo anche l'alterazione delle righe (dette anche spettro di emissione) verso lunghezze d'onda più grandi (frequenze più basse) detto redshift, e verso lunghezze d'onda più piccole (frequenze più grandi) detto blueshift.

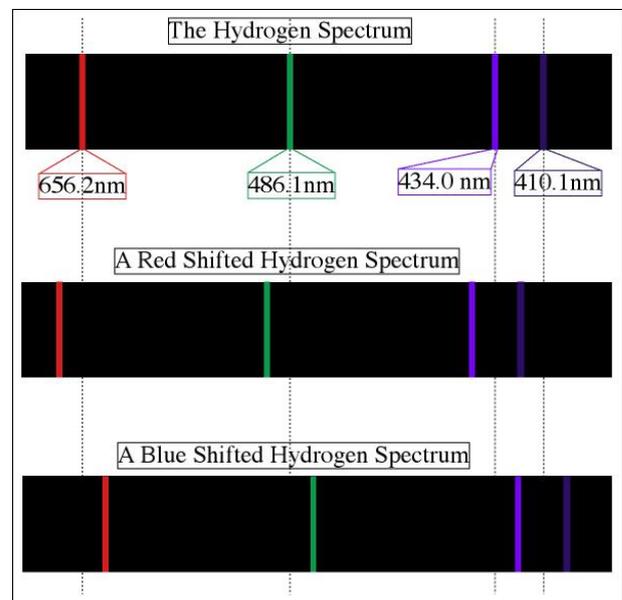


Figura 2: *Righe di emissione (o spettro di emissione) dell'idrogeno molecolare, H_2 . In figura anche lo spostamento dello spettro verso il rosso (redshift) e verso il blu (blueshift)*

Il redshift si quantifica tramite la variabile $z = \Delta\lambda/\lambda = -\Delta\nu/\nu$, dove λ (ν) è la lunghezza

d'onda (frequenza) naturale e $\Delta\lambda = \lambda_{\text{mis}} - \lambda$, λ_{mis} essendo la lunghezza d'onda misurata. Hubble non osserva nessun tipo di blueshift, ma misura un redshift in tutte le galassie osservate. Inoltre ricava una relazione sistematica fra il redshift z e la distanza dalla terra d della galassia osservata, la legge di Hubble:

$$z = \frac{H_0}{c} d \quad (1)$$

dove c è la velocità della luce e H_0 la costante di Hubble, il cui valore misurato odierno è $H_0 = 73 \pm 1.75$ (km/s)/Mpc*. Quindi il redshift aumenta linearmente con la distanza. Tale legge è valida per distanze superiori a circa 10 Mpc, che è la distanza tipica dello spazio extragalattico.

L'interpretazione più naturale della legge di Hubble è che il redshift sia collegato a una velocità di allontanamento delle galassie, per via di un effetto fisico chiamato effetto Doppler. Ognuno di noi ha esperienza dell'effetto Doppler nel caso delle onde sonore: la sirena di una ambulanza che si sta avvicinando ha un suono più acuto della stessa sirena quando l'ambulanza è ferma, mentre quando l'ambulanza si allontana la tonalità è più grave†. Nel caso del redshift, la teoria dell'effetto Doppler prevede che la velocità di recessione di una galassia sia data da $v = zc$. Indicando con il punto la derivata temporale si ottiene quindi la relazione (detta anch'essa legge di Hubble):

$$v = \dot{d} = H_0 d \quad (2)$$

Se si assume che l'universo sia omogeneo (cioè con uguali proprietà in tutti i punti) ed isotropo (cioè senza direzioni privilegiate), la legge di Hubble comporta l'espansione dell'universo, che descrivo più in dettaglio nel prossimo paragrafo.

Un universo in espansione non necessita di alcuna costante cosmologica, per cui Einstein stesso dopo le osservazioni di Hubble abbandona la costante cosmologica, qualificando la sua introduzione come un errore madornale‡.

*1 Mpc (Megaparsec) equivale a circa $3.3 \cdot 10^6$ anni luce, ovvero circa $3.1 \cdot 10^{19}$ km

†La stessa cosa succede nel caso di un jet che ci passi sopra la testa ad esempio.

‡È curioso notare come Hubble invece sia rimasto scettico per tutta la vita sulla interpretazione della relazione redshift-distanza in termini di espansione dell'universo, attribuendo tale relazione a 'un ignoto effetto fisico' [2]

La costante cosmologica diventa una curiosità accademica, e la comunità scientifica la mette nel cassetto per 70 anni. Ma una serie di osservazioni effettuate nel 1998 causano una rivoluzione nella comunità dei cosmologi. Le misure effettuate su una serie di oggetti astrofisici denominate Supernovae di tipo Ia indicano che l'espansione dell'Universo stia accelerando[1]. La cantonata più grande di Einstein non è affatto una cantonata, e l'antigravità esiste davvero.

Espansione e accelerazione

Per comprendere cosa significhi l'espansione dell'Universo, è utile una analogia con un panettone in lievitazione. In questa analogia un chicco di uvetta rappresenta una singola galassia, e la pasta fra un chicco e l'altro rappresenta lo spazio vuoto intergalattico. supponiamo che la pasta in lievitazione abbia forma sferica e rappresentiamo il suo raggio ad un certo istante con $R(t)$. Possiamo sempre scrivere:

$$\dot{R}(t) \equiv \frac{dR(t)}{dt} = H(t)R(t) \quad (3)$$

dove $H(t)$ dipende dall'istante t di osservazione.



Figura 3: *L'Universo in un panettone. Ogni chicco di uvetta rappresenta una galassia, mentre la pasta del panettone rappresenta lo spazio "vuoto", che poi tanto vuoto non è (vedi testo). La lievitazione del panettone corrisponde alla osservata espansione dell'Universo.*

Il punto cruciale adesso è supporre che il panettone che lievita sia omogeneo e isotropo. Ovviamente su scale dell'ordine delle dimensioni dei chicchi di uvetta ci sono disomogeneità dovute ai chicchi stessi, ma su scale più grandi possiamo supporre che il numero di chicchi in un

dato volume sia circa lo stesso in ogni porzione di panettone, che non esistano strutture con una direzione privilegiata, che l'agente lievitante sia distribuito in maniera uniforme. È facile allora rendersi conto che la (3) vale per *qualsiasi* scala di distanza. Cioè, data qualsiasi coppia di chicchi ad una distanza d , la variazione col tempo della distanza è data dalla legge di Hubble (2). Poiché tutte le distanze variano in maniera uniforme, possiamo definire un fattore di scala comune $a(t)$ adimensionale, tale che $d(t) = a(t)d_0$, dove d_0 è la distanza fra una coppia di oggetti al momento t_0 . Convenzionalmente si pone $a(0) = 1$ all'istante $t_0 = 0$ in cui il panettone (l'Universo) viene osservato nel presente.

L'equazione di Friedmann

Come accennato nel prologo, la dinamica dell'Universo, cioè le equazioni che dicono come il fattore di scala $a(t)$ varia nel tempo, è determinata dall'interazione di gravità. Tale interazioni sono descritte dalle equazioni di Einstein della relatività generale. Le equazioni si semplificano notevolmente per via dell'assunzione di omogeneità e isotropia, dando luogo all'equazione di Friedmann, che è l'oggetto di questo paragrafo:

$$\frac{1}{2}(\dot{d})^2 - \frac{4\pi G}{3} \rho d^2 = U \quad (4)$$

dove G è la costante di Newton e ρ la densità di energia, che come vedremo è composta di vari contributi. Curiosamente, questa equazione si può ricavare nell'ambito della descrizione Newtoniana della gravità ed ammette una interpretazione in termini di conservazione dell'energia: il primo termine del membro di destra rappresenta l'energia cinetica e il secondo l'energia potenziale; le due si sommano a dare l'energia totale U , che è costante. La gravità Newtoniana non fornisce alcuna espressione per U ; invece in relatività generale U dipende in maniera precisa dalla curvatura globale dell'Universo. Nel caso del nostro Universo, tutte le osservazioni concordano con il fatto che il nostro universo è globalmente piatto (raggio di curvatura infinito); in tal caso la relatività generale prevede $U = 0$ per cui d'ora in poi ometterò questo termine. Introdotto il fattore di

scala $a(t)$, l'equazione di Friedmann si riscrive:

$$H^2(t) = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \rho(t) \quad (5)$$

Anche in questa forma, l'interpretazione fisica è abbastanza chiara: la densità di energia ρ determina una curvatura, in questo caso temporale, che a sua volta genera una accelerazione attraverso la costante di Hubble.

Se supponiamo che l'universo sia composto principalmente di materia ordinaria, quale gli atomi, i nuclei, gli elettroni che compongono i pianeti e le stelle, man mano che l'universo si espande (cioè $a(t)$ aumenta con t), la densità ρ diminuisce: questo è una ovvia conseguenza del diluirsi della materia in un volume che aumenta. Di conseguenza la costante di Hubble determinata dalla (5) diminuisce col tempo. In questo caso l'espansione dell'Universo *rallenta*, cioè decelera: questo è ciò che ci si aspetta intuitivamente dato che la gravità è un'interazione attrattiva. Ma il caso della costante cosmologica è diverso: essa è una proprietà del tessuto spaziotemporale ed è appunto costante nel tempo e nello spazio, determinando una densità di energia ρ_Λ che non diminuisce col tempo. In tal caso l'espansione è *accelerata* ed anzi di tipo esponenziale $a(t) \sim e^{\alpha t}$ con $\alpha = \sqrt{\frac{8\pi G}{3} \rho_\Lambda}$. La densità di energia del vuoto è connessa con la costante cosmologica Λ che compare nelle equazioni di Einstein tramite la relazione:

$$\rho_\Lambda = \frac{\Lambda}{8\pi G} \quad (6)$$

Le attuali misure indicano una densità di energia del vuoto pari a:

$$\rho_\Lambda \approx 10^{-26} \text{ kg/m}^3 \approx 10^{-11} \text{ eV}^4 \quad (7)$$

che corrisponde a circa 7 atomi di idrogeno per centimetro cubo[§]. Da un punto di vista più moderno, per aprire la possibilità a piccole variazioni della densità di energia nel tempo e nello spazio, si utilizza il termine 'energia oscura'. Le attuali osservazioni implicano che la densità di energia oscura sia molto più grande di quelle della materia ordinaria e della altrettanto misteriosa

[§]Nella (7) si è espressa la densità anche in unità 'naturali' di uso comune nella fisica delle alte energie, nelle quali la massa e l'inverso della lunghezza si esprimono in unità di energia: gli elettronvolt (eV).

materia oscura, costituendo circa il 70 % della densità di energia totale dell'Universo.

L'energia del vuoto si può predire in base alle attuali teorie fisiche?

In breve, la risposta è "No, non si può". Tuttavia, ci sono buone ragioni per ritenere che la conoscenza delle teorie microscopiche fondamentali consenta quantomeno di stimare un'ordine di grandezza per l'energia del vuoto e la costante cosmologica. Intanto, cos'è il vuoto? Supponiamo di considerare una porzione di spazio che contenga energia sotto varie forme: materia, radiazione sotto forma di luce, neutrini ecc. Dal punto di vista della teoria dei campi, che unisce relatività ristretta e meccanica quantistica, tutte le forme di energia sono 'stati di eccitazione del vuoto', cioè stati con energia maggiore dello stato di energia minima, e contenenti una o più particelle. Ad esempio, la luce è costituita da un particolare tipo di particella, i fotoni. Ogni volta che aggiungo un fotone aumento l'energia del sistema, poiché i fotoni hanno energia positiva. Se supponiamo ora di poter estrarre da tale porzione di spazio tutte le forme di energia, cioè tutte le particelle, otteniamo lo stato di minima energia. Il vuoto è quindi uno stato in cui non ci sono particelle elementari.

Per capire quanto valga l'energia del vuoto, consideriamo per iniziare un sistema semplice: una molla, considerata ideale, in cui la forza di richiamo è proporzionale allo spostamento x dalla posizione di equilibrio: $F = -kx$ dove k è una costante. Questo sistema è detto anche oscillatore armonico. Dal punto di vista della meccanica classica l'energia di un oscillatore di massa m è la somma di energia cinetica ed energia potenziale:

$$E = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 \quad (8)$$

e può assumere qualunque valore maggiore o uguale 0. Ma in meccanica quantistica le cose cambiano. Non tutte le energie sono permesse, e i livelli energetici sono quantizzati, cioè assumono valori discreti e non più continui. Inoltre lo stato di energia minima non ha energia 0, bensì $E_0 = \frac{h}{4\pi}\omega$ dove $\omega = \sqrt{k/m}$ è la frequenza di risonanza dell'oscillatore e h la costante di Planck

(vedi Fig. 4). Quindi l'oscillatore armonico ha una energia fondamentale (o energia minima, o energia di punto zero) diversa da 0 e ben definita in termini dei parametri che lo descrivono. Si può intuire perché questo accada tenendo presente il principio di indeterminazione di Heisenberg, che sancisce che non si possono misurare contemporaneamente posizione (in questo caso x e velocità (in questo caso \dot{x}). Questo significa che il sistema non può esistere nello stato in cui vale $x = \dot{x} = 0$, che corrisponderebbe a energia nulla. Invece, il sistema fondamentale è caratterizzato da *fluttuazioni quantistiche* tali che sia il valor medio dell'energia cinetica che quello dell'energia potenziale sono diversi da zero. Il sistema si assesta in modo tale da assumere il valore più piccolo possibile dell'energia, che come osservato non è 0. In Fig. 4 sono graficate le densità di probabilità, cioè le funzioni che rappresentano la probabilità di trovare il sistema nella coordinata x : in meccanica quantistica infatti non si può predire il valore della coordinata alla quale si trova il sistema, ma solo la sua probabilità. Dalla figura si capisce come lo stato di minima energia non corrisponda a $x = 0$.

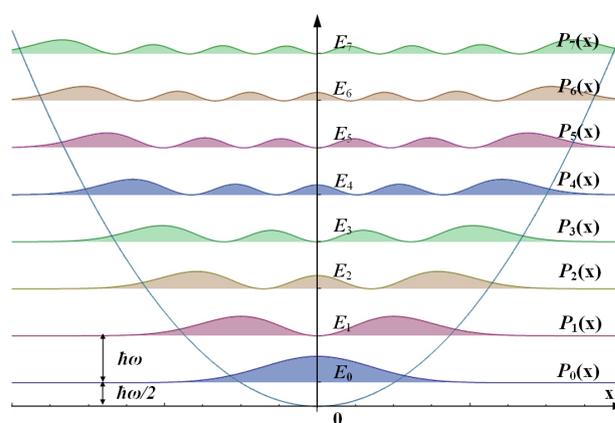


Figura 4: Livelli energetici e densità di probabilità dell'oscillatore armonico. Il livello fondamentale corrisponde ad un'energia $E_0 = \frac{h}{4\pi}\omega$ e a fluttuazioni della coordinata intorno al valore $x = 0$.

Passiamo ora a considerare quello che succede in teoria quantistica dei campi, che mette insieme meccanica quantistica e relatività ristretta. È possibile raffigurare un campo quantistico come un numero infinito di oscillatori armonici, ognuno definito in un punto dello spazio-tempo. Sommando le energie di punto zero di ognuno degli oscillatori otteniamo l'energia di punto zero del

vuoto. La teoria dei campi fornisce una espressione precisa per l'energia minima di un volume finito di spazio V e conseguentemente anche per la densità di energia:

$$\rho_{\text{vuoto}} = \frac{1}{V} E_{\text{vuoto}} = \int_0^\infty \omega^3 d\omega \quad (9)$$

Quindi a causa del numero infinito di gradi di libertà l'energia del vuoto quantistico è infinita!

Occorre a questo punto fare due osservazioni. La prima è che nella stragrande maggioranza dei casi il valore dell'energia del vuoto può essere ignorato, in quanto nei processi fisici compaiono sempre differenze di energie, e in tali differenze il valore dell'energia minima si cancella. Tuttavia non possiamo assumere lo stesso atteggiamento nel caso della gravità, in quanto le interazioni gravitazionali, come abbiamo visto, hanno come sorgente *tutte* le forme di energia: una energia infinita causa una curvatura infinita, ed è pertanto impossibile da interpretare. La seconda osservazione è che in teoria dei campi gli infiniti compaiono dappertutto, appunto per via del numero infinito di gradi di libertà. Ma esiste un metodo in grado di dare un significato alle divergenze ottenendo alla fine quantità fisiche finite. Tale metodo si chiama rinormalizzazione, e provo ad accennarlo. Il primo passo consiste nel regolarizzare, cioè dare un significato, all'integrale (9), introducendo una scala di energia massima Λ_{cut} , ottenendo così $\rho_{\text{vuoto}} \sim \Lambda_{\text{cut}}^4$. La scala di energia di cutoff compare come parametro intermedio, ma le quantità osservabili fisiche non possono dipendere da tale scala. La densità di energia del vuoto riceve poi un'altro contributo direttamente come parametro presente nella teoria, ρ_0 . In definitiva si ha:

$$\rho_\Lambda = \Lambda_{\text{cut}}^4 + \rho_0(\Lambda_{\text{cut}}) \quad (10)$$

Normalmente la dipendenza di ρ da Λ_{cut} viene fissata in modo che il valore fisico (in questo caso ρ_Λ) sia determinato dall'esperimento. Ad esempio la massa misurata dell'elettrone non è predetta dal Modello Standard delle interazioni fondamentali: essa è la somma del contributo 'bare' m_0 che è un parametro della teoria, e di un contributo divergente che dipende esplicitamente da Λ_{cut} . Nessuna teoria può dunque *predire* il valore di ρ_Λ . Tuttavia nel moderno modo di

vedere la rinormalizzazione, la scala di cutoff introdotta in precedenza ha il significato di 'scala massima di energia alla quale la teoria è valida'. Possiamo supporre che la nostra maniera di intendere la fisica delle particelle tramite la teoria dei campi e il Modello Standard sia valida fino alla scala di Planck: $\Lambda_{\text{cut}} = \Lambda_{\text{Planck}} \approx 10^{19}$ GeV, per cui $\Lambda_{\text{Planck}}^4 \approx 10^{112} \text{eV}^4$. Sappiamo infatti che oltre questa scala sono rilevanti effetti quantistici gravitazionali che non siamo capaci di descrivere. Inserendo il valore misurato (7) nella (10) otteniamo:

$$\rho_\Lambda \approx 10^{-11} \text{eV}^4 = 10^{112} \text{eV}^4 + \rho_0 \quad (11)$$

Questa equazione implica che i due termini $\Lambda_{\text{Planck}}^4$ e ρ_0 , a priori totalmente indipendenti, siano accidentalmente identici (e con segno opposto) fino alla centoventiresima cifra significativa! Potremmo richiedere un livello di cancellazione 'naturale' a livello dell'un per mille, basandoci su quello che accade normalmente nel Modello Standard per altre quantità fisiche. In questo caso la predizione 'naturale' è che i due termini di ordine 10^{112} eV si cancellino parzialmente a dare $\rho_\Lambda \approx 10^{109}$ eV. Ma questo valore 'naturale' è 120 ordini di grandezza più grande di quello osservato, una situazione descritta [3] come 'la peggior predizione nella storia della fisica'.



- [1] Vedi ad esempio [Accelerating expansion of the Universe](#) su Wikipedia.
- [2] https://en.wikipedia.org/wiki/Edwin_Hubble
- [3] M.P. Hobson, G.P. Efstathiou, A.N. Lasenby, *General Relativity: An introduction for physicists* (Reprint ed.). Cambridge University Press., Cambridge, (2006), p. 187.



Paolo Ciafaloni: è un ricercatore in fisica teorica presso l'INFN e l'Università del Salento; vive ad Arnesano in provincia di Lecce. Laureato a Pisa nel 1991, si occupa di fisica delle particelle e di cosmologia.

La quantità del nulla

*Yet zero was inexorably linked with the void – with nothing,
[1] p. 9.*

Charles Seife (1972 –)

Lorenzo Dello Schiavo

Institut für Angewandte Mathematik –

Rheinische Friedrich-Wilhelms-Universität Bonn – Bonn, Germania

Anna Baccaglioni-Frank

Dipartimento di Matematica – Università di Pisa

0 *zero*, ovvero la quantità del nulla. Un'idea che — assieme a quella gemella di insieme *vuoto* — ha segnato la storia della matematica attraverso i secoli e i continenti, ben al di là dell'introduzione dei sistemi numerici posizionali o degli interi relativi.

 *vuoto*, ovvero l'insieme privo di elementi ed unico sottoinsieme di sé, ogni elemento del quale soddisfa ogni proprietà: *adynaton* per la filosofia antica, *paradosso* per quella moderna, *commodité* irrinunciabile per la matematica contemporanea.

Dell'uso di questi due concetti, intimamente legati, cercheremo di fornire una storia breve, corredata dalle fonti originali dall'antichità ai giorni nostri ed arricchita da curiosità matematiche, accennando poi a come alcune tappe storiche segnino delicati passaggi cognitivi nel loro apprendimento.

1 Introduzione e breve storia

Pochi tra i concetti elementari della matematica moderna hanno una storia affascinante quanto quella dello zero, la quantità del nulla. Una storia all'inizio della quale, come spesso nell'antichità, la matematica, con i suoi enti e le sue costruzioni, non costituì per molti secoli una disciplina a sé stante, contesa piuttosto tanto da dottrine filosofiche o religiose, dall'ontologia all'etica, quanto da necessità pratiche e materiali, dall'agrimensura all'astronomia, dall'ingegneria al commercio. Una storia dunque, scritta da sacerdoti indiani su foglie di palma o da astronomi babilonesi su tavolette d'argilla, da mercanti italiani e da bibliotecari arabi, dove lo zero ed il vuoto, il nulla, furono di volta in volta sia concetti di uso comune, denotati con simboli propri, sia necessità inconsapevoli, accettate per comodità o per convenzione, ma spesso prive di identificazione, sia concetti ignorati ed osteggiati, dei quali si negava la stessa esistenza.

Di questa stessa storia, che riprendiamo brevemente di seguito nei suoi aspetti universali, tratteremo poi alcuni aspetti individuali, relati-

vi all'apprendimento del concetto di zero nella prima infanzia ed alle difficoltà cognitive che si presentano in relazione ad esso. È interessante notare come la nostra prospettiva storica si possa rileggere — scevra dalle molte sovrastrutture filosofico-religiose accumulate nel corso del tempo e dai tecnicismi della disciplina matematica contemporanea — proprio alla luce di tali difficoltà.

Etimologia Tra le ipotesi maggiormente accreditate riguardo l'origine del termine *zero* è la contrazione del veneziano *zevero* per il latino *zephyrum*, con il quale il Fibonacci (ca 1175 – ca 1235 d. C.) traduceva nel suo *Liber abbaci* l'arabo *صفر* *sifr*¹, orig. *vuoto*:

Novem figure indorum he sunt

9 8 7 6 5 4 3 2 1

*Cum his itaque novem figuris, et cum hoc signo 0, quod arabice zephyrum appellatur, scribitur quilibet numerus, ut inferius demonstratur.*²

quest'ultimo a sua volta per il sanscrito *śūnya*, dal medesimo significato di *vuoto*, nell'accezione³ di *privo di contenuto*, ma accostato, specialmente nella filosofia buddhista, anche a quello di *nullità*, *non-esistenza*.

Uso presso Babilonia e la Grecia Antica

ἀεὶ ἦν ὁ τι ἦν καὶ ἀεὶ ἔσται. εἰ γὰρ ἐγένετο, ἀναγκαῖόν ἐστι πρὶν γενέσθαι εἶναι μηδέν· εἰ [τύ] τοῖσιν μηδέν ἦν, οὐδαμὰ ἂν γένοιτο οὐδὲν ἐκ μηδενός.⁴

Melisso di Samo (sec. v a. C.)

¹La medesima etimologia ha anche l'italiano *cifra*.

²Le nove figure degli Indiani sono queste: 9 8 7 6 5 4 3 2 1. Con tali nove figure, e con il simbolo 0, in arabo detto *zephyro*, si scriverà qualsiasi numero, com'è dimostrato più avanti. [2, I] [trad.: L. D. S.].

³Alla voce *śūnya*, in particolare v. L220069 per il primo significato qui riportato e L220084 per il secondo [3].

⁴Simplicio (ca 490 – ca 560 d. C.) riportando Melisso: *Sempre era ciò che era e sempre sarà. Poiché se si fosse generato, allora necessariamente prima della sua generazione esso sarebbe stato nulla; e se dunque fosse stato nulla, in nessun caso dal nulla si sarebbe potuta generare cosa alcuna.* [trad.: L. D. S.] da [4, 20. B1, p. 148].

La *sententia* dei Presocratici οὐδέν ἐξ' οὐδενός, *oudèn ex' oudenòs*, ovvero *nulla* [viene] *dal nulla* — ripresa anche in ambito romano:

*Principium cuius hinc nobis exordia sumet, nullam rem e nihilo gigni divinitus umquam*⁵

Lucrezio (94 – ca 55 a. C.)

— tradisce tutta l'ostilità di gran parte del mondo classico verso il concetto di *nulla*. Lo zero, quantità di esso, non gode di maggior favore, e in effetti persino l'unità stessa, la *monade*, sarà, fino ad epoche molto successive, un concetto controverso:

*Il ponto [punto] in Geometria, è simile alla unità nella Arithmetica: la qual è principio del numero, & non è numero*⁶.

Niccolò Tartaglia (1499 – 1557 d. C.)

Sebbene presso i Greci del v – iii secolo a. C. lo zero fosse — al pari di quello di *infinito* o di numero *irrazionale* — un concetto ignorato, quando non esplicitamente avversato, esso non era tuttavia ignoto. In particolare, avrebbe trovato applicazione presso gli astronomi del ii secolo a. C. — gli inventori della macchina di Antikythera⁷ — i quali ne avevano adattato l'uso, importandolo dal sistema numerico assiro-babilonese dell'età alessandrina.

Quest'ultimo era un sistema cumulativo-posizionale⁸ a base mista decimale e sessagesimale, in cui la decina era rappresentata dal segno ' < ', l'unità dal segno ' | ' e i numeri da 1 a 59 erano indicati dalla giustapposizione di questi simboli, ad esempio 𐤆 𐤇 per 47.

Tale sistema raggiunse la maturità di una forma rigorosa durante la dinastia seleucide (secc.

⁵Tito Lucrezio Caro. *De rerum natura* I, 149–150: *Il suo fondamento prenderà per noi inizio da ciò: che, per volere divino, cosa alcuna si genera mai dal nulla.* Trad.: L. D. S..

⁶Commento alla definizione di punto negli *Elementi* di Euclide. [5, fol. 7^v]

⁷Si tratta di un sorprendente meccanismo ad orologeria (ca 150 a. C.), unico nel suo genere tra i reperti giunti sino a noi, mediante il quale era possibile predire, calcolando il tempo intercorso tra essi, certi eventi astronomici, quali le fasi lunari, i movimenti di alcuni pianeti e gli equinozi.

⁸Si veda [6], p. 12, per una classificazione dei sistemi numerici.

iv – i a. C.), divenendo così particolarmente adatto al calcolo delle cosiddette *frazioni astronomiche*⁹, motivo per il quale fu infine adottato anche in Grecia. A partire da questo periodo [6], p. 253, si registra infatti l'introduzione di un carattere speciale, '  ', avente duplice funzione: la prima, di indicare una posizione vuota nella scrittura posizionale di un numero frazionario, ad esempio per distinguere un intero da una frazione propria¹⁰; la seconda, di essere impiegato come simbolo meramente epigrafico (cioè privo di valore numerico) allo scopo di impedire letture erranee (ad esempio ϰϑ potrebbe leggersi sia 47 sia $2407 = 40 \cdot 60 + 7$, mentre la grafia $\text{ϰϑ} \text{  }$ precluderebbe la prima lettura).

Una delle prime apparizioni sistematiche dello zero in ambito greco, proprio con la prima funzione dello zero assiro-babilonense, si registra nell'opera del matematico, astronomo e geografo greco-romano Claudio Tolomeo (ca 100 – ca 170 d. C.), ideatore del modello astronomico oggi noto come sistema aristotelico-tolemaico (v. Tabella 1).

Dall'India all'Arabia all'Europa È opinione diffusa¹¹ che l'introduzione dello zero come valore numerico, sebbene con uso dissimile da quello moderno, sia ascrivibile al matematico e astronomo indiano Brahmagupta di Bhillamāla (ca 598 – d. il 665 d. C.) nella sua *Brāhmasphuṭasiddhānta*, ovvero la *dottrina di Brahma propriamente*

⁹La conoscenza astronomica del tempo era diffusa in gran parte mediante tavole astronomiche o *almanacchi* (dall'arabo *المنامخ* *al-manākh*, orig.: *clima*), le quali riportavano la posizione relativa degli astri. Da queste deriveranno le successive tavole astronomiche arabe dette *زنج* *zīj*, letteralmente *corde*, tra cui la *Zīj al-Sindhind al-kabīr* di al-Ḥwārizmī (v. infra).

¹⁰Tuttavia, mai in posizione finale, ad esempio per distinguere la decina dall'intero, da cui si è soliti dedurre che lo zero assiro-babilonense fosse concepito esclusivamente come una sorta di separatore (cf. [6], p. 252), come nelle prime e più antiche attestazioni testuali, dove ha un ruolo assimilabile a quello del moderno punto e virgola.

¹¹Si veda ad esempio [7], p. 242, riferito a Brahmagupta: *The systematized arithmetic of the negative numbers and zero is, in fact, first found in his work.*

περιφερειῶν	εὐθειῶν			ἑξήκοστων			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
κϑ	λ	β	μδ	ο	α	ο	μη
κϑζ'	λ	λγ	η	ο	α	ο	μδ

Tabella 1: Parte della cosiddetta Tavola delle corde, dall'*Almagesto* di Tolomeo, trattato di astronomia e matematica. Nella tavola (v. [8, cap. 1, sez. xi pp. 48ss.]) sono riportati i valori approssimati della lunghezza delle corde sottese agli angoli al centro di un cerchio del raggio di sessanta unità. L'angolo al centro di θ° (ovvero l'arco di circonferenza o περιφερειῶν, peripherèion) varia da 1° a 180° ad intervalli di mezzo grado ed è indicato, così come gli altri valori, secondo la numerazione alfabetica greca antica. Ad esempio $\kappa = 20$, $\vartheta = 9$, dunque $\kappa\vartheta = 29$ denota un angolo di 29° , mentre il segno ζ' indica il mezzo grado. La lunghezza $\text{chord}(\theta^\circ)$ della corda sottesa (εὐθειῶν, euthèion) è riportata in base sessagesimale fino alla seconda cifra frazionaria, dunque fino all'ordine di $1/3600 = 1/60^2$; ad esempio $\lambda = 30$, $\beta = 2$, $\mu\delta = 44$, ovvero una lunghezza di $30 + 2/60 + 44/60^2 = 30,04(5)$. La terza colonna (ἑξήκοστων, hexekostòn, la sessantesima parte) riporta invece il numero medio di sessantesimi dell'unità

$$\frac{\text{chord}((\theta+1/2)^\circ) - \text{chord}(\theta^\circ)}{30}$$

che devono essere aggiunti a $\text{chord}(\theta^\circ)$ per ottenere la lunghezza della corda $\text{chord}(\theta^\circ 1')$ sottesa all'angolo corrispondente aumentato di un minuto primo, per ogni angolo tra θ° e $(\theta + 1/2)^\circ$ ad intervalli di un minuto primo. L'assenza della cifra corrispondente in base sessagesimale viene riportata mediante un cerchio 'ο', ad indicare che lo spazio nella relativa colonna della tavola è lasciato intenzionalmente vuoto.

dimostrata, compendio di aritmetica ed algebra ove si danno, inter alia, i rudimenti del calcolo aritmetico con lo zero. Tra questi è senz'altro degna di nota — seppure incompatibile con la moderna aritmetica — la regola

$$\S 23 \text{ [...] zero, diviso per zero, dà nulla }^{12}$$

in simboli solitamente interpretata con $0/0 = 0$.

La *Brāhmasphuṭasiddhānta* era già nota al matematico (anche astronomo, astrologo e cartografo)

¹²[9], p 339, trad. dall'inglese: L. D. S.. Si noti l'uso, nell'originale inglese, di *cipher* per zero, distinto da *nought* per nulla.

persiano Muḥammad ibn Mūsā al-Ḥwārizmī¹³ (ca 780 – ca 850 d. C.) autore del celebre *Libro compendiario sul calcolo mediante completamente e bilanciamento*¹⁴ e di altri trattati¹⁵ di matematica. Su questi si formò quasi certamente il già citato Leonardo Pisano *filius* Bonacij, cui si attribuisce il merito di aver introdotto¹⁶ lo zero in Europa, ovvero nell’Europa cristiana, dal momento che l’idea di zero aveva quasi certamente già da tempo raggiunto la Spagna omayyade.

Uso indiano e nell’Europa medioevale Come si evince, in ambito europeo, dal maestro Jacopo da Firenze († ca 1320)

*Et dovete sapere chel zeuero per se solo non significa nulla ma è potentia di fare significare*¹⁷

è chiara sin da Brahmagupta la funzione, propria del solo zero, di alterare il significato delle altre nove cifre, ed altresì il fatto che esso, preso a sé stante, non designa alcun numero. A riprova si dà che il termine sanscrito *añka*, col significato (tra gli altri¹⁸) di *segno* o *cifra*, denota anche il numero *nove* dunque presumibilmente, per estensione, l’insieme delle nove cifre decimali [6], p. 196, da cui lo zero risulta escluso. In tal

¹³Orig. arabo: محمد بن موسى الخوارزمي. *al-Ḥwārizmī*: dalla regione persiana del Ḥwārizm, ital. Corasmia, parte degli attuali Uzbekistan, Kazakistan e Turkmenistan. Dalla storpiatura di questa *nisba* (nell’onomastica araba: il luogo di appartenenza o di provenienza geografica), latinizzata in *Algoritmi*, deriva l’italiano *algoritmo*.

¹⁴Orig.: الكتاب المختصر في حساب الجبر والمقابلة *al-kitāb al-muḥtaṣar fī ḥisāb al-ğabr wa-l-muqābāla*, dove il *completamento*, *الْجَبْر* *al-ğabr* (letteralmente: *ricomposizione*, spec. di fratture ossee, si veda [10], pp. 177 ss, per una spiegazione dettagliata del termine), da cui l’italiano *algebra*, denota l’aggiunta di un termine additivo ad ambo i lati di un’equazione a fini di cancellazione.

¹⁵Tra questi l’*Algoritmi de numero Indorum*, ovvero *Al-Ḥwārizmī sui numeri indiani*, pervenutoci solo in traduzione latina. Per un resoconto dettagliato dell’opera di Al-Ḥwārizmī si veda [11].

¹⁶Più propriamente, è di Fibonacci il primo tentativo europeo di introdurre un calcolo sistematico con lo zero, sebbene la cifra in sé fosse probabilmente già nota.

¹⁷Jacopo da Firenze, *Tractatus algorismi*, 1307. Cit. in [13], p.5, si veda anche [14].

¹⁸Anche col significato di *uno*, si veda [3] alla voce *añka*, in particolare le accezioni L1419 e L1422, oppure col significato di *gancio* (L1409) o *curva* (L1411) o *gomito* (L1414), da cui il greco ἀγκύλος *ankylos*, *ricurvo*, il latino *angulus* e l’italiano *angolo*.

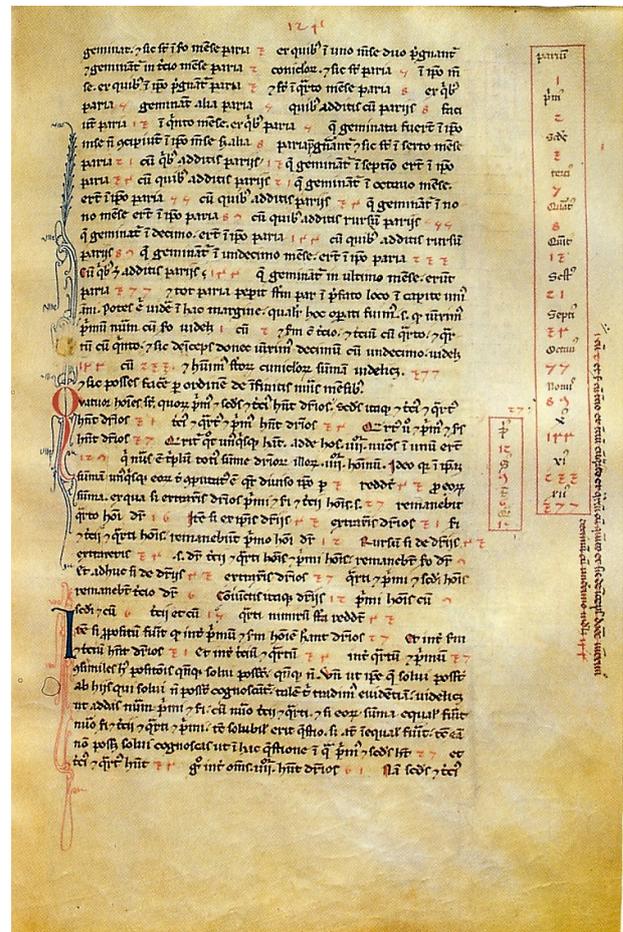


Figura 1: Il celebre passo sulla riproduzione dei conigli dal Liber Abbaci di Fibonacci e i primi valori 1, 2, 3, 5, 8, 13, 21, ecc. dell’omonima successione numerica, indicati a lato (Museo della Biblioteca Nazionale di Firenze, Codice magliabechiano Conv. Soppr. C 1, 2616, fol. 124^r. Per la versione a stampa si veda [2], p. 284). Fonte: [12].

senso sono da leggersi anche il Fibonacci, per il quale lo zero è *signum*, non *figura*, ed i versi del poeta indiano Bihārī (1595 – 1663 d. C.):

*Il punto [tilaka]¹⁹ sulla fronte di lei incrementa la sua bellezza di dieci volte così come il punto-vuoto [śūnya-bindu] incrementa un numero di dieci volte.*²⁰

Come si evince dalla citata regola di Brahmagupta, se il *śūnya* indiano — come pure lo *zephīro* di Fibonacci — non è ancora numero, esso è tuttavia un’entità con cui dover “fare i conti”.

Per Brahmagupta, il cui fine ultimo è di ricondurre l’enunciazione della ‘dottrina di Brahma’

¹⁹Nella tradizione induista, segno pigmentale decorativo ed identificativo usualmente portato sulla fronte.

²⁰Bihārī, cit. in [15], p. 81, nota 2, trad. dall’inglese: L. D. S..

ad un calcolo rigoroso, il far discendere l'aritmetica allora conosciuta da regole elementari assume una connotazione sacra, che come tale non può prescindere da un concetto di nulla —

*L'esistenza, nella più antica età degli Dèi,
[scaturì dalla Non-esistenza].*²¹

Rgveda, x, LXXII, v. 5 (secc. xx – xv a. C.)

— antitetico a quello di Melisso.

Al contrario, Fibonacci — e in qualche misura anche al-Hwārizmī — si pone un fine assai più pragmatico: di istruire il mercante pisano, suo lettore ideale, nell'acquisto di merce e moneta. Materia centrale del Liber Abbaci sono infatti *l'acquisto e la vendita delle merci e simili, i baratti delle merci, l'acquisto di monete e simili* e persino *le società fatte tra consoci* (capitoli VIII – X) — così il capitolo VIII del Libro compendiale è باب المعاملات *bāb al-muāmalāt*, ovvero il *capitolo delle transazioni*.

In ultima analisi, proprio questa necessità di calcolare con lo zero segna l'inizio della sua storia come concetto distinto dal 'segnaposto vuoto' dei sistemi numerici più antichi.

Uso presso i Maya Parallelamente ed indipendentemente dal pensiero indo-arabo-europeo, lo zero fu anche un'invenzione delle civiltà dell'America precolombiana, in particolare dei Maya (cf. [16]), presso i quali esso costituiva non solo un segnaposto nel sistema numerico in uso, bensì anche il primo numero a partire dal quale iniziare il computo, ad esempio dei giorni del mese. Questi ultimi erano denotati da una serie di glifi antropomorfi di valenza rituale, incluso uno per lo zero²², indicato anche da glifi accessori²³, come  (probabilmente un quadrifoglio, cf. [16], v. Fig. 2).

²¹Orig. sanscrito: *devānām yuge prathame.asataḥ sad-ajāyata*, dove *asataḥ* denota *ciò che non esiste*. Altri traducono: *La realtà [...] scaturì dalla non-realtà*.

²²Ovvero il glifo T1085, cf. [16, Fig. 5].

²³V. [16, *passim*] e in particolare p. 53: *This range of visual expression for a single number or word is standard within the Maya hieroglyphic corpus made by scribes who delighted to form their written language out of syllables, words, and numerical signs that intertwine and change like the animate elements of the tropical, mountainous, and coastal region in which they lived*.

In effetti, i Maya disponevano anche di un sistema numerico avanzato, più adatto al calcolo, classificabile²⁴ come cumulativo-posizionale a base mista 5 e 20, dove l'unità era indicata con '·', il numero 5 con '—' e lo zero solitamente denotato con una conchiglia²⁵ '☉'; il numero 1217 si sarebbe scritto ad esempio:

$$\begin{array}{|c|} \hline \dots \\ \hline \text{☉} \\ \hline \text{—} \\ \hline \text{·} \\ \hline \end{array} = \begin{array}{l} 3 \cdot 1 \cdot 20^2 + \\ 0 \cdot 20^1 + \\ (2 \cdot 1 + 3 \cdot 5) \cdot 20^0 \end{array} = 1217.$$

Simboli per lo zero Storicamente, l'introduzione di un simbolo atto a denotare un concetto algebrico è in genere di molto successiva a quella del concetto stesso: la notazione sintetica e largamente univoca cui siamo abituati è da considerarsi una conquista moderna e contemporanea.

La tradizione filologica iniziata dal filologo e storico tedesco Georg Nesselmann (1811 – 1881) distingue in tal senso un'*algebra retorica* (*rethorische Algebra*), caratterizzata dall'uso di lunghe perifrasi per indicare le più semplici formule, cui si sostituì via via un'*algebra sincopata* (*synkopirte Algebra*), ove si affiancava all'uso del linguaggio naturale quello sistematico di abbreviazioni per denotare i concetti di uso più frequente, ed infine la moderna *algebra simbolica* (*symbolische Algebra*).²⁶ Celeberrimo esempio italiano della prima è il *Capitolo di cose e cubo eguale a numero*²⁷ di Scipione del Ferro (1465–1526 d. C.), formulazione²⁸ dell'equazione di terzo grado $x^3 + px = q$.

²⁴V. [6, p. 12].

²⁵Presumibilmente appartenente alla famiglia delle *Olividae* (cf. [16], p.11) o delle *Cypraeidae*. V. anche Figura 2.

²⁶V. [17, p. 302]. Si ritiene oggi che questa suddivisione sia incompleta, poiché essa non tiene conto della cosiddetta *algebra geometrica*, ovvero della trattazione e/o dimostrazione di risultati algebrici mediante l'ausilio di figure della geometria. Si consideri ad esempio la nota dimostrazione grafica del calcolo del quadrato della somma di due termini.

²⁷Ove 'capitolo' denota la somma delle incognite x , 'cose', al cui numero dato p ci si riferisce come alla *quantità delle cose*, e del cubo di un'incognita medesima.

²⁸Riportata nelle note di tale cavalier Pompeo Bolognetti († 1568 d. C.), lettore ad *praxim mathematicae* nello *Studium* bolognese. V. [18, p.157].

Non deve dunque sorprendere il fatto che la prosa di al-Hwārizmī sia del tutto priva di simboli convenzionali, persino per denotare le cifre decimali, tanto che i numeri vi sono scritti per esteso ([7] p. 288, [10], *passim*). Tuttavia, è ragionevole supporre che i calcoli venissero effettuati con l'ausilio di una notazione simbolica concisa, all'interno della quale si rendeva necessario un simbolo per lo zero.



Figura 2: Lato est della stele C di Quiriguá (parte superiore). I primi sette glifi A2, B2, A3, fino ad A5 della stele, successivi al glifo introduttivo A1, riportano una data del cosiddetto Lungo computo del calendario Maya, un sistema di calcolo del tempo in base mista, la cui unità più piccola è il giorno, K'in. Seguono cicli di 20 giorni, uinal; di 18 uinal, ovvero 360 giorni, tun; di 20 tun, ovvero 7200 giorni, k'atun; infine di 20 k'atun, ovvero 144 000 giorni, b'ak'tun. I Maya ritenevano l'universo dinamico e soggetto a creazioni e distruzioni periodiche ad opera degli Dèi: tredici b'ak'tun corrispondono alla durata di uno di questi cicli. La stele riporta 13 b'ak'tun 0 katun 0 tun 0 uinal 0 k'in 4 Ahau 8 Kumku, corrispondente all'11 Agosto dell'anno 3114 a. C., la data della seconda più recente creazione del mondo. Il glifo , parte del glifo B3, è uno dei simboli usati nel Lungo computo per indicare lo zero. Lo stesso è denotato invece nei glifi B2, A3 e A4 da una mano recante una conchiglia (cf. [16], Fig. 3). Fonte: [19].

Dei diversi simboli utilizzati dalle varie culture nel corso dei secoli, i principali tre ben esemplifi-

cano le complementari concezioni antiche dello zero (e del vuoto):

‘ ’ la spaziatura, mero segno grafico con il quale palesare l'assenza della cifra corrispondente ad una data posizione nei sistemi numerici posizionali, ad esempio in quello assiro-babilonese antico (ovvero prima dell'introduzione del citato segno , v. [6], p. 251);

- *sūnya-bindu*, il *punto-vuoto*²⁹, in arabo denotato con •, ovvero, almeno inizialmente, una cifra segnaposto, e successivamente un carattere modificatore delle cifre circostanti;

- *xīng*, il carattere Zétiān³⁰ per *stella*, successivamente entrato nell'uso, in Cina e Giappone, come simbolo secondario per zero, più simile al segno indo-arabo 0 in uso in Occidente e propriamente numero in tutte le sue funzioni.

(Ri-)partiamo da zero Concludiamo la nostra breve trattazione storica con l'idea del “partire da zero”:

The story, presumably apocryphal, is that once when he [Wacław Sierpiński] was traveling, he was worried that he'd lost one piece of his luggage. «No, dear!» said his wife, «All six pieces are here.» «That can't be true,» said Sierpiński, «I've counted them several times; zero, one, two, three, four, five.» [21], p. 265.

Eccezione fatta per i Maya, la concezione moderna dello zero quale primo dei numeri *naturali* è relativamente recente ed in effetti successiva all'introduzione dell'assiomatizzazione dei naturali in uso a tutt'oggi, dovuta a Giuseppe Peano (1858 – 1932), il quale, nel 1889, dava come primo

²⁹La prima attestazione risale (cf. [6], p. 196) al romanzo d'amore *Vāsavadattā* (ca 600 d. C.) di Subandhu (seconda metà del sec. VI d. C.), si veda [20], p. 100, nota 10 e riferimenti ivi citati.

³⁰Detto dei caratteri ideografici cinesi introdotti durante il regno (624 – 705 d. C.) dell'imperatrice Wu Zétiān (624 – 705 d. C.), oggi prevalentemente in disuso.

$$1 \in \mathbb{N}.$$

Uno dei maggiori contributi verso la concezione dello zero come ‘iniziale’ è la seguente definizione induttiva dei numeri *ordinali* proposta nel 1923 da John von Neumann (1903 – 1957):

„Jede Ordnungszahl ist die Menge der ihr vorangehenden Ordnungszahlen.“³¹

In simboli, si pone per definizione

$$\begin{aligned} 0 &= \emptyset \\ \{0\} &= 1 = \{\emptyset\} \\ \{0, 1\} &= 2 = \{\emptyset, \{\emptyset\}\} \\ \{0, 1, 2\} &= 3 = \{\emptyset, \{\emptyset, \{\emptyset\}\}\} \\ \{0, 1, 2, 3\} &= 4 = \{\emptyset, \{\emptyset, \{\emptyset, \{\emptyset\}\}\}\} \end{aligned}$$

e la relativa generalizzazione $\kappa := [0, \kappa)$ per ogni ordinale κ .

Limitandosi a considerare i numeri naturali, questi possono essere riguardati come le cardinalità degli insiemi finiti presi a meno di equivalenza (ovvero a meno di biiezione). È dunque naturale che si prenda inizio da 0, che denota la cardinalità dell’insieme vuoto, seguito da 1, la cardinalità di ogni insieme con un unico elemento, e così via.

2 Uso moderno

Nella matematica contemporanea sono molti gli enti cui si dà, per ragioni diverse, il nome di ‘zero’, come ad esempio l’elemento neutro di una qualsiasi operazione associativa e commutativa. Di seguito esponiamo, in modo informale, le proprietà di alcuni di questi enti utilizzando il linguaggio della teoria delle categorie.

Neutralità e assorbenza

§21 *Negativo, meno zero, dà negativo, positivo* [meno zero], [dà] *positivo, zero* [meno zero], [dà] *nulla*. [9], p. 339, trad.: L. D. S.

³¹“Ogni numero ordinale coincide con l’insieme dei numeri ordinali ad esso precedenti.” [23], p. 199. Trad.: L. D. S.

Sin dalle prime apparizioni dei rudimenti del calcolo con gli interi è noto che per ogni intero a

$$a + 0 = a = 0 + a.$$

In termini moderni, 0 è l’*elemento neutro della somma* nell’anello \mathbb{Z} ³² dei numeri interi. Per questo motivo si è soliti denotare con lo stesso simbolo l’elemento neutro dell’operazione di somma in un qualsiasi anello, qui indicato invece con **0**.

Analogamente, si denota solitamente con 1, qui con **1**, l’*elemento neutro*, qualora presente, del prodotto di un anello. In tal caso, quest’ultimo si dice *anello con unità*.

Come noto, la neutralità dello zero è legata al concetto di ‘somma vuota’. Comunque dato un insieme J di indici ed un insieme di interi $A := \{a_j\}_{j \in J}$ siamo soliti denotare con il simbolo³³ $\sum_{j \in J} a_j$ la somma degli elementi di A . Se J , dunque A , è l’insieme vuoto, poniamo per convenzione

$$\sum_{j \in \emptyset} = 0.$$

In modo analogo, il prodotto vuoto si fa convenzionalmente coincidere con l’unità, elemento neutro del prodotto in \mathbb{Z} ; in simboli:

$$\prod_{j \in \emptyset} a_j = 1,$$

da cui discende anche la convenzione $0! = 1$, dove $n!$ denota il *fattoriale* di n , definito come il prodotto dei numeri naturali da 1 ad n .

Come si vedrà più avanti, questa scelta, convenzionale nel contesto degli anelli, può in altri contesti essere propriamente compresa nelle definizioni di somma e prodotto.

Una seconda proprietà fondamentale dello 0, anch’essa già enunciata da Brahmagupta, [9] p. 339, è quella di essere l’unico elemento

³² \mathbb{Z} , dal tedesco *Zahlen*, *numeri*.

³³A rigore, tale scrittura è non ambigua grazie alla commutatività della somma in \mathbb{Z} .

assorbente per il prodotto di \mathbb{Z} , ovvero tale che

$$a \cdot 0 = 0 = 0 \cdot a$$

per ogni intero a . Nel lessico della teoria degli anelli si dirà che l'insieme $\{0\}$, costituito dal solo 0 , è un *ideale* dell'anello \mathbb{Z} , ovvero un sottoinsieme chiuso rispetto alla somma ed assorbente rispetto al prodotto.

Vedremo nella prossima sezione come tutte queste proprietà dello zero possano essere rilette anche in contesti più generali.

Oggetti iniziali, terminali e zero-oggetti

Tendenza di parte considerevole della matematica degli ultimi due secoli è stata quella di spostare l'attenzione dagli enti alle relazioni che intercorrono tra di essi. Parte di questa tendenza è lo sviluppo della *teoria delle categorie*, all'interno della quale ha trovato una sua rielaborazione anche il concetto di 'zero iniziale'.

Informalmente, una *categoria* è il dato di una collezione $\text{ob}(\mathcal{C})$, i cui elementi sono detti *oggetti* (di \mathcal{C}), e di una collezione $\text{hom}(\mathcal{C})$, i cui elementi sono detti *morfismi* (o *mappe* o *freccie*, di \mathcal{C}) e soddisfano le seguenti proprietà. Ogni morfismo f ha associati un unico oggetto d'origine a ed un unico oggetto di destinazione b , entrambi in $\text{ob}(\mathcal{C})$. Un tale morfismo si indica solitamente con $f: a \rightarrow b$, e si denota con $\text{hom}_{\mathcal{C}}(a, b)$ la collezione di tutti i morfismi di \mathcal{C} con origine a e destinazione b . Ai morfismi si richiedono alcune proprietà di composizione, tra cui l'esistenza di un'identità $\text{id}_a: a \rightarrow a$ per ogni $a \in \text{ob}(\mathcal{C})$ e l'associatività di tale operazione, denotata con

$$\circ: \text{hom}_{\mathcal{C}}(a, b) \times \text{hom}_{\mathcal{C}}(b, c) \rightarrow \text{hom}_{\mathcal{C}}(a, c),$$

e tale che $g \circ f \in \text{hom}_{\mathcal{C}}(a, c)$ per ogni scelta dei morfismi $f \in \text{hom}_{\mathcal{C}}(a, b)$ e $g \in \text{hom}_{\mathcal{C}}(b, c)$ e per ogni tripla di oggetti $a, b, c \in \text{ob}(\mathcal{C})$.

Un oggetto o di una categoria \mathcal{C} si dice

- *iniziale* (originariamente indicato da alcuni autori con $\mathbf{0}$, più modernamente con \emptyset) se

per ogni oggetto a di \mathcal{C} esiste uno ed un solo morfismo f in $\text{hom}_{\mathcal{C}}(o, a)$;

- *terminale* (a volte indicato con $\mathbf{1}$) se per ogni oggetto a di \mathcal{C} esiste uno ed un solo morfismo f in $\text{hom}_{\mathcal{C}}(a, o)$;
- *zero* se esso è sia iniziale che terminale.

Consideriamo l'insieme vuoto, denotato con \emptyset , e la *mappa vuota*, l'unica mappa tra insiemi il cui dominio è l'insieme vuoto, anch'essa denotata con il medesimo simbolo. L'insieme vuoto è, mediante la mappa vuota, l'oggetto iniziale della fondamentale categoria degli insiemi Set , i cui oggetti sono gli insiemi e i cui morfismi sono le mappe tra insiemi. Nella stessa categoria, ogni insieme costituito da un solo elemento è un oggetto terminale, mediante la relativa mappa costante.

Nella categoria degli insiemi si dà inoltre la seguente definizione di *prodotto* (la quale è possibile generalizzare ad ogni categoria), ivi detto *cartesiano*:

dati un insieme J di indici ed una collezione $\mathcal{A} := \{A_j\}_{j \in J}$ di insiemi, il prodotto cartesiano $\prod_{j \in J} A_j$ di \mathcal{A} è l'insieme

$$\left\{ f: J \rightarrow \prod_{j \in J} A_j \mid f(j) \in A_j \right\}.$$

In questo contesto, la definizione di prodotto include il caso in cui $J = \emptyset$, per il quale si ha

$$\prod_{j \in \emptyset} A_j = \left\{ f: \emptyset \rightarrow \prod_{j \in J} A_j \right\} = \{\emptyset\}.$$

Ne segue che il prodotto vuoto è l'insieme con unico elemento la mappa vuota e dunque un oggetto terminale, o $\mathbf{1}$, della categoria Set .

Possiamo rileggere in questo contesto anche la proprietà di assorbimento dello zero. Infatti, scegliendo $J \neq \emptyset$ ed $A_j = \emptyset$ per qualche $j \in J$ nella definizione precedente, si ha

$$\prod_{j \in J} A_j = \emptyset$$

poichè non esiste alcuna mappa di insiemi che abbia come dominio J e come codominio l'insie-

me vuoto. D'altra parte, se in questo caso anche J fosse vuoto, si avrebbe allora

$$\prod_{j \in \emptyset} A_j = \{\emptyset\}$$

poichè esiste unica la mappa di insiemi di dominio e codominio vuoti, cioè la mappa vuota. Quest'ultimo esempio corrisponde, nel contesto dei numeri interi, all'operazione di elevamento a potenza 0^0 , cui si attribuisce convenzionalmente il valore 1.

Si noti da un lato come le precedenti osservazioni motivino la scelta di indicare gli oggetti iniziali con 0 , quelli terminali con 1 . Dall'altro, le definizioni date sono tali per cui le proprietà della somma e del prodotto vuoti discendono direttamente da esse.

Come ultimo esempio, consideriamo ora la categoria degli anelli non unitari Rng , i cui oggetti sono gli anelli non unitari ed i cui morfismi sono gli omomorfismi di anelli. Indichiamo con $\{0\}$ l'anello banale, costituito dal solo elemento 0 . In modo analogo a quanto detto per Set , detto *morfismo nullo* l'unico omomorfismo di anelli con dominio assegnato che assume identicamente il valore 0 nell'anello codominio, l'anello banale è, mediante il morfismo nullo, l'oggetto zero della categoria Rng .

Tuttavia, quando ci limitiamo a considerare la sottocategoria Ring degli anelli con unità (ai cui oggetti si richiede di essere anelli con unità, 1 , non necessariamente distinta da 0 e ai cui morfismi si richiede di preservare 1), la situazione è differente. L'oggetto iniziale di questa categoria è infatti l'anello degli interi \mathbb{Z} , mentre l'anello banale è soltanto oggetto terminale, sempre attraverso il morfismo nullo.

3 Difficoltà legate all'uso e all'apprendimento dello zero alla luce della sua evoluzione storica

Ripercorrere l'evoluzione storica di nozioni matematiche può essere utile per capire alcune difficoltà che gli studenti incontreranno nell'appropriarsi della nozione stessa, come affermava Kline nel 1970, parlando, in particolare, della nascita del concetto di numero negativo:

There is not much doubt that the difficulties the great mathematicians encountered are precisely the stumbling blocks that the students experience and that no attempts to smoothen these difficulties with logical verbiage will succeed. If it took mathematicians 1000 years from the time that first class mathematics appeared to arrive at the concept of negative numbers, and it did, and if it took another 1000 years for mathematicians to accept negative numbers, as it did, we may be sure that students will have difficulties with negative numbers. Moreover, the students will have to master these difficulties in about the same way that the mathematicians did, by gradually accustoming themselves to the new concepts, by working with them and by taking advantage of all the intuitive support that the teacher can master [24], p. 270.

M. Kline (1908 – 1992)

L'apprendimento dello zero in età prescolare

In questa sezione conclusiva, vogliamo sottolineare alcune delle prime difficoltà che gli studenti molto giovani (ma a volte anche quelli meno giovani) incontrano nell'apprendimento dello zero, mettendo in relazione queste difficoltà con tappe dell'evoluzione storica della nozione. Ci concentreremo sulle *prime* difficoltà che i bambini incontrano, già a partire dai tre anni, ma soprattutto nei primi anni di scuola primaria. Ovviamente a queste difficoltà possono sommarsene altre legate ai significati e alle proprietà dello zero che si incontrano nei successivi livelli scolari.

Innanzitutto, gli studi di carattere cognitivo concordano sul fatto che, mentre imparare la sequenza di numeri da 1 a 9 sia in generale poco problematico, l'apprendimento dello zero sembra essere più difficoltoso [25, 26]. Considerando lo sviluppo storico della nozione, si può notare come i primi usi di segni (o spazi vuoti) in qualche modo associabili allo zero attuale riguardassero la rappresentazione di nessuna quantità o la mancanza di quantità, come i segni sviluppati nel sistema numerico assiro-babilonese e poi nel sistema greco di età post-ellenistica. Anche per bambini di tre o quattro anni questo sembra essere il principale significato associato a 'zero'. A questa età i bambini tendono a indicare la mancanza di elementi come un pugno chiuso (o entrambi i pugni chiusi), usando espressioni verbali del tipo: «Non ce ne sono», o, alcuni, magari con qualche incertezza in più: «Ce ne sono zero». Tuttavia, associare il segno '0' a tale significato sembra richiedere tempi più lunghi rispetto a quelli legati ai segni convenzionali associati a quantità non nulle. A tre o quattro anni i bambini molto più difficilmente usano il segno '0' per indicare la mancanza di oggetti, pur essendo circondati dal segno (almeno nella cultura occidentale), tanto quanto dagli altri nove segni che rappresentano le cifre, su cellulari, etichette al supermercato, calendari, quadranti di orologi digitali, ascensori, termometri, ecc.. Ad esempio, in uno studio di Ewers-Rogers e Cowan [27], bambini di tre e quattro anni dovevano annotare quantità di mattoncini in scatole di latta. Lo studio voleva investigare quali fossero le notazioni usate dai bambini a questa età come strumenti comunicativi-referenziali. I risultati furono molto simili a quelli di altri studi aventi il medesimo obiettivo: i bambini usano diverse notazioni, per lo più analogiche; alcuni usano, anche correttamente, i simboli convenzionali, ma tra questi pochissimi usano '0' per indicare la situazione di mancanza di mattoncini in una scatola. A questa età i bambini preferiscono usare espressioni verbali usando la parola 'niente'. Secondo

L. Tolchinsky [28] il segno grafico 0 non è una notazione abbastanza esplicita per molti bambini per indicare la mancanza di oggetti. Ricordiamo come anche nello sviluppo storico del concetto, vi siano state interpretazioni per cui zero non designava alcun numero, come per Jacopo da Firenze. Dunque, è possibile che lo zero non sia concepito come numero durante queste prime fasi dell'apprendimento, a differenza, per esempio, di uno o due. Questa interpretazione è coerente con la resistenza iniziale ad usare il segno 0 per indicare l'assenza di una quantità.

[...] zero è l'ultima cifra che è venuta in mente all'uomo. E non è tanto strano, perché lo zero è il numero più raffinato di tutti.

H. M. Enzensberger, *Il Mago Dei Numeri* [29].

In effetti, anche storicamente, trovare modi per indicare l'assenza di quantità è stato utile nel momento in cui serviva indicare una posizione vuota nella scrittura posizionale, facendo quindi cambiare significato ai segni circostanti, per eliminare scritture ambigue. Questa esigenza non è sentita da un bambino che gestisce solo quantità piccole.

Nei primi anni di scuola primaria, si riscontrano altre difficoltà con l'uso del segno 0, questa volta legate alla lettura di numeri a più cifre, contenenti il segno 0 per indicare posizioni vuote. Infatti, usato in questo modo, come segnaposto, lo zero è, a differenza delle altre cifre, "muto". In molte lingue, come l'italiano (la lettura di '203' non contiene la parola 'zero'), esso gioca comunque un ruolo importante cambiando la lettura (e il significato) delle altre cifre. In lingue come il Taiwanese, in cui si leggono anche gli zeri nei numeri a più cifre, pare che i bambini commettano molti meno errori nella lettura di numeri a tre o a quattro cifre rispetto ai bambini inglesi. È interessante notare che la *trasparenza* di una lingua rispetto alla notazione posizionale decimale in generale favorisca processi di traduzione dal linguaggio verbale alla notazione simbolica e viceversa [30] [28].

Negli adulti si ritrovano ancora tracce della complessità cognitiva legata al segno 0: alcuni studi mostrano che solo per leggere 0 all'interno di numeri sono richiesti negli adulti tempi eccezionalmente lunghi rispetto a quelli richiesti per leggere numeri senza la cifra 0 [31]. Tempi eccezionalmente lunghi si riscontrano anche nella scrittura di numeri a più cifre contenenti 0 come cifra. I bambini commettono più frequentemente errori nella scrittura dei numeri che hanno 0 tra le loro cifre; inoltre, quando devono scrivere 0 i bambini spesso eseguono movimenti differenti rispetto a quando scrivono le altre cifre [32].

Le difficoltà più significative nella gestione del segno 0 riguardano il suo uso nella notazione posizionale e, in particolare, l'interiorizzazione del fatto che il valore di una cifra cambia in base alla posizione che occupa all'interno di un numero. Anche in pazienti con lesioni cerebrali, lo zero è oggetto di errori selettivi rispetto al ruolo sintattico o lessicale che assume nel numero [33]. Nella nostra discussione storica, abbiamo visto tornare più volte, e in diverse parti del mondo, l'accostamento del riferimento a 'niente' o 'quantità nulla' con il riferimento alla possibile alterazione del significato di altre cifre. Alla luce di quanto detto da Kline, tali difficoltà non devono stupire.

La linea dei numeri Finora abbiamo discusso le prime principali difficoltà con l'uso del segno 0 in relazione al significato di mancanza e in relazione al significato di segnaposto nella notazione posizionale. Concludiamo questa trattazione con un terzo significato dello zero, che si lega anche al suo ruolo nel calcolo aritmetico: zero come punto di partenza nell'allineamento dei numeri (naturali). È interessante notare che questo significato è comparso per la prima volta nel sistema numerico utilizzato dai Maya, senza prendere piede in altre culture dove, invece, lo sviluppo del concetto e di segni ad esso associato è passato per gli altri due significati discussi e per la necessità di eseguire calcoli, come in Fibonacci.

A livello cognitivo, la linea dei numeri sembra corrispondere ad una rappresentazione intuitiva e ad una traduzione naturale della sequenza in una dimensione spaziale [34, 35]. Il modello della linea dei numeri trova conferma in vari effetti comportamentali come l'effetto SNARC (*Spatial Numerical Association of Response Codes*) usato per documentare l'effetto dello spazio sulla rappresentazione dei numeri [36]. Il modello mentale della linea dei numeri costituisce al tempo stesso una forma più astratta di rappresentazione rispetto al conteggio degli oggetti, in quanto avvia la possibilità di contare qualsiasi elemento. Sembra che la rappresentazione della linea dei numeri evolva con lo sviluppo cognitivo dell'individuo, anche in seguito alle influenze culturali [37], e che quindi un'appropriata esposizione porti ad una migliore rappresentazione e conseguente elaborazione numerica.

Una buona rappresentazione mentale della linea dei numeri consente un accesso rapido ed efficace alle informazioni numeriche necessarie per moltissimi compiti sia numerici che aritmetici. Alcune ricerche [38, 39] suggeriscono che l'uso della linea numerica visiva, come rappresentazione esterna di supporto, favorisce la manipolazione di quantità e l'acquisizione e costruzione di concetti e procedure matematiche più avanzate. Tuttavia, è noto che fornire modelli fisici di linee dei numeri contenenti la tacca dello 0 può provocare errori legati al conteggio delle tacche: invece che contare i segmenti, magari considerandone l'estremo destro (la tacca che lo segna), alcuni bambini continuano a contare le tacche, inclusa la prima, cioè quella dello 0, dicendo «uno». Quindi vi è uno sfasamento tra i numeri pronunciati durante il processo di conteggio (o di calcolo) e la posizione corretta di tali numeri sulla linea. Per gestire correttamente rappresentazioni della linea dei numeri, che eventualmente diventeranno rappresentazioni della retta reale, sembra utile passare per l'idea di zero come 'punto di partenza'. Nell'uso di modelli fisici della linea dei numeri, questo significato

può essere costruito usando delle pedine, la cui posizione di partenza è la tacca dello 0 [40].

Ci è parso interessante mettere in relazione i risultati discussi di studi in ambito cognitivo sulle difficoltà legate alla gestione dello zero con aspetti dello sviluppo storico del concetto e di segni ad esso associati. Tali risultati sembrano essere assolutamente coerenti con quanto affermato da Kline, e dovrebbero essere presi in considerazione da qualunque maestro (o guida didattica³⁴) per costruire e rendere più fondato e solido il suo 'supporto intuitivo' ('*intuitive support*') da fornire ai giovani allievi che si affacciano alla nozione di zero.

Riferimenti bibliografici

- [1] C. Seife: *Zero – The Biography of a Dangerous Idea*. Penguin Books, Londra (2000).
- [2] L. Fibonacci: *Il Liber abbaci di Leonardo Pisano pubblicato secondo la lezione del Codice Magliabechiano C. 1, 2616, Badia Fiorentina, n. 73 da Baldassarre Boncompagni, socio ordinario dell'Accademia pontificia de' nuovi Lincei*. Tipografia delle scienze matematiche e fisiche, Roma (1857). <http://www.e-rara.ch/zut/doi/10.3931/e-rara-34354>.
- [3] Sir M. Monier-Williams: *A Sanskrit–English dictionary etymologically and philologically arranged with special reference to Greek, Latin, Gothic, German, Anglo-Saxon, and other cognate Indo-European languages*. The Clarendon Press, 2nd edition, Oxford (1899). <http://www.sanskrit-lexicon.uni-koeln.de/scans/MWScan/2014/web/index.php>.
- [4] H. Diels, W. Kranz: *Die Fragmente der Vorsokratiker*. Weidmannsche Buchhandlung, Berlin (1903). <https://archive.org/details/diefragmenteder00krangoog>.
- [5] Euclide di Megara: *Euclide megarense acutissimo philosopho, solo introduttore delle scienze matematiche. Diligentemente rassettato, et alla integrità ridotto, per il degno professore di tal scienze Nicolo Tartalea brisciano. Secondo le due tradottioni. Con vna ampla esposizione dello istesso tradottore di nuouo aggiunta*. C. Trojano, Venezia (1565). <http://gallica.bnf.fr/ark:/12148/bpt6k61831x.r=Euclide?rk=236052;4>.
- [6] S. Chrisomalis: *Numerical Notation – A comparative History*. Cambridge University Press, Cambridge (2010).
- [7] C. B. Boyer *A History of Mathematics*. John Wiley & Sons, Inc., 2nd edition, New York (1991).
- [8] C. Tolomeo: *Claudii Ptolomaei opera quae exstant omnia – edidit J. L. Heiberg, Professor Hauniensis*. B. G. Teubner, Lipsia (1898).
- [9] H. T. Colebrooke: *Algebra, with Arithmetic and Mensuration, from the Sanscrit of Brahmegeupta and Bhāscara*. John Murray, London (1817). <https://archive.org/details/algebrawitharith00brahuoft>.
- [10] Muḥammad ibn Mūsā al-Ḥwārīzīmī: *The Algebra of Mohammed ben Musa – edited and translated by Frederic Rosen*. J. L. Cox, London (1831).
- [11] N. Ambrosetti: *L'eredità arabo-islamica nelle scienze e nelle arti del calcolo dell'Europa medievale*. LED Edizioni Universitarie, Milano (2008).
- [12] Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Liber_abbaci_magliab_f124r.jpg, consultato in data 14 Ottobre 2017.
- [13] G. Gaetano: *Dell'origine delle parole zero e cifra*. Società Anonima Cooperativa Tipografica, Napoli (1903).
- [14] J. Høyrup *Jacobus de Florentia, Tractatus algorismi (1307), the chapter on algebra (Vat. Lat. 4826, fols 36^v–45^v)*, Centaurus, 2000.
- [15] B. Swami, A. N. Singh: *History of Hindu mathematics: a source book*. Asia Publishing House, Seul (1962).
- [16] A. Blume: *Maya Concepts of Zero*, Proc. Am. Phil. Soc., **155(1)** (2011) 51–88.
- [17] G. H. F. Nesselmann: *Versuch einer kritischen Geschichte der Algebra*. G. Reimer ed., Berlin (1842).
- [18] E. Bortolotti: *L'algebra nella scuola matematica bolognese del secolo XVI*, Periodico di Matematiche, **5(3)** (1925) 147–192.
- [19] Wikimedia Commons. https://commons.wikimedia.org/wiki/File:East_side_of_stela_C,_Quirigua.PNG, consultato in data 14 Ottobre 2017.
- [20] V. Subandhu: *A Sanskrit romance by Subandhu – translated, with an introduction and notes, by Louis H. Gray, Ph.D.*. Columbia University Press, New York (1913). <https://archive.org/details/vasavadattasansk00suba>.
- [21] J. H. Conway, R. K. Guy: *The book of numbers*. Springer-Verlag, New York (1996).
- [22] G. Peano: *Aritmetices Principia – nova methodo exposita*. F.lli Bocca, Roma, Firenze (1889). <https://archive.org/details/aritmeticespri00peangoog>.
- [23] J. von Neumann: *Zur Einführung der transfiniten Zahlen*, Acta litterarum ac scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio scientiarum mathematicarum, **1** (1923) 199–208.
- [24] M. Kline: *Logic versus Pedagogy*, The American

³⁴Per esempio, durante il progetto PerContare è stata sviluppata una guida didattica per la classe prima primaria che tiene conto di tali difficoltà legate alla nozione di zero (v. [40, 41]).

Mathematical Monthly, 77(3) (1970) 264–281.

- [25] M. Hughes: *Children and number: Difficulties in learning*. Wiley–Blackwell, New York (1986).
- [26] H. M. Wellman, K. F. Miller: *Thinking about nothing: Development of concepts of zero*, British Journal of Developmental Psychology, 4 (1986) 31–42.
- [27] J. Ewers-Rogers, R. Cowan: *Children as apprentices to number*, Early childhood development and care, 125(15) (1996) 15–17.
- [28] L. Tolchinsky: *The Cradle of Culture and What Children Know About Writing and Numbers Before Being Taught*. Developing Mind Series, Psychology Press (2003).
- [29] H. M. Enzensberger: *Il mago dei numeri*. Giulio Einaudi Editore, Torino (1997).
- [30] S. Dehaene: *The number sense*. Oxford University Press, Oxford (1997).
- [31] M. Brysbaert: *Arabic number reading: On the nature of the numerical scale and the origin of phonological reading*, Journal of Experimental Psychology: General, 124 (1995) 343–447.
- [32] A. Lochy, A. Pillon, P. Zesiger, X. Seron: *Verbal structure of numerals and digits in handwriting: New evidence from kinematics*, The Quarterly Journal of Experimental Psychology, 55A (2002) 263–288.
- [33] A. Granà, A. Lochy, L. Girelli, X. Seron, C. Semenza: *Transcoding zero within complex numerals*, Neuropsychologia, 41(12) (2003) 1611–1618.
- [34] P. Pinel, M. Piazza, D. Le Bihan, S. Dehaene: *Distributed and overlapping cerebral representation of number, size, and luminance during comparative judgments*, Neuron, 41(6) (2004) 983–993.
- [35] X. Seron, M. Pesenti, Mauro, M-P. Noël, G. Deloche, J-A. Cornet: *Images of numbers or “When 98 is upper left and 6 sky blue”*, Cognition, 44 (1992) 159–196.
- [36] S. Dehaene, S. Bossini, P. Giroux: *The mental representation of parity and numerical magnitude*, Journal of Experimental Psychology: General, 122 (1993) 371–396.
- [37] M. Zorzi, K. Priftis, C. Umiltà: *Neglect disrupts the mental number line*, Nature, 417 (2002) 138–139.
- [38] S. Dehaene: *Subtracting Pigeons: Logarithmic or Linear?*, Psychological Science, 12 (2001) 244–246.
- [39] Y. Okamoto, Yukari, R. Case: *The Role of Central Conceptual Structures in the Developments of Children’s Thought, volume 1-2, Monographs of the Society for Research in Child Development*. Blackwell, Malden (1996).
- [40] A. Baccaglini-Frank: *Trattamento dello zero nel Progetto PerContare*, L’Insegnamento della Matematica e delle Scienze Integrate, 37A(3) (2014) 257–282.
- [41] A. Baccaglini-Frank, M. G. Bartolini Bussi: *Buone pratiche didattiche per prevenire falsi positivi nelle diagnosi di discalculia: Il progetto PerContare*, Form@re, 15(3) (2016) 170–184.

Lorenzo Dello Schiavo: è dottorando in analisi stocastica presso l’Università di Bonn, Germania. I suoi interessi di ricerca riguardano principalmente gli spazi metrici di misura e la dinamica stocastica di sistemi a infinite particelle.

Anna Baccaglini-Frank: è ricercatrice in Didattica della Matematica presso l’Università di Pisa. I suoi interessi di ricerca riguardano l’apprendimento della matematica in ambienti digitali e difficoltà d’apprendimento di natura cognitiva. È autrice di numerosi articoli scientifici e di testi di supporto alla didattica, tra cui il recentissimo *Didattica della Matematica*, scritto insieme a Pietro Di Martino, Roberto Natalini e Giuseppe Rosolini, e la rubrica ‘*Strane Storie Matematiche*’, che tiene insieme a Pietro Di Martino per la rivista *Archimede*.

Etica senza verità

Dario Antiseri

LUISS – Libera Università Internazionale di Studi Sociali – Roma

1. Nel terzo libro del *Treatise of Human Nature* di David Hume c'è un passo sulla inderivabilità logica del dover essere dall'essere talmente chiaro da giustificare pienamente l'universale richiamo a Hume, a quella che, appunto, viene chiamata "legge di Hume". Ed ecco il celebre brano: «In ogni sistema di morale con cui ho avuto finora a che fare, ho sempre notato che l'autore procede per un po' nel modo ordinario di ragionare, e stabilisce l'esistenza di un bene, oppure fa delle osservazioni circa le faccende umane; quando all'improvviso mi sorprendo a scoprire che, invece di trovare delle proposizioni, rette come di consueto dai verbi è e non è, non incontro che proposizioni connesse con dovrebbe e non dovrebbe. Questo mutamento è impercettibile, ma è della massima importanza. Poiché questi dovrebbe e non dovrebbe esprimono una relazione o affermazione nuova, è necessario che si adduca una ragione di ciò che sembra del tutto inconcepibile, cioè del modo in cui questa nuova relazione può essere dedotta dalle altre, che sono totalmente diverse da essa»[1]. In realtà, questa ragione non può essere addotta, giacché da proposizioni descrittive possono venir logicamente dedotte unicamente proposizioni descrittive: l'informazione non produce imperativi. E, dunque, non si passa dall'essere al dover essere. Questa, in breve, è la legge di Hume, la grande divisione tra asserzioni indicative e asserzioni prescrittive, tra fatti e valori - un principio di fondamentale rilevanza in ambito etico, prima, e di conseguenza in campo politico.

Che esistano, che sia possibile individuare ed

enunciare, e poi razionalmente fondare valor etici universalmente validi: è questa, sostanzialmente, la pretesa centrale della tradizione giunaturalistica. Sennonché, se si pone attenzione alle diversità (esistite nel passato ed esistenti oggi) delle concezioni circa il bene e il male, e se si volge lo sguardo sulla storia delle vicende umane e degli umani conflitti, il non-cognitivista è pronto a ripetere con Pascal che «il furto, l'incesto, l'uccisione dei figli o dei padri, tutto ha trovato posto tra le azioni virtuose»; «singolare giustizia, che ha come confine un fiume! Verità di qua dei Pirenei, errore di là» [2]. Con Weber: nel campo dei valori, partendo dalla pura esperienza si giunge al politeismo, nel senso che «tra i diversi valori che presiedono all'ordinamento del mondo il contrasto è inconciliabile» ([3], p. 31). E, stando sempre al non-cognitivista, nella realtà della vita si giunge al politeismo dei valori per la fondamentale ragione che dalla prospettiva logica i valori e le norme etiche sono proposte (di «ideali di vita», di azioni «corrette», di leggi «giuste», di istituzioni «valide», ecc.) e non proposizioni indicative. L'etica non de-scrive; essa pre-scrive. L'etica non spiega e non prevede; l'etica valuta. Difatti, non esistono spiegazioni etiche. Esistono soltanto spiegazioni scientifiche; e valutazioni etiche. Né si danno previsioni etiche (o estetiche). L'etica non sa. L'etica non è scienza. L'etica è senza verità [4].

Se, come sostengono i non-cognitivisti, l'esser vero (o falso) è un predicato delle proposizioni indicative, se dunque la verità è un attributo della scienza (e di altre proposizioni indicative), questa scienza - tutta la scienza e qualsiasi altra teoria descrittiva, magari metafisica - non può

logicamente produrre etica. E non lo può per la ragione che da proposizioni descrittive non è possibile dedurre asserti prescrittivi. Conseguentemente, dall'intera scienza non è possibile spremere un grammo di morale. Tutto il sapere concepibile non produce valori, né può smentirli. Si chiede Max Weber: «Chi vorrà provarsi a "confutare scientificamente" l'etica del Sermone della Montagna, per esempio la massima: "non far resistenza al male" oppure l'immagine del "porgere altra guancia"?» ([3], p. 32).

2. La scienza sa (sempre parzialmente e in modo congetturale); l'etica valuta. L'etica non sa; la scienza non valuta. Le norme (etiche e giuridiche) non sono logicamente derivabili da asserti descrittivi. L'etica non è scienza. Di seguito, alcune autorevoli prese di posizione a favore di questa tesi, a cominciare da due fisici: Henri Poincaré e Albert Einstein.

Henri Poincaré: «Non ci può essere morale scientifica, ma non ci può nemmeno essere scienza immorale. E la ragione è semplice. E una ragione - come dire? - puramente grammaticale. Se le premesse di un sillogismo sono tutte e due all'indicativo, la conclusione sarà anch'essa all'indicativo. Perché la conclusione potesse essere all'imperativo bisognerebbe che almeno una delle due premesse fosse all'imperativo. Ora, i principi scientifici, i postulati della geometria sono e non possono essere che all'indicativo, le verità sperimentali sono tutte al modo indicativo e alla base della scienza non c'è né ci può essere altro. Ciò posto, il più sottile dei dialettici potrà fare con quei principi tutte le acrobazie che crede, combinandoli, mettendoli l'uno sull'altro; tutto quello che otterrà sarà all'indicativo. Non otterrà mai una proposizione che dica: fa questo, non far quello; che è quanto dire una proposizione che confermi o contraddica la morale. Questa difficoltà i moralisti la incontrano da molto tempo. Essi si sforzano di dimostrare la legge morale; e bisogna perdonarglielo perché quello è il loro mestiere; vogliono appoggiare la morale su qualcosa come se essa potesse poggiare su altro che non su se stessa! [...]. Ogni morale dogmatica, ogni morale dimostrativa è votata a un insicuro insuccesso; essa è come una macchina che avesse

tutte le trasmissioni di movimento e mancasse di energia motrice. Il motore morale, quello che può mettere in azione tutto l'apparecchio delle bielle e degli ingranaggi, non può essere che un sentimento» [5].

Albert Einstein: «La scienza [...] può solo accertare ciò che è, ma non ciò che dovrebbe essere, e al di fuori del suo ambito restano necessari i giudizi di valore di ogni genere» [6].

E, dopo Poincaré ed Einstein, ancora un richiamo a Max Weber e una annotazione di Karl Popper.

Max Weber: «Il medico cerca con tutti i mezzi di conservare la vita al moribondo, anche se questi implora di essere liberato dalla vita [...]. La scienza medica non si pone la domanda se e quando la vita valga la pena di essere vissuta. Tutte le scienze naturali danno una risposta a questa domanda: che cosa dobbiamo fare se vogliamo dominare tecnicamente la vita? Ma se vogliamo e dobbiamo dominarla tecnicamente, e se ciò, in definitiva abbia veramente un significato, esse lo lasciano del tutto in sospeso oppure lo presuppongono per i loro fini» ([3], p. 26-27).

Karl R. Popper: «è impossibile dedurre un'asserzione che enuncia una norma o una decisione ovvero una proposta per una politica da un'asserzione che enuncia un fatto; il che equivale a dire che è impossibile dedurre norme o decisioni o proposte dai fatti» ([7], p. 100).

«Certo, è impossibile dimostrare la giustezza di qualsivoglia principio etico o anche argomentare in suo favore esattamente allo stesso modo in cui argomentiamo in favore di un enunciato scientifico. L'etica non è scienza» ([7], p. 313).

3. In ogni caso, colui che ha posto la «legge di Hume» a base della sua grande opera di filosofia del diritto e di teoria della democrazia è stato Hans Kelsen. «La causa della democrazia - ha scritto Kelsen - risulta disperata se si parte dall'idea che sia possibile la conoscenza della verità assoluta, la comprensione di valori assoluti. Infatti, di fronte all'autorità del bene assoluto che tutto domina, a coloro cui questo bene porta la salute non resta che l'ubbidienza, l'ubbidienza incondizionata e grata a colui che, in possesso del bene assoluto, conosce e vuole tale bene; un'ubbi-

dienza che, senza dubbio, non può che poggiare sulla fiducia che l'autorità del legislatore sia in possesso del bene assoluto, nello stesso modo in cui, in senso inverso, si ammette che la conoscenza di questo bene resta interdotta alla grande massa dei sudditi» ([8], p. 139). E ancora un noto brano sempre di Kelsen: «Nel XVIII capitolo del Vangelo di San Giovanni viene descritto il processo di Gesù. La semplice storia nella sua ingenua esposizione è un brano sublime della letteratura e, senza volerlo, assume a tragico simbolo dell'antagonismo tra assolutismo e relativismo.

Erano i giorni della Pasqua ebraica quando Gesù, accusato di pretendere di essere il Figlio di Dio e il re dei Giudei, venne portato davanti a Pilato, procuratore romano. E Pilato ironicamente chiese a Gesù, che agli occhi del Romano non era altro che un povero pazzo: "Sei tu, così, il re dei Giudei?". Ma Gesù prese la questione molto seriamente e, tutto penetrato dalla fiamma della sua missione divina, rispose: "Tu lo dici che io sono re. Per questo io sono nato e per questo io sono venuto al mondo: per rendere testimonianza alla verità. Chiunque è dalla verità ascolta la mia voce". Allora Pilato domandò: "Che cos'è la verità?". E poiché egli, lo scettico relativista, non sapeva cosa fosse la verità, l'assoluta verità in cui questo uomo credeva, si affidò in perfetta coerenza alla procedura democratica rimettendo la decisione del caso al voto popolare. Egli andò incontro ai Giudei, riporta il Vangelo, e disse loro: "Io non trovo in lui nessuna colpa. Ma voi avete l'usanza che io vi rilasci uno in occasione della Pasqua. Volete dunque che vi liberi il re dei Giudei?". Allora tutti gridarono dicendo: "Non costui, ma Barabba". Il Vangelo aggiunge: "Barabba poi era un ladro". Per quelli che credono nel Figlio di Dio e nel re dei Giudei quale testimone dell'assoluta verità, questo plebiscito è certo un forte argomento contro la democrazia. E questo argomento noi scienziati politici dobbiamo accettarlo. Ma soltanto a una condizione: di essere così sicuri della nostra verità politica da imporla, se necessario, con il sangue e con le lacrime, di essere così sicuri della nostra verità, come lo era, della sua, il Figlio di Dio» ([8], pp. 452-453 e pp. 264-266).

4. Se l'etica non è scienza; se l'etica non è conoscenza; se l'etica non si può dire: non si può dire scientificamente; ebbene, stabilite queste cose, viene subito da chiedersi: cosa può fare la ragione nell'etica? E, senza andare troppo per le lunghe, affermiamo che la ragione nell'etica può fare molto. Così, per esempio, la ragione può analizzare e stabilire i mezzi più efficaci e meno costosi per raggiungere fini scelti e desiderati; essa può farci vedere che certi fini sono irrealizzabili all'epoca o di principio; può mostrare come la realizzazione di un valore possa farci calpestare un altro valore anch'esso accettato per buono (come potrebbe essere il caso del risanamento di un bilancio aziendale a scapito dell'occupazione); può eliminare quei disaccordi di atteggiamento che dipendono da disaccordi di credenza; può condurre all'analisi del maggior numero di alternative nella soluzione di un problema etico; può renderci più responsabili ponendoci sotto gli occhi le conseguenze delle nostre scelte - scelte che, così, si configurano come «scelte a occhi aperti»; può farci capire che l'etica dell'intenzione (*fiat justitia, pereat mundus*) non basta, in quanto essa deve pur sempre fare i conti con l'etica che bada ai risultati, con l'etica cioè della responsabilità (*fiat justitia ne pereat mundus*): l'etica dell'intenzione «non è razionalmente sufficiente», se non altro a motivo delle inevitabili conseguenze inintenzionali che possono portare a esiti non solo diversi ma addirittura contrari agli scopi intesi e voluti. In ogni caso, tuttavia, le cose più importanti che la ragione può fare nel cameo dell'etica stanno nel farci vedere che l'etica non è scienza e che i valori non trovano un fondamento razionale ultimo e certo. Non è possibile fondare razionalmente nessun sistema di norme etiche, nessun valore supremo, nessun principio etico ultimo. Difatti: una norma si fonda, all'interno di un'argomentazione, presupponendone almeno un'altra; e questa la si accetta perché ne viene presupposta ancora un'altra. E così in avanti, fino a che si giunge a quella norma (o a quell'insieme di norme) che fonda (o fondano) il sistema, ma che da parte sua (loro), non si fonda (o non si fondano) su nulla. Fondano il sistema nel senso che le altre norme derivano da esse. Ma esse non derivano logi-

camente da altro. Sono norme poste, pre-poste all'intero sistema: sono proposte etiche. E queste non si fondano né si confutano. Si accettano o si respingono. Si potrebbe dire che la legge di Hume e la base logica della libertà di coscienza. I valori supremi sono oggetto di scelte di coscienza: non sono né teoremi «dimostrati» né assiomi «autoevidenti» o «autofondantisi» [9].



- [1] D. Hume: *A Treatise of Human Nature.* , London (1739-1740). libro III, parte I, 1.
- [2] B. Pascal: *Pensieri.* Edizioni Paoline, Milano (1961). pp. 178-179.
- [3] M. Weber: *La scienza come professione, in Il lavoro intellettuale come professione.* Einaudi, Torino (1967). p. 31.
- [4] *Etica senza verità* è il significativo titolo di un libro di U. Scarpelli, edito da il Mulino, Bologna, (1982).
- [5] H. Poincaré: *Dernières Pensées.* , Paris (1917). Il saggio di Poincaré da cui è tratto il brano citato si intitola La morale e la scienza, ed è tradotto in italiano nell'antologia Poincaré, a cura di F. Severi, L'Arco, Firenze, 1949, pp. 214-215.
- [6] A. Einstein, *Scienza e religione*, in *Pensieri degli anni difficili*, Bollati Boringhieri, Torino, (1965) p. 134.
- [7] K.R. Popper: *La società aperta e i suoi nemici.* Armando, Roma (v).ol. 1, p. 100.
- [8] H. Kelsen: *La democrazia.* Il Mulino, Bologna (1966). p. 139.
- [9] Cfr. su tale problematica H. Kelsen, *Lineamenti di dottrina pura del diritto*, Einaudi, Torino, (1952); R. Hare, *Il linguaggio della morale*, Ubaldini, Roma, (1968); N. Bobbio, *Giusnaturalismo e positivismo giuridico*, Laterza, Bari, (2011), U. Scarpelli, *Che cos'è il positivismo giuridico*, Comunità, Milano, (1965), U. Scarpelli, *Filosofia analitica, norme e valori*, Comunità, Milano, (1961).



Dario Antiseri: si è laureato in Filosofia nel 1963 a Perugia discutendo una tesi sul passaggio dal primo al secondo Wittgenstein. Dal 1963 al 1967 ha studiato logica matematica e filosofia della scienza presso le università di Vienna, Münster, Oxford. Ha insegnato discipline filosofiche presso le Università di Roma "La Sapienza",

Siena, Padova, e LUISS. È attualmente membro del consiglio direttivo dell'Alta Scuola di Studi del Collegio S. Carlo di Modena. Tra le sue pubblicazioni – molte delle quali tradotte in più lingue: *Teoria Unificata del Metodo* (1981 e succ. edizioni); *Ragioni della Razionalità* (voll. I & II, 2004); *Karl Popper* (2002); *La Vienna di Popper* (2000); *Trattato di Metodologia delle Scienze Sociali* (2007); *Come si ragiona in Filosofia* (2011); *Dalla parte degli Insegnanti* (2013). È autore con Giovanni Reale di una diffusa storia della filosofia: *Il Pensiero Occidentale dalle Origini ad Oggi* (più volte edito, tradotto in portoghese, spagnolo e kazako, ed in corso di traduzione in cinese) e dell'opera in due volumi *Cento Anni di Filosofia* (2015).

Sfere e gruppi topologici

Francesco Esposito

Dipartimento di Matematica e Fisica "Ennio De Giorgi" - Università del Salento

Si è spesso interessati a dotare oggetti matematici di strutture sempre più ricche che consentano una maggiore comprensione dell'oggetto stesso. Ovviamente tale operazione non è sempre possibile. Un risultato degli anni Quaranta del secolo scorso mostra, ad esempio, come le uniche sfere ad ammettere una struttura di gruppo topologico siano S^0 , S^1 e S^3 . In questo articolo si discuterà tale proposizione dandone una parziale, ma elementare, dimostrazione.

Introduzione

Alcune definizioni e risultati preliminari utili in seguito:

Definizione 1. Sia G uno spazio topologico su cui è definita un'operazione \cdot per cui (G, \cdot) è un gruppo. G è detto un **gruppo topologico** se il prodotto

$$G \times G \rightarrow G, \quad (g, g') \mapsto g \cdot g'$$

e l'inversione

$$G \rightarrow G, \quad g \mapsto g^{-1}$$

sono applicazioni continue.

Quello che si richiede, pertanto, è che le strutture di gruppo e di spazio topologico si "raccordino" in modo coerente.

Proposizione 2. Se (G, \cdot) è un gruppo topologico, per ogni $a \in G$ la moltiplicazione a sinistra (risp. a

destra) definita da

$$L_a : G \rightarrow G, \quad g \mapsto a \cdot g \\ (\text{risp. } R_a : G \rightarrow G, \quad g \mapsto g \cdot a)$$

è un omeomorfismo. Inoltre, l'applicazione che ad ogni elemento di G associa l'inverso è un omeomorfismo.

Dimostrazione. Sia $a \in G$. Notiamo che L_a è continua poiché G è un gruppo topologico e che l'inversa è data da $L_{a^{-1}}$, anch'essa continua. Pertanto L_a è un omeomorfismo. Analogamente, si dimostra che R_a è un omeomorfismo. Inoltre, l'inversa

$$G \rightarrow G, \quad g \mapsto g^{-1}$$

per definizione di gruppo topologico, è continua e coincide con la propria inversa; pertanto si ha la tesi. \square

Definizione 3. Uno spazio topologico X si dice **omogeneo** se per ogni $x, y \in X$ esiste un omeomorfismo $f : X \rightarrow X$ tale che

$$f(x) = y.$$

La proprietà di omogeneità, quindi, indica che ogni porzione dello spazio topologico è "indistinguibile" da un'altra dal punto di vista topologico.

I gruppi topologici sono spazi omogenei: infatti, se (G, \cdot) è un gruppo topologico e $g, g' \in G$, allora

$$f = L_{g^{-1}} \circ R_{g'}$$

è un omeomorfismo (segue dalla proposizione precedente, considerando la composizione di omeomorfismi) tale che

$$f(g) = g^{-1} \cdot g \cdot g' = g'.$$

Non è tuttavia vero il viceversa: si può dimostrare, ad esempio, che la retta reale \mathbb{R} munita della topologia di Sorgenfrey (che ha come base di aperti gli intervalli semichiusi $[a, b]$) è uno spazio topologico omogeneo che non ammette struttura di gruppo topologico.

Un altro esempio di spazio topologico omogeneo ma che, come vedremo, non ammette struttura di gruppo è la sfera S^2 .

Struttura di gruppo topologico sulle sfere S^0, S^1, S^3

Ricordiamo la definizione di sfera n -dimensionale:

$$S^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} : \sum_{j=1}^{n+1} x_j^2 = 1\}$$

Ogni sfera $S^n \hookrightarrow \mathbb{R}^{n+1}$ ammette una struttura di spazio topologico, quella di sottospazio indotta da \mathbb{R}^{n+1} . Come vedremo, non è invece sempre possibile munire S^n della struttura algebrica di gruppo. Discutiamo ora i seguenti casi.

- $n = 0$.

Dalla definizione di sfera, segue che

$$S^0 = \{+1, -1\} = \mathbb{Z}/2.$$

- $n = 1$.

Notiamo che, nel caso della circonferenza nel piano, è possibile l'identificazione $\mathbb{R}^2 \cong \mathbb{C}$. In tal senso,

$$S^1 = \{z \in \mathbb{C} : |z| = 1\}.$$

Possiamo pertanto considerare su S^1 la seguente operazione

$$\cdot : S^1 \times S^1 \rightarrow S^1$$

così definito: per ogni $z_1, z_2 \in S^1$

$$z_1 \cdot z_2 = z_1 z_2.$$

Ossia consideriamo il prodotto di numeri complessi unitari, che risulta ancora essere un complesso unitario poiché

$$|z_1 z_2| = |z_1| |z_2| = 1.$$

Vale banalmente la proprietà associativa dell'operazione \cdot in quanto essa vale in \mathbb{C} . Inoltre se $z \in S^1$, allora l'inverso di z si dimostra essere

$$z^{-1} = \bar{z},$$

mentre l'elemento neutro è, chiaramente, 1; pertanto (S^1, \cdot) è un gruppo (abeliano).

- $n = 3$.

Anche nel caso di $S^3 \hookrightarrow \mathbb{R}^4$ è possibile un'utile identificazione dello spazio ambiente.

Consideriamo $\{1, i, j, k\}$ base canonica di \mathbb{R}^4 e definiamo un'operazione bilineare \cdot avente 1 come elemento neutro e tale che sulla base valga

$$\begin{aligned} ij &= k = -ji, \\ jk &= i = -kj, \\ ki &= j = -ik, \\ i^2 &= j^2 = k^2 = -1. \end{aligned}$$

Con la somma componente per componente e l'operazione di prodotto ora definita, $\mathbb{H} = (\mathbb{R}^4, +, \cdot)$ risulta essere un corpo, noto come il corpo dei **quaternioni**.

Tale struttura fu introdotta nel 1843 da Sir William Hamilton, nel tentativo di estendere al caso tridimensionale la forte connessione che vi è, nel caso bidimensionale, tra numeri complessi e geometria nel piano (connessione di cui abbiamo fatto uso nella discussione del precedente caso $n = 1$).

Se $q \in \mathbb{H}$ allora esistono unici $q_1, q_2, q_3, q_4 \in \mathbb{R}$ tali che

$$q = q_1 + q_2 i + q_3 j + q_4 k;$$

definiamo il **coniugato** \bar{q} di q come il quaternionione

$$\bar{q} = q_1 - q_2 i - q_3 j - q_4 k.$$

Svolgendo i calcoli si ottiene che se $p, q \in \mathbb{H}$ e $\alpha \in \mathbb{R}$:

$$\overline{p+q} = \bar{p} + \bar{q}, \quad \overline{pq} = \bar{q}\bar{p}, \quad \bar{\bar{q}} = q, \quad \bar{\alpha} = \alpha.$$

Inoltre, si definisce **norma** di q il numero reale non negativo

$$N(q) = q\bar{q} = q_1^2 + q_2^2 + q_3^2 + q_4^2.$$

È possibile definire l'**inverso** di un quaternionione q come il quaternionione

$$q^{-1} = \frac{1}{N(q)}\bar{q}.$$

Possiamo ora discutere il caso della sfera 3-dimensionale. Si ha che

$$S^3 = \{q \in \mathbb{H} : N(q) = 1\}.$$

Consideriamo su S^3 la restrizione del prodotto \cdot definito su \mathbb{H} : questo è possibile poiché, se $p, q \in S^3$, allora

$$N(pq) = pq\bar{p}\bar{q} = pq\bar{q}\bar{p} = p\bar{p} = 1.$$

Pertanto $pq \in S^3$.

Inoltre, se $q \in S^3$, si verifica subito che

$$N(q^{-1}) = \frac{\bar{q}}{N(q)} \frac{\overline{\bar{q}}}{N(q)} = \bar{q}q = N(q) = 1.$$

E pertanto anche l'inverso di ogni elemento di S^3 appartiene ad S^3 . Quindi (S^3, \cdot) risulta essere un gruppo. Si noti che, a differenza dei due casi precedenti, questo gruppo non è abeliano: infatti, $i, j, k \in S^3$ ma

$$ij = k \neq -k = ji.$$

- $n = 7$.

In completa analogia al caso precedente, anche su \mathbb{R}^8 può essere definita un'applicazione bilineare che dia allo spazio un'operazione interna. Se infatti $\{1, e_1, \dots, e_7\}$ è la base canonica di \mathbb{R}^8 , si può definire un'applicazione bilineare \cdot avente 1 come unità e tale che

\cdot	e_1	e_2	e_3	e_4	e_5	e_6	e_7
e_1	-1	e_4	e_7	$-e_2$	e_6	$-e_5$	$-e_3$
e_2	$-e_4$	-1	e_5	e_1	$-e_3$	e_7	$-e_6$
e_3	$-e_7$	$-e_5$	-1	e_6	e_2	$-e_4$	e_1
e_4	e_2	$-e_1$	$-e_6$	-1	e_7	e_3	$-e_5$
e_5	$-e_6$	e_3	$-e_2$	$-e_7$	-1	e_1	e_4
e_6	e_5	$-e_7$	e_4	$-e_3$	$-e_1$	-1	e_2
e_7	e_3	e_6	$-e_1$	e_5	$-e_4$	$-e_2$	-1

$\mathbb{O} = (\mathbb{R}^8, +, \cdot)$ ha una struttura di Algebra, detta **Algebra degli Ottonioni**. Le nozioni di coniugato, norma e inverso sono del tutto analoghe al caso dei quaternioni. Pertanto, anche nel caso di S^7 , si può scrivere:

$$S^7 = \{o \in \mathbb{O} : N(o) = 1\}.$$

Tuttavia, l'applicazione bilineare sopra introdotta non definisce un prodotto associativo su S^7 . Infatti, ad esempio,

$$(e_1e_2)e_3 = e_4e_3 = -e_1$$

mentre

$$e_1(e_2e_3) = e_1e_5 = e_6.$$

Questo, ovviamente, non dimostra che su S^7 non si possano definire strutture di gruppo. Tuttavia indica che, data un'algebra di dimensione n , costruendone un'altra di dimensione doppia $2n$ (secondo quella che è nota in generale come *costruzione di Cayley-Dickson*), si riesce a dotare di struttura di gruppo solo le sfere di cui abbiamo discusso.

Osservazione 4. Si noti che i prodotti definiti sulle sfere S^0, S^1, S^3 risultano essere continui. Infatti, nel caso 0-dimensionale, la topologia considerata su $S^0 \times S^0$ è quella discreta ed il prodotto è sicuramente continuo. Nel caso 1-dimensionale, si è considerata la restrizione su S^1 del prodotto tra numeri complessi, il quale risulta essere continuo. Infatti, considerando $\mathbb{C} = \mathbb{R} \times \mathbb{R}$, il prodotto tra i numeri complessi $z_1 = (x_1, y_1)$ e $z_2 = (x_2, y_2)$ è dato dal numero complesso $z_1 \cdot z_2 = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1)$. Se indichiamo con

$$p : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C} = \mathbb{R} \times \mathbb{R}$$

il prodotto tra numeri complessi, e π_1, π_2 le proiezioni (continue) su prima e seconda componente nell'identificazione $\mathbb{C} = \mathbb{R} \times \mathbb{R}$, si ottiene che

$$p(z_1, z_2) = (\pi_1(z_1)\pi_1(z_2) - \pi_2(z_1)\pi_2(z_2), \pi_1(z_1)\pi_2(z_2) + \pi_1(z_2)\pi_2(z_1)).$$

Così, poiché una funzione a valori nel prodotto di due spazi è continua se e solo se le sue componenti sono continue e poiché somme e prodotti di funzioni reali continue sono continue, si ottiene la tesi.

Rimane da vedere il caso 3–dimensionale. Con una dimostrazione del tutto analoga a quella vista per i numeri complessi, considerando

$$\mathbb{H} = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$$

e $\pi_1, \pi_2, \pi_3, \pi_4$ proiezioni (continue) su ognuna delle componenti, si ottiene che il prodotto tra quaternioni è continuo e di conseguenza lo è anche il prodotto definito su S^3 , restrizione del prodotto in \mathbb{H} .

Una dimostrazione parziale

La dimostrazione data da H. Samelson [3] nel 1940 del fatto che le uniche sfere ad ammettere struttura di gruppo topologico sono S^0, S^1 ed S^3 , non è elementare ed utilizza la teoria dei gruppi di Lie. La definizione di gruppo di Lie è simile a quella di gruppo topologico e prevede l'utilizzo di una varietà differenziabile al posto di uno spazio topologico, richiedendo inoltre che la moltiplicazione e l'inversione siano applicazioni differenziabili (al posto della continuità). In particolare, ogni gruppo di Lie è un gruppo topologico.

La dimostrazione che si propone in questo lavoro, apparsa in [2], è certamente più semplice, seppur parziale. Infatti si dimostrerà la tesi unicamente per le sfere di dimensione pari S^{2n} con $n \geq 0$.

Premettiamo alcune definizioni e risultati preliminari.

Definizione 5. Sia G un gruppo ed X uno spazio topologico. Si dice che G agisce su X se esiste un omomorfismo

$$\varphi : G \rightarrow \text{Hom}(X)$$

dove $\text{Hom}(X)$ è il gruppo degli omeomorfismi di X . Tale omomorfismo è detto **azione del gruppo G su X** e per ogni elemento g del gruppo indichiamo $\varphi(g) = \varphi_g$.

Inoltre, se per ogni $g \in G \setminus \{e\}$ risulta che φ_g non ammette punti fissi, l'azione è detta **libera**.

Si noti che ogni gruppo topologico G agisce liberamente su se stesso. Infatti se consideriamo la seguente applicazione

$$\varphi : G \rightarrow \text{Hom}(G), \quad g \mapsto L_g,$$

risulta che φ è un omomorfismo ed inoltre, se $g \in G$, per ogni $x \in G$ si ha

$$\varphi_g(x) = x \iff gx = x \iff g = e$$

e pertanto l'azione risulta essere libera.

La dimostrazione che si propone si incentra su questa osservazione e fa uso di alcuni risultati della topologia algebrica: è noto che il gruppo di omologia n –dimensionale della sfera risulta essere isomorfo al gruppo degli interi relativi,

$$H_n(S^n) \simeq \mathbb{Z}.$$

Per ogni intero n , l'omologia consente di associare ad uno spazio topologico X un gruppo $H_n(X)$, detto *gruppo di omologia n –dimensionale*. Inoltre, se $f : X \rightarrow Y$ è un'applicazione continua tra spazi topologici, per ogni intero n , risulta definito un omomorfismo di gruppi

$$f_* : H_n(X) \rightarrow H_n(Y).$$

Tali omomorfismi indotti soddisfano le seguenti proprietà

$$(Id_X)_* = Id_{H_n(X)} \quad \text{e} \quad (f \circ g)_* = f_* \circ g_*.$$

Si ha inoltre che i gruppi di omologia sono invarianti topologici.

In maniera più informale, si potrebbe dire che i gruppi di omologia “contano i buchi dello spazio topologico in una data dimensione”: se consideriamo, ad esempio, S^2 , allora $H_2(S^2) \simeq \mathbb{Z}$ (poiché la sfera è vuota all'interno). Questo ragionamento (che non rappresenta certamente una prova formale dell'asserzione fatta) è valido in generale per ogni $n \in \mathbb{N}$.

Sia ora $f : S^n \rightarrow S^n$ un'applicazione continua. Per quanto detto, f induce un endomorfismo

$$f_* : H_n(S^n) \rightarrow H_n(S^n).$$

Poiché $H_n(S^n) \simeq \mathbb{Z}$, si ha che $H_n(S^n)$ risulta essere ciclico e quindi, per ogni $\alpha \in H_n(S^n)$, esiste $d \in \mathbb{Z}$ tale che

$$f_*(\alpha) = d\alpha.$$

Tale intero d dipende unicamente da f ed è detto **grado di f**

$$\text{deg}(f) = d.$$

L'applicazione che ad una funzione continua da S^n in sé associa il grado della funzione soddisfa le seguenti proprietà

1) $\deg(I_{S^n}) = 1;$

2) se $f, g : S^n \rightarrow S^n$ sono due applicazioni continue, allora

$$\deg(f \circ g) = \deg(f) \deg(g);$$

inoltre f e g sono omotope (ossia esiste un'applicazione continua $H : [0, 1] \times S^n \rightarrow S^n$, t.c.

$$H(0, x) = f(x), H(1, x) = g(x)$$

per ogni $x \in S^n$) se e solo se

$$\deg(f) = \deg(g);$$

3) se f è la simmetria rispetto ad un iperpiano passante per il centro della sfera, allora

$$\deg(f) = -1;$$

4) l'applicazione

$$I^- : S^n \rightarrow S^n, \quad x \mapsto -x$$

(che ad ogni punto associa l'antipodale) ha grado

$$\deg(I^-) = (-1)^{n+1};$$

5) se $f : S^n \rightarrow S^n$ non ha punti fissi, allora

$$\deg(f) = (-1)^{n+1}.$$

Pertanto, se $f : S^n \rightarrow S^n$ è un'applicazione che non ha punti fissi, poiché

$$\deg(f) = (-1)^{n+1} = \deg(I^-),$$

essa risulta essere omotopa all'applicazione che associa ad ogni punto l'antipodale. Non vale tuttavia il viceversa, poiché se n è dispari

$$\deg(I^-) = 1 = \deg(I_{S^n})$$

mentre se n è pari

$$\deg(I^-) = -1 = \deg(f)$$

dove f è la simmetria rispetto ad un iperpiano passante per il centro della sfera; in ognuno dei due casi, le applicazioni hanno punti fissi (se n è dispari tutti i punti sono fissi, mentre se n è pari i punti dell'equatore individuato dall'iperpiano sono lasciati fissi).

Possiamo ora provare il seguente

Teorema 6. $\mathbb{Z}/2$ è l'unico gruppo non banale, a meno di isomorfismi, ad agire liberamente su S^n , per n pari.

Dimostrazione. Sia n pari e sia

$$\varphi : G \rightarrow \text{Hom}(S^n)$$

una qualsiasi azione libera del gruppo G su S^n . Poiché per ogni $g \in G$, $\varphi_g \in \text{Hom}(S^n)$, possiamo calcolarne il grado $\deg(\varphi_g)$. Osserviamo che, poiché φ è un omomorfismo e l'operazione in $\text{Hom}(S^n)$ è quella di composizione, si ha, sfruttando le proprietà del grado,

$$\begin{aligned} \deg(\varphi_g) \deg(\varphi_{g^{-1}}) &= \deg(\varphi_g \circ \varphi_{g^{-1}}) = \\ &= \deg(\varphi_e) = \deg(I_{S^n}) = 1. \end{aligned}$$

e quindi, poiché per definizione il grado è un intero,

$$\deg(\varphi_g) = \deg(\varphi_{g^{-1}}) = \pm 1.$$

Possiamo pertanto considerare la seguente applicazione

$$d : G \rightarrow \{\pm 1\} \simeq \mathbb{Z}/2, \quad g \mapsto d(g) = \deg(\varphi_g).$$

Tale applicazione è ben definita per la precedente osservazione ed inoltre, se $g, h \in G$

$$\begin{aligned} d(hg) &= \deg(\varphi_{hg}) = \deg(\varphi_h \circ \varphi_g) = \\ &= \deg(\varphi_h) \deg(\varphi_g) = d(h)d(g), \end{aligned}$$

e pertanto d è un omomorfismo. Poiché l'azione è libera si ha che per ogni $g \in G \setminus \{e\}$, φ_g non ammette punti fissi, mentre $\varphi_e = I_{S^n}$ banalmente lascia tutti i punti fissi: pertanto, per le proprietà del grado, $d(g) = -1$ e $d(e) = 1$ dimostrando così che d è suriettiva ed ha nucleo banale. Pertanto, d è un isomorfismo e

$$G \simeq \mathbb{Z}/2.$$

□

Corollario 7. Se $n > 0$, allora S^{2n} non è un gruppo topologico.

Dimostrazione. Abbiamo visto che ogni gruppo topologico agisce liberamente su se stesso. Se S^{2n} fosse un gruppo topologico, per il precedente Teorema,

$$S^{2n} \simeq \mathbb{Z}/2$$

e quindi $n = 0$. □

In altre parole, tra le sfere S^n di dimensione pari, solo S^0 ammette una struttura di gruppo topologico.



- [1] **Manetti, Marco**, *Topologia*, 2a edizione, Springer, (2008);
- [2] **Megía, Ignacio Santa-Maria**, *Which spheres admit a topological group structure*, Rev. R. Acad. Cienc. Exactas Fis. Quim. Nat. Zaragoza (2), 62, 75-79, (2007);
- [3] **Samelson, Hans**, *Über die Sphären, die als Gruppenräume auftreten*, Commentarii Mathematici Helvetici, 13, 144-155, (1940).



Francesco Esposito: Si è laureato in Matematica presso il Dipartimento di Matematica e Fisica "Ennio De Giorgi" dell'Università del Salento discutendo una tesi di Geometria Riemanniana. Attualmente è un dottorando del Dipartimento di Matematica e Fisica dell'Università del Salento.

Numero X Anno 2017

Viaggio nella Scienza

Ithaca

Il vuoto

