

# A CIEN AÑOS DE LA PUBLICACIÓN DEL ARTÍCULO DE EINSTEIN SOBRE LOS CUANTOS DE LUZ

Luis Hernández, Facultad de Física, Universidad de La Habana, La Habana, Cuba

## RESUMEN

Se presenta una breve reseña histórica de los antecedentes que originaron las ideas de Einstein sobre los cuantos de luz, al mismo tiempo que se resalta el significado e impacto del famoso artículo de Einstein de 1905. En el presente trabajo se evidenciará que el concepto de cuanto de luz elaborado por Einstein, fue la coherente concepción de un programa de investigación, independiente de las ideas de Planck, que habría conducido irremediamente a la ley del cuerpo negro y al concepto de los cuantos de energía. Se evidenciará como Einstein utilizó la mecánica estadística para fundamentar la termodinámica del equilibrio con base en la mecánica clásica. La descripción einsteiniana, mecánica-estadística de la interacción radiación-sustancia conllevó a la búsqueda de los componentes elementales de la radiación electromagnética, los cuantos de luz.

## ABSTRACT

A brief historical description regarding the antecedents of the Einstein original ideas about the quantum of light is shown. The present work also stands out the meaning and impact of the Einstein's famous article of 1905. It will be demonstrated that the quantum light idea, elaborated by Einstein, was the coherent conception of a research programme, independent of that developed by Planck, that would have hopelessly lead to the law of the black body and to the concept of the energy quantum. In addition, it will be shown how like Einstein used the statistical mechanical description in order to explain the electromagnetic radiation-substance interaction and to found the elementary components of the electromagnetic radiation: the quantum of light.

## I. INTRODUCCIÓN

Siempre que se enseña la teoría de Albert Einstein (1879-1955) sobre los cuantos de luz, la cual explica el efecto fotoeléctrico, los estudiantes se sienten inmediatamente atraídos, no sólo porque todo aquello que lleva el nombre de Einstein posee un encanto mítico, sino también porque es una teoría sencilla de entender para explicar un problema de gran trascendencia e irresoluto hasta ese entonces.

El artículo de 1905, fue publicado en *Annalen der Physik* (ADP) e intitulado "Sobre un punto de vista heurístico referente a la emisión y a la transformación de la luz" [1] y desde su título ya se está infiriendo cual será su propósito, pues el concepto de heurístico se utiliza como asistencia práctica en la resolución de un problema sin que esté plenamente justificado su empleo.

Hoy se conoce este artículo, como aquel en donde se explica el efecto fotoeléctrico y de seguro que continuará siendo así, pero en verdad ese no fue el objetivo central de Einstein. De hecho antes de 1905, el efecto fotoeléctrico no estaba bien establecido, era muy difícil realizar el experimento con mediciones precisas y reproducibles, en particular por la necesidad de una alta calidad de las superficies metálicas ya que si se encuentran

depositadas pequeñas cantidades de impurezas se afectan grandemente los valores de la frecuencia de corte. Sólo en 1916, el físico norteamericano Robert Millikan [2,3], con trabajos de extraordinaria calidad pudo comprobar experimentalmente la ecuación de Einstein. Confirmada experimentalmente la teoría, Einstein recibió el Premio Nobel en 1921, por el efecto fotoeléctrico y no por su famosa teoría de la relatividad. Millikan lo recibió dos años más tarde.

En el presente trabajo se evidenciará que el concepto de cuanto de luz elaborado por Einstein, fue la coherente concepción de un programa de investigación, independiente de las ideas de Planck, que habría conducido irremediamente a la ley del cuerpo negro y al concepto de los cuantos de energía. En sus estudios iniciales, Einstein utilizó la mecánica estadística para fundamentar la termodinámica del equilibrio con base en la mecánica clásica. La descripción einsteiniana, mecánica-estadística de la interacción radiación-sustancia conllevó a la búsqueda de posibles componentes elementales de la radiación electromagnética y a la existencia de procesos elementales entre los mismos. Los componentes elementales, aunque no en su forma definitiva, se insinúan en el artículo de 1905, pero se establecieron definitivamente en 1916-1917, junto con los procesos elementales de emisión y absorción [4].

## II. ANTECEDENTES

En 1900, después de 50 años de haberse establecido experimentalmente por Gustav Kirchhoff (1824-1887) y otros científicos la distribución espectral de la energía de la radiación del cuerpo negro, Max Planck (1858-1947), alumno de H.L. Helmholtz (1821-94) y del propio Kirchhoff, profesor entonces en la Universidad de Berlín, en un “acto de desesperación”, según él confesó, encontró su famosa ley para explicar la radiación Electromagnética confinada e intercambiando energía en equilibrio con los átomos de la cavidad. En lugar de deducir la fórmula de la radiación electromagnética partiendo de la teoría, Planck propuso, dentro del marco de la termodinámica, ajustar los datos experimentales a una ecuación empírica. Posteriormente, para justificar físicamente la expresión tuvo que suponer dos aspectos fundamentales, primero, admitir la estadística de Maxwell-Boltzmann y por tanto las consideraciones probabilísticas de la física. Segundo, supuso que los átomos se comportaban como osciladores eléctricos que absorbían y emitían energía radiante. Cada oscilador tenía su propia frecuencia de oscilación emitiendo y absorbiendo la radiación a esa frecuencia. Todas las frecuencias están presentes en el espectro, dado el enorme número de diferentes osciladores, y por tanto el espectro es continuo. Planck introdujo una suposición ad hoc para justificar su ecuación empírica: un resonador atómico podía absorber o emitir solamente cantidades discretas de energías que eran proporcionales a su frecuencia de oscilación, “aceptando” la naturaleza discontinua de la energía, aunque para él era más importante las magnitudes continuas de la termodinámica. Planck uno de los exponentes más fieles en su época a la física clásica se convirtió sin proponérselo en su detractor. En 1910, después del trabajo de Einstein, Planck aún escribía: “la teoría de la luz sufriría un retraso de siglos, a los tiempos en que los seguidores de Newton y Huygens luchaban entre sí por la aceptación de la teoría corpuscular u ondulatoria de la luz. Toda la obra de Maxwell quedaría invalidada si se acepta el cuanto de energía radiante”. La teoría cuántica de Planck fue una victoria de su honestidad intelectual sobre sus convicciones científicas.

En el otoño de 1900, Albert Einstein aprobó los exámenes finales y obtuvo el diploma de la facultad de pedagogía del Politécnico de Zürich (E.T.H.) que entonces preparaba profesores de física y matemáticas. Cuatro años antes había ingresado en el instituto después de un primer intento infructuoso. A pesar de las buenas notas (por el sistema de seis puntos): física teórica 5, teoría de las funciones 5.5,

astronomía 5: tesis de grado 4.5; puntuación general 4.9 y de una reputación de investigador talentoso, a Einstein no lo mantuvieron en el Politécnico. Él no asistía frecuentemente a clases y en el laboratorio desdeñaba las instrucciones para la ejecución de los experimentos ejecutándolas a su modo. Pero la más grave violación fue la oportunidad en que llamó a Weber\*, sólo “señor Weber” y no el respetado “señor profesor” [5].

Después de varias suplencias como profesor de física y matemáticas en escuelas preuniversitarias, así como de clases particulares de estas disciplinas, en el verano de 1902 Einstein ocupa una plaza fija en la Oficina de Patentes de Suiza en Berna y en esa época su vida debió ser algo paradisíaca. Allí creó la teoría del movimiento browniano, la teoría de los cuantos de luz y la teoría especial de la relatividad. Sólo época tan fértil científicamente puede compararse con la que pasó Isaac Newton (1642-1727) en Woolsthorpe (1665-1667). Newton llegó en esa época a las ideas del cálculo diferencial, a la ley de gravitación universal y a su modelo corpuscular de la luz.

Es muy significativo que estando Einstein fuera de la Academia, en una simple oficina de patentes, creara tanta revolución en el pensamiento científico. En sus años juveniles de Berna, Einstein no se encontró atado a ninguna institución científica, al tiempo que poseía una libertad de creación fabulosa. Con anterioridad y también con posterioridad a Einstein en Berna, los eminentes físicos han creado sus teorías siempre vinculados a la Academia ya sea como investigador o como profesor. Einstein a lo largo de toda su vida continuó en esencia con la tradición de Berna: elaboraba los problemas de turno sin pensar nunca en la valoración de los resultados, pero esto ya lo pudo hacer en calidad de investigador profesional, profesor en Praga, en Zürich, en Berlín, en Princeton, cuando alcanzó notoriedad por su teoría de la relatividad. En los comienzos de su camino creador, un trabajo en la oficina de patentes no vinculado estrechamente con la física, le facilitó la completa absorción de su intelecto por el contenido de los problemas [6].

Tres días después de que Einstein se iniciara en su nuevo trabajo, el 26 de junio se recibió en ADP, su artículo “Teoría cinética del equilibrio térmico y del segundo principio de la termodinámica” [7], el primero de una triada, en los que se pretendía imprimir una interpretación de la entropía desde un punto de vista microscópico. En el mismo inicio del primer artículo, Einstein establece claramente cual es el objetivo de su investigación:

---

\*Heinrich Friedrich Weber, profesor de física y eminente electrotécnico, pero en la esfera de la física teórica no enseñaba nada nuevo a Einstein.

“No obstante las grandes conquistas de la teoría cinética del calor en el campo de la teoría de los gases, la mecánica no ha logrado proveer una base suficiente para la teoría general del calor al no haber demostrado aún las leyes del equilibrio térmico y el segundo principio utilizando exclusivamente las ecuaciones mecánicas y la teoría de la probabilidad, si bien las teorías de Maxwell y de Boltzmann han estado muy cerca de alcanzar ese fin. El objetivo de este artículo es resolver ese problema. Además proporcionaremos una extensión del segundo principio que es de importancia para las aplicaciones de la termodinámica. También deduciremos la expresión matemática de la entropía desde el punto de vista mecánico”.

El 26 de enero de 1903, siete meses después del primero, se recibe en ADP su segundo artículo, intitulado “Una teoría sobre los fundamentos de la termodinámica” [8]. Otra vez Einstein dejaba claro en la introducción cuales serían los objetivos de su trabajo, al tiempo que precisaba su conexión con el artículo de 1902, en donde había establecido las leyes del equilibrio térmico y el significado de la entropía con ayuda de la teoría cinética del calor. Este artículo estaba consagrado a demostrar que sin el uso de la teoría cinética es posible alcanzar los mismos resultados pero para ello es necesario utilizar “condiciones más generales”. Estas incluían hipótesis adicionales a las de naturaleza mecánica, así como una descripción probabilística más acabada de los sistemas que contenía el cálculo de los promedios y su interpretación física a través de magnitudes establecidas, es decir las bases sobre la cual se asienta la mecánica estadística clásica del equilibrio.

Su último artículo de la triada fue recibido, nuevamente en ADP, el 29 de marzo de 1904, “Sobre una teoría molecular general del calor” [9]. Este artículo es necesario analizarlo un poco más en detalle, porque es aquí donde incorpora Einstein la radiación del cuerpo negro como sistema termodinámico en equilibrio para encontrar los *componentes elementales*.

El artículo, aparte la introducción habitual, está dividido en cinco epígrafes. El primero está referido a la deducción de la expresión de L. Boltzmann (1844-1906), para la entropía, en el segundo, al igual que en los dos artículos anteriores, demuestra nuevamente el segundo principio de la termodinámica. El análisis del significado de la constante  $\chi^*$ , Einstein lo presenta en el tercer epígrafe, el cuarto está dedicado al significado de la misma constante  $\chi$

en las fluctuaciones y el último y más importante para ver su vínculo con el artículo de 1905 [1], es la utilización de la radiación del cuerpo negro como sistema termodinámico en equilibrio para estudiar las fluctuaciones.

A partir del cuarto epígrafe, “Significado general de la constante  $\chi$ ”, es donde Einstein trata por primera vez sus ideas sobre la radiación del cuerpo negro y que lo conducirían hacia la formulación de su teoría sobre los cuantos de luz. Einstein parte de un sistema gaseoso, en equilibrio térmico a la temperatura T (K), en donde el diferencial de la distribución de estados dW en el intervalo de energía E y E + dE viene dado por la actual conocida expresión:

$$dW = Ce^{-\frac{E}{2\chi T}} \omega(E) dE \quad (1)$$

donde C es una constante,  $\omega(E)$  es la hoy denominada densidad de estados y  $\chi$  es la constante universal cuyo significado físico Einstein intentaba esclarecer, y que traducida a su presente equivalente es la mitad de la constante de Boltzmann. Es cierto que en el epígrafe anterior, basándose en la teoría cinética de los gases, Einstein encontró un valor de  $\chi = 6,5 \cdot 10^{-17}$  erg/K\*\*, contra el actual valor de  $2\chi = 1,38 \cdot 10^{-16}$  erg/K, pero él pretendía encontrar un significado de  $\chi$  más general, de manera que entonces empleó el concepto de las fluctuaciones de la energía entre sistemas en equilibrio térmico, un tema que en aquel tiempo no tenía la notabilidad de hoy en día. Por supuesto, que no se detallará, por aparecer en cualquier texto de física estadística, el cálculo mediante el cual, a partir de ecuación 1 se recaba la expresión para las fluctuaciones de la energía, y que Einstein encontró en este epígrafe, a saber:

$$\overline{E^2} - \overline{E}^2 = \overline{\epsilon^2} = 2\chi T^2 \frac{d\overline{E}}{dT} \quad (2)$$

donde

$$E = \overline{E} + \epsilon \quad (3)$$

de manera que  $\overline{E}$  es la energía media recabada mediante la distribución de la ecuación 1, y  $\epsilon$  es la fluctuación de la energía, que a propósito y tal vez como constancia de la poca relevancia de entonces, él la definió entre comillas. Después de demostrar la expresión (2), Einstein recalca que  $\overline{\epsilon^2}$  simboliza una medida numérica de la estabilidad térmica del

\*Se utilizará la misma nomenclatura de los artículos originales, pero se enumerarán las ecuaciones por requisito de las normas de la RCF.

\*\*En su artículo original Einstein no escribió las unidades de  $\chi$  ni asoció explícitamente el valor de N a la constante de Avogadro.

sistema y que ésta se encuentra establecida por la constante  $\chi$  la cual alcanzaba un significado físico concluyente. Tal vez sería conveniente señalar que tanto Boltzmann como J.W. Gibbs (1839-1903) habían estudiado las fluctuaciones de la energía, pero lo hicieron relacionadas a unas meras descripciones estadísticas y coligadas a propiedades un tanto "indeseables". Es cierto que para sistemas de un gran número de componentes elementales, como es el caso de la mayoría de los sistemas termodinámicos macroscópicos, el valor de las fluctuaciones es pequeño y muchas veces intrascendente en la mecánica estadística del equilibrio.

La genialidad de Einstein fue considerar las fluctuaciones  $\overline{\varepsilon^2}$  en un sistema donde fuera relevante su valor y posible realizar su medición, a la vez que permitiera determinar la constante  $\chi$ . El sistema no fue otro que la radiación del cuerpo negro, y la argumentación la brinda el propio Einstein en su artículo:

"Si se considera la radiación térmica en un espacio vacío de dimensiones lineales muy grandes comparadas con la longitud de onda correspondiente al máximo de energía radiada a determinada temperatura, el valor de las fluctuaciones de la energía media será muy pequeño en comparación con la energía media de la radiación en ese espacio. En contraste, si las dimensiones del espacio son del orden de la longitud de onda, la fluctuaciones de la energía serán del mismo orden de magnitud que la energía de la citada radiación".

Con estas consideraciones Einstein encuentra que:

$$\overline{\varepsilon^2} = \overline{E^2} \quad (4)$$

El valor de  $\overline{E^2}$  lo obtuvo a partir de la ley de Stefan-Boltzmann

$$\overline{E} = cVT^4 \quad (5)$$

donde V es el volumen en  $\text{cm}^3$  de la cavidad que contiene a la radiación y c es una constante, cuyo valor es  $7,06 \cdot 10^{-15} \text{ erg/k.cm}^3$ . Einstein sustituyó las ecuaciones. (4) y (5) en ecuación (2), obteniendo:

$$\sqrt[3]{V} = \frac{2 \cdot \sqrt[3]{\chi}}{T} = \frac{0,42}{T} \quad (6)$$

donde las constantes  $\chi$  y c se sustituyeron por sus valores. Si  $\lambda_m$  la longitud de onda correspondiente al

máximo de energía radiada es  $\approx \sqrt[3]{V}$ , el resultado de ecuación (6) es cercano a:

$$\lambda_m = \frac{0,293}{T} \quad (7)$$

la ley del desplazamiento de W. Wien (1864-1928), comprobada experimentalmente por aquella época\*. Este fue un resultado asombroso y que apoyaba la hipótesis de Einstein de utilizar la radiación del cuerpo negro como sistema termodinámico para estudiar las fluctuaciones. En el final del artículo, él escribía:

"Ello demuestra que, mediante la teoría molecular general del calor, no sólo se puede determinar la dependencia de  $\lambda_m$ , respecto a la temperatura, sino también su orden de magnitud, y yo creo que, dada la gran generalidad de nuestras hipótesis, la concordancia no debe ser atribuida a la casualidad".

Lo primero a destacar en este párrafo es la honestidad de Einstein. En la triada de artículos reseñados, siempre escribió en tercera persona, pero aquí escribe en primera persona para resaltar que no cree en la casualidad de este resultado y además que se responsabiliza totalmente con las conclusiones.

Por supuesto, no era muy convincente para aquella época aceptar volúmenes del orden de  $\lambda^3$  ( $\sim 1 \mu\text{m}^3$ ), cuando en la termodinámica siempre se pensaba en volúmenes macroscópicos y por lo tanto continuaron sin aceptación las fluctuaciones. Pero con estas ideas, veinte años más tarde, Einstein estaba preparado para aceptar el nuevo espacio de fase de S. Bose (1894-1974) y a desarrollar la nueva estadística cuántica.

### III. EL ARTÍCULO DE 1905 [1]

En los antecedentes se evidencia como Einstein se encaminaba en su proyecto de investigación a determinar los componentes elementales de la radiación electromagnética del cuerpo negro e independiente de las ideas de Planck. La esencia del artículo de 1905 es mostrar la conclusión de su programa, esto es, que la energía del haz luminoso está discontinuamente distribuida en el espacio. Una vez más, Einstein expone brillantemente en la introducción los objetivos que perseguía. Aunque en verdad merece una cita completa, por cuestión de espacio sólo se reproducirán dos párrafos de la introducción:

\*Si emplean los valores actuales de las constantes de Boltzmann, Stefan-Boltzman y de Wien las ecuaciones 6 y 7 tienen mayor concordancia.

“La teoría ondulatoria de la luz, que opera con funciones espaciales continuas, ha funcionado bien al nivel de fenómenos puramente ópticos y probablemente nunca será sustituida por otra teoría. Debería tenerse presente, no obstante, que las observaciones ópticas están relacionadas a promedios temporales más que a valores instantáneos. A pesar de la total confirmación experimental de la teoría cuando se aplica a la difracción, reflexión, refracción, dispersión, etc., resulta que la teoría de la luz que opera con funciones espaciales continuas puede presentar contradicciones con la experiencia cuando se aplica a fenómenos de emisión y transformación de la luz.

Me parece que las observaciones asociadas a la radiación del cuerpo negro, a la fluorescencia, a la producción de rayos catódicos por luz ultravioleta y otros fenómenos parecidos relacionados con emisión o transformación de la luz se entienden más propiamente si se supone que la energía de la luz está discontinuamente distribuida en el espacio. De acuerdo con la hipótesis que vamos a admitir en este trabajo, la energía de un rayo de luz que emerge de un foco puntual no está continuamente distribuida sobre un espacio cada vez mayor, sino que consiste en un número finito de quanta de energía que están localizados en puntos del espacio, que se mueven sin dividirse y que únicamente pueden ser producidos y absorbidos en unidades completas”.

Dos aspectos de vital trascendencia resaltan de inmediato en esta introducción. Primero, el concepto de quanta de energía de Einstein es muy diferente al de Planck, éste comprende a toda la radiación luminosa y no sólo a la absorción o emisión de los resonadores de la cavidad. Segundo, Einstein deja intacta la validez de la teoría ondulatoria de la luz para explicar fenómenos ópticos como difracción, reflexión, etc. Es casi la idea actual de la unidad onda-partícula para explicar el comportamiento de la luz. Otro aspecto importante de la introducción y que ha sido su hilo conductor, es la referencia al carácter de promedio temporal asociado a las observaciones ópticas que permitiría a Einstein introducir el cálculo de los valores medios de las magnitudes ópticas utilizando la mecánica estadística y relacionarlos, como hizo antes en los artículos precedentes, con los valores experimentales. Entonces su programa utilizando la mecánica estadística queda establecido, primero encontrar los componentes elementales de la radiación del cuerpo negro y explicar posteriormente de la manera más convincente su interpretación física.

El artículo de 1905, aparte de la introducción contiene nueve secciones, la primera, Einstein la intituló “Sobre una dificultad en la teoría de la radiación del cuerpo negro” y se refirió al tratamiento clásico de la radiación que Planck había utilizado en su artículo de 1900 [10]. En esta sección se

consideró un sistema clásico, una cavidad con paredes perfectamente reflectoras y conteniendo moléculas de gas, electrones libres, ligados que sufren colisiones, emitiendo y absorbiendo radiación electromagnética. Bajo esas consideraciones Einstein planteó que si existe el equilibrio dinámico y no se ponen limitaciones a las frecuencias de los osciladores la radiación de dicho espacio debe ser idéntica a la radiación del cuerpo negro. El equilibrio entre el gas y los osciladores, Einstein lo describió mediante el principio de equipartición de la energía y entonces el valor de la energía media de un oscilador viene dada por:

$$\bar{E} = \frac{R}{N} T \quad (8)$$

donde R es la constante universal de los gases y N el número de Avogadro (aunque Einstein nuevamente no identifica la constante por ese nombre).

La electrodinámica clásica había determinado el valor de la energía media para un oscilador de frecuencia  $\nu$  cuando se establecía el equilibrio entre los osciladores y la radiación electromagnética dentro de la cavidad. Esta expresión fue también utilizada en el artículo de Planck, a saber:

$$\bar{E}_\nu = \frac{L^3}{8\pi\nu^2} \rho_\nu \quad (9)$$

donde L es velocidad de la luz en el vacío y  $\rho_\nu d\nu$  es la energía por unidad de volumen de la radiación con frecuencia comprendida en el intervalo  $(\nu, \nu + d\nu)$ . Einstein entonces igualó las ecuaciones 8 y 9 ya que el equilibrio en la cavidad impone tal condición:

$$\frac{R}{N} T = \bar{E} = \bar{E}_\nu = \frac{L^3}{8\pi\nu^2} \rho_\nu, \quad (10)$$

$$\rho_\nu = \frac{R}{N} \frac{8\pi\nu^2 T}{L^3}$$

Einstein expresó su inconformidad con este resultado porque no sólo no coincidía con la experiencia práctica sino que también resultaba incongruente por la divergencia de energía en todo el espectro de frecuencia, como a continuación demostró:

$$\int_0^\infty \rho_\nu d\nu = \frac{R}{N} \frac{8\pi}{L^3} T \int_0^\infty \nu^2 d\nu = \infty \quad (11)$$

La aplicación del principio de equipartición de la energía fue un paso consecuente en su programa de investigación, de claro carácter mecánico-estadístico, del cual por supuesto Planck en su trabajo no hizo mención alguna y por lo tanto

no alcanzó la expresión (10) ni vislumbrar la consecuencia de la ecuación (11). De manera que en este primer apartado Einstein demuestra que si la mecánica estadística presupone la validez de la ecuación (8), la electrodinámica clásica de Maxwell establece la ecuación (9) no existe alternativa desde el punto de vista clásico que no se conduzca a la catástrofe ultravioleta. Como se verá, Einstein recurre nuevamente a la mecánica estadística para resolver la contradicción.

En el segundo apartado Einstein estudia el significado de las constantes en la expresión de Planck para  $\rho_v$ , esto es:

$$\rho_v = \frac{\alpha v^3}{\beta v e^{\frac{v}{T}} - 1} \quad (12)$$

donde,  $\alpha = 6.10 \cdot 10^{-56}$  y  $\beta = 4.866 \cdot 10^{-11}$ \*. Para el caso de longitudes de ondas largas y altas temperaturas, la ecuación (12) se transforma en:

$$\rho_v = \frac{\alpha}{\beta} v^2 T \quad (13)$$

Entonces igualando las ecuaciones (10) y (12), halla el valor de N a partir de las constantes de Planck,  $\alpha$  y  $\beta$ , coincidiendo su valor con el reportado experimentalmente. De esta forma Einstein demuestra que el principio de equipartición de la energía está implícitamente contenido en la función de Planck, es decir su línea de investigación contiene la expresión de Planck.

En el epígrafe 3, Einstein nuevamente aborda el cálculo de la entropía, esta vez el sistema que estudia es la radiación electromagnética. En un tratamiento muy formal, se relacionan los conceptos de energía, temperatura y entropía, a través de una función  $\phi(\rho, v)$ , mediante la cual es posible recabar tanto la entropía como la temperatura:

$$S = V \int_0^{\infty} \phi(\rho, v) dv \quad (14)$$

$$\frac{d\phi}{d\rho} = \frac{1}{T}$$

donde nuevamente,  $v$  es la frecuencia,  $\rho$  es la densidad espectral de la radiación y  $V$  el volumen donde se encuentra contenida.

Einstein opera en el epígrafe 4 de forma inversa a como lo haría en artículos subsiguientes. A partir de una densidad espectral de la radiación, determina la

función  $\phi(\rho, v)$  y posteriormente encuentra el valor de la variación de la entropía. Es decir, conocido el resultado macroscópico ( $\rho$ ), se calcula la entropía del sistema e inmediatamente lo esencial, obtener información sobre los *componentes elementales* de la radiación. Lo sorprendente para este autor, es que Einstein en su cálculo no utilizara la Ley de Planck sino la ley de Wien:

$$\rho = \alpha v^3 e^{-\beta \frac{v}{T}} \quad (15)$$

donde  $\alpha$  y  $\beta$  son constantes de ajuste entre los valores experimentales y teóricos. Para justificar su elección, Einstein escribe:

“De los experimentos sobre la radiación del cuerpo negro es claro que la ley postulada por Wien no es exactamente válida. Sin embargo, para grandes valores de  $v/T$ , ha sido bien confirmada experimentalmente. Basaremos nuestro análisis en esta fórmula, teniendo en cuenta que nuestros resultados sólo serán válidos dentro de ciertos límites”.

¿Por qué Einstein no empleó la Ley de Planck, si ésta no tenía las limitaciones de la Ley de Wien y además había sido verificada experimentalmente? La respuesta no puede si quiera especularse, pero lo que es bien conocido, que Planck para alcanzar la ley de la radiación del cuerpo negro, empleó una metodología termodinámica, y asumió una hipótesis cuántica ad hoc, sin investigar su coherencia con el resto de la física. A diferencia, Einstein desarrolló un programa basado en la mecánica estadística que partiendo de la ley de Wien arribó a su objetivo final, esto es, la determinación de los componentes elementales de la radiación. Einstein escribía un año después (1906), en una de sus célebres introducciones:

“Entonces, la teoría de la radiación de Planck me parecía, de alguna manera, opuesta a la de mi trabajo. Sin embargo, nuevos motivos, los cuales presentamos en la primera sección de este trabajo, me han convencido, de que las bases teóricas sobre lo que se funda la teoría de la radiación de Planck son diferentes de las de la teoría de Maxwell y de la teoría de los electrones. Y precisamente, la diferencia radica en que la teoría de Planck se sirve implícitamente de la hipótesis del cuanto de luz que nosotros hemos introducido” [11].

Regresando a los cálculos, Einstein a partir de la ley de Wien encontró por el método anteriormente descrito, la variación de la entropía de la radiación del cuerpo negro ante un cambio de volumen de la cavidad a energía constante  $E$ , a saber:

\*Nuevamente aparecen las constantes sin unidades.

$$S - S_0 = \frac{E}{\beta v} \ln \left( \frac{V}{V_0} \right) \quad (16)$$

Justo a continuación escribe:

“Esta ecuación muestra que la entropía de la radiación monocromática de densidad suficientemente baja varía con el volumen de la misma forma que lo hace la entropía de un gas ideal o de una solución diluida. En lo adelante, esta ecuación se interpretará de acuerdo al principio introducido en física por Boltzmann, según el cual la entropía de un sistema es una función de la probabilidad de su estado”.

Con esto, Einstein dejaba bien claro sus propósitos de emplear un enfoque mecánico-estadístico, de manera que el quinto epígrafe lo dedicó a resolver un problema típico que hoy aparece en muchos textos, esto es, el cálculo de la variación de la entropía de un gas ideal con  $n$  moléculas y energía constante en dependencia de su volumen. Así, Einstein encontró:

$$S - S_0 = R \frac{n}{N} \ln \left( \frac{V}{V_0} \right) \quad (17)$$

Comparando las ecuaciones 16 y 17, Einstein plantea su hipótesis sobre los cuantos:

“La radiación monocromática de baja densidad (dentro del rango de validez de la ley de radiación de Wien) se comporta termodinámicamente como si estuviera constituida por cuanto de energía mutuamente independientemente de valor  $R\beta v/N$ ”.

Es decir, las ecuaciones 15 y 16, son análogas en dependencia, una describe la variación de entropía en función del volumen de un gas ideal, el cual como se sabe está constituido por *componentes elementales*, sus moléculas, en tanto la otra describe la variación de entropía de una radiación monocromática de baja densidad y entonces por analogía la radiación debe también tener sus *componentes elementales*, los cuantos.

Como se ha resaltado, el método mecánico-estadístico fue aplicado por Einstein en su triada de trabajos de 1902-04 al estudio de los gases, en particular al gas ideal, en donde los componentes elementales eran moléculas que cumplen las leyes del movimiento de Newton, logrando explicar desde un nivel microscópico los principales conceptos de la entropía de los sistemas termodinámicos. De tal manera que fue el gas ideal el sistema apropiado para encontrar analogías y al tiempo justificar el enfoque mecánico-estadístico de un sistema tan diferente por aquel entonces como lo es la radiación del cuerpo negro.

Por supuesto la idea de Einstein sobre los cuantos no podría estar basada en la inspiración de ecuaciones análogas; seguramente él tenía el convencimiento físico de que la radiación electromagnética estaba constituida por componentes elementales y el camino fue la interpretación de la entropía desde un punto de vista microscópico. Pero la idea de los cuantos no podría ser remotamente aceptada si no explicara fenómenos experimentales existentes y en su artículo de 1905 Einstein estudió tres de ellos: la regla de Stokes, la emisión de rayos catódicos por iluminación de cuerpos sólidos (efecto fotoeléctrico) y la ionización de gases por luz ultravioleta.

#### IV. EL EFECTO FOTOELÉCTRICO

De los tres hechos experimentales que estudió Einstein para justificar su teoría de los cuantos, sólo se presentará el efecto fotoeléctrico, y esto por dos razones fundamentales, una en aras de no extender en demasía el presente artículo y la otra, por la trascendencia e importancia que alcanzó con el tiempo.

La primera vez, el efecto fotoeléctrico fue presentado por P. Lenard [12], reportando que cuando se ilumina una superficie metálica, la energía de los electrones emitidos resulta independiente de la intensidad de la radiación. Es bien conocido, la forma en que Einstein explica este fenómeno a partir de su teoría de los cuantos, pero lo que resalta de manera asombrosa es que la gran mayoría de los actuales textos prácticamente utilizan las mismas palabras que empleó Einstein: la radiación luminosa constituida de cuantos de energía  $(R/N)\beta v$ , incide sobre la superficie del metal y cada cuanto interactúa con un único electrón, cediéndole la energía para extraerlo del metal y para que alcance una energía cinética. Einstein escribió entonces, que la energía cinética de cada electrón es igual a  $(R/N)\beta v - P$ , donde  $P$  es el trabajo que realiza el electrón para atravesar la superficie. Después de relacionar la energía cinética de los electrones con el potencial  $\Pi$  aplicado entre la superficie catódica metálica y un electrodo anódico, arribó a su famosa ecuación:

$$\Pi \varepsilon = \frac{R}{N} \beta v - P \quad (18)$$

donde  $\varepsilon$  es la carga del electrón. Inmediatamente, utilizando los valores de la carga del ion, Einstein comparó su ecuación (18), con los resultados experimentales que había reportado Lenard, observando coincidencia en el “orden de magnitud”. Inmediatamente propone que en una gráfica de  $\Pi$  versus  $v$  se obtendrá una recta cuya pendiente no depende de las propiedades de la sustancia. A continuación escribía cautelosamente, en concor-

dancia con su enfoque heurístico: “Por todo esto, me parece observar que no existen contradicciones entre nuestras concepciones y las propiedades del efecto fotoeléctrico detectadas por Lenard”.

## V. EPÍLOGO

La teoría de los cuantos de Einstein se demoró en ser aceptada. Su fundamento teórico tenía poco sustento y considerar que la luz transportaba paquetes de energía significaba un gran desacato a toda teoría precursora. Un ejemplo de ello resalta cuando, apoyado por Planck y Nernst entre otros, en 1913 Einstein es promovido a miembro correspondiente de la Academia Prusiana, y en un párrafo de la presentación se puede leer:

“En suma, puede decirse que de los grandes problemas en que es tan rica la física moderna, difícilmente exista uno al que Einstein no haya hecho una contribución notable. Que alguna vez errara en el blanco en sus especulaciones, como por ejemplo en sus hipótesis sobre los cuantos de luz, no puede esgrimirse realmente demasiado en su contra, porque no es posible introducir ideas realmente nuevas, ni aun en las ciencias más exactas, sin correr a veces algún riesgo” [13].

Pero lo más significativo para este autor, fue el talante de Millikan, quien por más de diez años intentó demostrar que la ecuación 18 no era válida, pero contrario a sus pretensiones, en 1916 publicó un par de artículos con su aceptación [2,3], pero eso sí, sin someterse a su fundamento teórico. En ese año Millikan escribió:

“La ecuación fotoeléctrica de Einstein parece predecir exactamente en todos los casos los resultados observados. Pero la teoría semicorpuscular mediante la cual Einstein llegó a la ecuación parece hoy completamente insostenible”.... “A pesar del éxito aparentemente completo de la ecuación de Einstein, la teoría física, de la que estaba destinada a ser expresión simbólica, se halló tan insostenible, que el mismo Einstein, según creo, ya no la mantiene”.

Brevemente es necesario señalar que Einstein continuó ahondando en las ideas relativas a los cuantos de luz y en 1916-17 publica unos trabajos trascendentes donde deduce la ley de Planck y las leyes que gobiernan los procesos de absorción y emisión de la luz que sirvieron de fundamento teórico para la construcción del láser. De tal manera, que tanto el efecto fotoeléctrico como el efecto láser tienen sus cimientos en trabajos esencialmente teóricos, desarrollados por Einstein, la más pura ciencia básica, y hoy constituyen aplicaciones tecnológicas habituales de la vida moderna.

Es sabido que Einstein fue un detractor de la teoría cuántica, ya que no aceptaba el principio del indeterminismo y de aquí su celebre frase de que “Dios no juega a los dados”. Pero este sería otro acercamiento a Einstein y los cuantos, entonces valga sólo esta mención para concluir con sus palabras:

“Un total de cincuenta años de especulación consciente no me ha acercado a la respuesta a la pregunta: ¿qué son los cuantos de luz? Es cierto que hoy día cualquier bribón cree saber la respuesta, pero se equivoca” [15].

## VI. CONCLUSIONES

Una breve reseña histórica de los antecedentes que originaron las ideas de Einstein sobre los cuantos de luz se ha presentado en los epígrafes anteriores. A lo largo del presente trabajo se evidencia que fue la coherente concepción de un programa de investigación, independiente de las ideas de Planck, el que condujo a Einstein a la ley del cuerpo negro y al concepto de los cuantos de energía de la radiación luminosa. La descripción einsteniana, mecánica-estadística de la interacción radiación-sustancia conllevó a la búsqueda de posibles componentes elementales de la radiación electromagnética que Einstein los identificó como cuanto de energía localizados en el espacio.

## AGRADECIMIENTOS

El autor agradece la revisión crítica realizada por la profesora M. T. Pérez Maldonado.

## REFERENCIAS

- [1] EINSTEIN, A. (1905): “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt”, **Ann. Phys.**, 17, 132.
- [2] MILLIKAN, R.A. (1916): “Einstein’s Photoelectric Equation and Contact Electromotive Force”, **The Physical Review**, 7, 18.
- [3] \_\_\_\_\_ (1916): “A Direct Photoelectric Determination of Planck’s Constant “h”, **Physical Review**, 7, 355.

- [4] NAVARRO VEGUILLAS, L. (1990): **Einstein profeta y hereje**, Tusquets Editores, Barcelona, 95.
- [5] KUZNETSOV, B. (1990): **Einstein Vida, Muerte, Inmortalidad**, Editorial Ciencias Sociales, La Habana, 35-42.
- [6] Ibid., pág 43-44.
- [7] EINSTEIN, A. (1902): "Kinetische Theorie des Wärmegleichgewichtes und des zweiten Hauptsatzes der Thermodynamik", **Ann Phys.**, 9, 417.
- [8] \_\_\_\_\_ (1903): "Eine Theorie der Grundlagen der Thermodynamik", **Ann Phys.**, 11, 170.
- [9] \_\_\_\_\_ (1904): "Zur allgemeinen molekularen Theorie der Wärme". **Ann Phys.**, 14, 351.
- [10] PLANCK, M. (1900): **Ann Phys.** 1, 99.
- [11] EINSTEIN, A. (1906): "Zur Theorie der Lichterzeugung und Lichtabsorption", **Ann Phys.**, 20, 199.
- [12] LENARD, P. (1902): **Ann Phys.**, 8, 169.
- [13] Citado por Navarro Veguillas, L. Ibid, véase pág. 111 y ref. pág, 25, 153.
- [14] Citado por Navarro Veguillas, L. Ibid, véase pág. 112.
- [15] Citado por Navarro Veguillas, L. Ibid, véase pág. 151 y ref.1 pág. 152.

# EPR, Bell, GHZ, AND HARDY THEOREMS, AND QUANTUM MECHANICS<sup>a</sup>

M. Socolovsky, Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, Ciudad Universitaria, México D.F., México

## ABSTRACT

We review the theorems of Einstein-Podolsky-Rosen (EPR), Bell, Greenberger-Horne-Zeilinger (GHZ), and Hardy, and present arguments supporting the idea that quantum mechanics is a complete, causal, non local, and non separable theory.

## RESUMEN

Revisamos el teorema de Einstein-Podolsky-Rosen (EPR), Bell, Greenberger-Horne-Zeilinger (GHZ), y Hardy, y se presentan argumentos soportando la idea de que la teoría mecánica es completa, causal, no local, y una teoría no separable.

## I. INTRODUCTION

In 1935, Einstein, Podolsky and Rosen (EPR) published the today 70 years old paper "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete<sup>1</sup>", in which they gave a negative answer to this question. The incompleteness refers to an individual quantum system, not to an ensemble of identical systems (section 2). To arrive at this conclusion, EPR assumed locality and a criterion for the result of a measurement of a physical quantity to be considered an element of physical reality prior to the measurement, and established a necessary condition for a physical theory to be considered complete. This led to the search of a "complete theory" by adding "hidden" variables to the wave function in order to implement realism, the most celebrated of this kind of theories being the De Broglie-Bohm theory.<sup>27</sup>

It was not until 29 years later, that Bell published his famous paper "On the Einstein-Podolsky-Rosen paradox"<sup>13</sup> in which, taking as starting points the EPR hypothesis plus the assumption of the existence of hidden variables, he derived an inequality between two-particle correlation functions (averages of products of spin projections along arbitrary directions in the case of atoms, nucleons and electrons, or polarizations in the case of photons) which for some range of angles or polarizations is violated by the quantum formulae for the correlations (section 3). For a particular value of the angle between the spin directions or of photon polarizations, as in the EPR case, the correlation is perfect, that is, the result of the measurement on one particle allows the prediction with certainty -i.e. with probability

equal to 1- of the result of the measurement on the other particle, coincides with the EPR result and does not contradict the quantum prediction; for the other values of angles or polarizations the correlation is imperfect i.e. only probabilities -less than 1- of the outcomes of the measurements on the 2<sup>nd</sup> particle are predicted from the result of the measurement on the 1<sup>st</sup> particle. The Bell's result offered the possibility to decide experimentally -and theoretically if one believes a priori in the quantum formulae-between the orthodox view<sub>5</sub> of quantum mechanics and the interpretation based on the EPR hypothesis plus, possibly, hidden variables. The experimental evidence in the last decades<sup>15-18</sup> favours the first alternative.

Within the framework of quantum mechanics one can give arguments to support the statement that the theory *is non separable* (subsection 2.1.a) and *relativistically causal* but *non local* in the sense of EPR-Bell (subsection 2.2.g). Then an analysis of the Greenberger-Horne-Zeilinger theorem ("Bell's theorem without inequalities") of 1989 shows a contradiction between quantum mechanics and local realism, but for perfect correlations of three or more particles. One concludes that quantum mechanics is a complete theory in the sense that it provides the most complete description of a quantum system, without appealing to the concept of elements of reality and/or to hidden variables (section 4).

Hardy's theorem (1992) (section 5) is a Bell's theorem for perfect correlations between two particles i.e. establishes a contradiction between quantum mechanics and local realism for this system, and shows that EPR elements of reality corresponding to Lorentz invariant observables are

<sup>a</sup> Based on a talk given by the author at the Simposio de la Sociedad Cubana de Física, La Habana, Cuba, May 19<sup>th</sup>, 2005.

**E-mail:** socolovs@nuclecu.unam.mx

not Lorentz invariant. However, since quantum mechanics is causal, no signal can be sent backwards in time and therefore there is no need of a special reference frame to block causal paradoxes.

## 2. EPR theorem (1935)

The EPR theorem<sup>1</sup> says that *quantum mechanics does not provide a complete description of an individual quantum system*. This result, commonly known as the EPR paradox since Schrödinger<sup>2</sup> considered as such the situation already pointed out in the original paper by EPR, namely that to a distant physical system one could associate two (or more) different quantum states or wave functions, was derived assuming the validity of the following hypothesis:

- i) Quantum mechanics (QM) provides a correct description of nature.
- ii) Locality: "..., since at the time of measurement the two systems no longer interact, no real change can take place in the second system in consequence of anything that may be done to the first system"<sup>1</sup> or "But on one supposition we should, in my opinion, absolutely hold fast: the real factual situation of the system  $S_2$  is independent of what is done with the system  $S_1$ , which is spatially separated from the former."<sup>3</sup>
- iii) Physical reality (sufficient condition): "If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity"<sup>1</sup>. This concept, from the physical point of view, is rather obscure, since "physical reality" was not defined before; it rests rather on intuition.
- iv) Completeness of a theory: any element of the physical reality has to have a counterpart in the theory.

In the literature, the set of hypothesis ii) and iii) is called *local realism*.

### Proof of the theorem

We shall consider here the finite dimensional Hilbert space approach of Bohm<sup>4</sup>, who replaced the continuum spectrum variables  $p$  and  $q$  by the spin  $\frac{1}{2}$  components of a pair of particles. A particle of total spin  $S = 0$  decays into two spin  $\frac{1}{2}$  particles of the same mass, which move apart in opposite directions; according to quantum mechanics, the spin wave function of the pair of particles is given by

$$\psi = \frac{1}{\sqrt{2}}(\uparrow_1 \otimes \downarrow_2 - \downarrow_1 \otimes \uparrow_2) \quad (2.1)$$

where  $\uparrow_k$  ( $\downarrow_k$ ) is the wave function of particle  $k$  ( $l$ ) with spin projection  $+\frac{1}{2}$  ( $-\frac{1}{2}$ ) in the space direction specified by the unit vector  $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ . A more precise notation is

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\hat{n}\rangle_1 \otimes |-\hat{n}\rangle_2 - |-\hat{n}\rangle_1 \otimes |\hat{n}\rangle_2) \quad (2.1')$$

where  $|\hat{n}\rangle_k$  are kets in the Hilbert space of each particle. In terms of the polar angles  $\theta$  and  $\phi$ ,

$$\begin{aligned} |\hat{n}\rangle &= \cos\frac{\theta}{2}|\uparrow\rangle + e^{i\phi}\sin\frac{\theta}{2}|\downarrow\rangle, \\ |-\hat{n}\rangle &= \sin\frac{\theta}{2}|\uparrow\rangle - e^{i\phi}\cos\frac{\theta}{2}|\downarrow\rangle, \end{aligned} \quad (2.1'')$$

where  $|\uparrow\rangle$  and  $|\downarrow\rangle$  are spin states in directions  $\hat{z}$  and  $-\hat{z}$  respectively. Notice however that this does not mean that each particle is in a pure state (see below 2.2.f.). A state like (2.1) which is not a product of states but a sum of products of states, is called an *entangled* state. According to Schrödinger<sup>2</sup>, entanglement is "the characteristic trait of quantum mechanics". *If one measures* the spin component of particle 1 with a Stern-Gerlach (SG) apparatus in the direction  $\hat{m}$  and finds  $+\frac{1}{2}$  ( $-\frac{1}{2}$ ), then one can predict with certainty that the spin component in the same direction of particle 2 is  $-\frac{1}{2}$  ( $+\frac{1}{2}$ ). Since by assumption the two particles do not interact, according to the hypothesis ii) and iii) the spin component of particle 2 is an element of physical reality. It is important to realize that this way of thinking implies that  $\psi$  has suffered one of the following two *collapses*:

$$\psi \rightarrow \uparrow_1 \otimes \downarrow_2 \text{ or } \psi \rightarrow -\downarrow_1 \otimes \uparrow_2 \quad (2.2)$$

But the choice of direction  $\hat{m}$  in the measurement of particle 1 was arbitrary, and one *could* have chosen for example the directions  $\hat{z}$ ,  $\hat{y}$  or  $\hat{x}$ , respectively giving physical reality to the spin components  $S_z$ ,  $S_y$  or  $S_x$  of particle 2. But according to QM, since the spin operators  $\hat{S}_k$ ,  $k = x, y, z$  do not commute and satisfy

$$[\hat{S}_k, \hat{S}_l] = i\hat{S}_m \quad (2.3)$$

with  $k, l, m$  in cyclic order, only one component can have a precise value. Then QM does not provide a complete description of the physical reality of any of the two particles. QED

### 2.1. Some definitions

2.1.a. A quantum system consisting of  $N$  non interacting parts is called *separable* if its wave

function  $\psi$  is the product of the wave functions  $\psi_i$  of its constituent parts  $i = 1, \dots, N$ :

$$\Psi = \Psi_1 \dots \Psi_N,$$

independently whether these parts have interacted or not among themselves before.

The existence of entangled states like (2.1) shows that QM *is a non separable theory*.

It is clear that a separable theory is local in the sense of ii) above, but not the other way around, that is, separability is a stronger property than locality. Then non separability does not imply non locality: a non separable theory can be local or non local.

2.1.b. By *relativistic causality* or simply *causality*, we understand that no signal (or information) can travel faster than light in the vacuum. This is equivalent to the statement that a particle can be acted only by other particles or fields lying in its past light cone. A theory can be non local (and then non separable) but however causal. This is the case of QM (see 2.2.g.).

## 2.2. Observations and criticisms

2.2.a. The EPR theorem *does not contradict* neither QM (only claims its incompleteness) nor mathematics.

2.2.b. The EPR *correlation* between particles is *perfect*, that is, the conditional probability of the result of the measurement on the second particle, given the probability ( $1/2$ ) that the result of the measurement on the first particle is  $+1/2$  or  $-1/2$ , is 1 i. e. *certainty*: this is a result of QM, and in this language, related to the concept of collapse, one can write (with obvious notation)

$$p(\uparrow_1, \downarrow_2) = p(\downarrow_2 \mid \uparrow_1) \times p(\uparrow_1) = 1 \times \frac{1}{2} = \frac{1}{2} \quad (2.4)$$

where the  $p$ 's denote probabilities, and  $p(A \mid B)$  denotes the conditional probability of A given B.

2.2.c. A direct calculation in QM shows that, without passing through the intermediate steps of first computing a probability and then a conditional probability, in the state (2.1) the *joint probability* for, say  $|\hat{m} >_1 \otimes |-\hat{m} >_2$  is given by

$$p(|\hat{m} >_1, |-\hat{m} >_2) = \frac{1}{2} \quad (2.5)$$

In fact, the probability amplitude for finding, in the singlet state (2.1), particle 1 (2) with spin projection in direction  $\hat{a}(\hat{b})$  is

$$A(|\hat{a} >_1, |-\hat{b} >_2) = \frac{1}{\sqrt{2}} \langle \hat{a} | \otimes \langle -\hat{b} |$$

$$\langle \hat{a} | \otimes \langle -\hat{b} | (|\hat{n} >_1 \otimes |-\hat{n} >_2 - |-\hat{n} >_1 \otimes |\hat{n} >_2)$$

$$= \frac{1}{\sqrt{2}} (1 \langle \hat{a} | \hat{n} >_1 \langle -\hat{b} | -\hat{n} >_2 - \langle \hat{a} | -\hat{n} >_1 \langle -\hat{b} | \hat{n} >_2)$$

$$= \frac{1}{\sqrt{2}} ((\cos \frac{\theta_a}{2} \langle \uparrow | + e^{-i\varphi_a} \sin \frac{\theta_a}{2} \langle -\downarrow |)$$

$$\langle \downarrow |) (\cos \frac{\theta_n}{2} |\uparrow \rangle + e^{i\varphi_n} \sin \frac{\theta_n}{2} |\downarrow \rangle)$$

$$\times (\cos \frac{\theta_b}{2} \langle \uparrow | + e^{-i\varphi_b} \sin \frac{\theta_b}{2} \langle -\downarrow |) (\sin \frac{\theta_n}{2} |\uparrow \rangle - e^{i\varphi_n} \sin \frac{\theta_n}{2} |\downarrow \rangle)$$

$$- (\cos \frac{\theta_a}{2} \langle \uparrow | + e^{-i\varphi_a} \sin \frac{\theta_a}{2} \langle -\downarrow |) (\sin \frac{\theta_n}{2} |\uparrow \rangle + e^{i\varphi_n} \cos \frac{\theta_n}{2} |\downarrow \rangle)$$

$$\times (\cos \frac{\theta_b}{2} \langle \uparrow | + e^{-i\varphi_b} \sin \frac{\theta_b}{2} \langle -\downarrow |) (\cos \frac{\theta_n}{2} |\uparrow \rangle + e^{i\varphi_n} \sin \frac{\theta_n}{2} |\downarrow \rangle)$$

$$= \frac{1}{\sqrt{2}} ((\cos \frac{\theta_a}{2} \sin \frac{\theta_n}{2} + e^{-i(\varphi_a - \varphi_n)} \sin \frac{\theta_a}{2} \sin \frac{\theta_n}{2})$$

$$(\cos \frac{\theta_b}{2} \sin \frac{\theta_n}{2} - e^{-i(\varphi_b - \varphi_n)} \sin \frac{\theta_b}{2} \cos \frac{\theta_n}{2})$$

$$- (\cos \frac{\theta_a}{2} \sin \frac{\theta_n}{2} - e^{-i(\varphi_a - \varphi_n)} \sin \frac{\theta_a}{2} \cos \frac{\theta_n}{2}) (\cos \frac{\theta_b}{2} \cos \frac{\theta_n}{2} +$$

$$+ e^{-i(\varphi_b - \varphi_n)} \sin \frac{\theta_b}{2} \sin \frac{\theta_n}{2}))$$

$$= \frac{1}{\sqrt{2}} (e^{-i(\varphi_a - \varphi_n)} \sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} - e^{-i(\varphi_b - \varphi_n)} \sin \frac{\theta_b}{2} \cos \frac{\theta_a}{2}). \quad (2.6)$$

which gives the probability

$$p(|\hat{a} >_1, |\hat{b} >_2) = |A(|\hat{a} >_1, |\hat{b} >_2)|^2 = \frac{1}{2} (\sin^2 \frac{\theta_a}{2} \cos^2 \frac{\theta_b}{2} +$$

$$+ \sin^2 \frac{\theta_b}{2} \cos^2 \frac{\theta_a}{2} - \frac{1}{2} \sin \theta_a \sin \theta_b \cos(\varphi_a - \varphi_b)). \quad (2.7)$$

For the particular case  $\hat{a} = \hat{m}$  and  $\hat{b} = -\hat{m}$ , (2.7) leads to (2.5). For the other cases  $p(|-\hat{m} >_1, |\hat{m} >_2) = 1/2$ ,  $p(|\hat{m} >_1, |\hat{m} >_2) = p(|-\hat{m} >_1, |-\hat{m} >_2) = 0$ . This method does *not* require the concept of collapse and is at the basis of the *orthodox position* with respect to QM: (2.5) or, in general, (2.6) and (2.7), is the unique information that QM provides on the physical system and so the conclusion of incompleteness of QM by EPR and the idea of an action at a distance involving superluminal transmission of signals are false.

2.2.d. Moreover, one of the most powerful criticisms to the arguments of EPR is the meaning of the words "can predict" in the hypothesis iii). As employed by EPR, these words are used in the *weak sense*, which means that it is enough with the possibility of making the measurements, even if they are not performed, that the realities of the physical quantities in the remote system occur (*counterfactual alternative*). The *strong sense* however, requires that the measurements on the first system have to be actually performed and this is not possible if the observables in question have no common eigenvalue, as for example  $S_x^{(1)}$  and  $S_z^{(1)}$ . In this case,  $S_x^{(2)}$  and  $S_z^{(2)}$  can not be elements of the physical reality and the conclusion of incompleteness is invalidated. This is essentially the criticism of Bohr<sup>5</sup>; see also the references 6 and 7. *This argument allows the coexistence of QM with relativistic causality, but destroys the EPR sufficient criterion of reality.*

2.2.e. We shall see that the EPR hypothesis ii) and iii), for statistical correlations (Bell's theorem) between two particles (section 3), and for perfect correlations between three (or more) particles (section 4) and for two particles (section 5) contradict QM.

2.2.f. The most general description of the state of a quantum mechanical system is through the *density or statistical operator*  $\hat{\rho}$ . If the system is in a *pure state*, then its description is given by a normalized ket  $|\psi\rangle \in \mathcal{H}$  ( $\mathcal{H}$  is the Hilbert space), and  $\hat{\rho}$  is given by

$$\hat{\rho} = |\psi\rangle \otimes \langle\psi| \in \mathcal{H} \otimes \mathcal{H}^*, \quad (2.8)$$

$\mathcal{H}^*$  being the dual space of  $\mathcal{H}$ . Clearly,

$$\hat{\rho}^2 = \hat{\rho}. \quad (2.8.a)$$

Instead, a *mixed state* is described by the operator

$$\hat{\rho} = \sum_i \omega_i |\psi_i\rangle \otimes \langle\psi_i| \quad (2.9)$$

with  $\omega_i \in (0,1)$  and  $\sum_i \omega_i = 1$  (the different normalized kets  $|\psi_i\rangle$  need not be orthogonal). If  $\{|a\rangle\}$  is a complete set, then

$$\begin{aligned} \text{tr}(\hat{\rho}) &= \sum_a \langle a | \hat{\rho} | a \rangle = \sum_a \sum_i \omega_i \langle a | \psi_i \rangle \langle \psi_i | a \rangle = \\ &= \sum_i \omega_i \sum_a \langle \psi_i | a \rangle \langle a | \psi_i \rangle = \sum_i \omega_i \|\psi_i\|^2 = \sum_i \omega_i = 1. \end{aligned} \quad (2.10)$$

Then

$$\text{tr}(\hat{\rho}^2) = 1 \text{ for a pure state.} \quad (2.11)$$

Clearly,  $\hat{\rho}^\dagger = \hat{\rho}$  i.e.  $\hat{\rho}$  is *hermitian*, and  $\langle \psi | \hat{\rho} | \psi \rangle \geq 0$  for all  $|\psi\rangle \in \mathcal{H}$  i.e.  $\hat{\rho}$  is a *non negative* operator with eigenvalues satisfying  $1 \geq \rho_k \geq 0$  and  $\sum_k \rho_k = 1$ . Then

$$\text{tr}(\hat{\rho}^2) < 1 \text{ if and only if } \hat{\rho} \text{ represents a mixed state.} \quad (2.12)$$

Both for a pure or mixed state, the average value of an observable A is given by

$$\langle \hat{A} \rangle = \text{tr}(\hat{\rho} \hat{A}) \quad (2.13)$$

where  $\hat{A}$  is the hermitian operator corresponding to A. In fact, if  $\{|a\rangle\}$  are its eigenkets, then

$$\begin{aligned} \langle \hat{A} \rangle &= \sum_i \omega_i \langle \hat{A} \rangle_i = \sum_i \omega_i \langle \psi_i | \hat{A} | \psi_i \rangle = \\ &= \sum_i \sum_a \sum_{a'} \omega_i \langle \psi_i | a' \rangle \langle a | \psi_i \rangle \langle a' | \hat{A} | a \rangle = \\ &= \sum_i \sum_a \sum_{a'} \omega_i \langle \psi_i | a' \rangle \langle a | \psi_i \rangle a \langle a' | a \rangle = \\ &= \sum_i \sum_a \sum_{a'} \omega_i \langle \psi_i | a' \rangle \langle a | \psi_i \rangle a \delta_{aa'} = \\ &= \sum_i \sum_a \omega_i \langle \psi_i | a \rangle \langle a | \psi_i \rangle a = \\ &= \sum_a \langle a | (\sum_i \omega_i |\psi_i\rangle \otimes \langle\psi_i|) \hat{A} | a \rangle = \sum_a \langle a | \hat{\rho} \hat{A} | a \rangle. \end{aligned}$$

Also,

$$\begin{aligned} \langle a | \hat{\rho} | a \rangle &= \langle a | \sum_i \omega_i |\psi_i\rangle \otimes \langle\psi_i| | a \rangle = \\ &= \sum_i \omega_i \langle a | \psi_i \rangle \langle \psi_i | a \rangle = p(a). \end{aligned} \quad (2.14)$$

Finally, it can be easily shown that since each  $|\psi_i\rangle$  satisfies the Schrödinger equation, then  $\hat{\rho}$  satisfies

$$\frac{d\hat{\rho}}{dt} = i[\hat{\rho}, \hat{H}] \quad (2.15)$$

where  $\hat{H}$  is the hamiltonian. So  $\hat{\rho}$  evolves with time with an opposite sign as the evolution of the Heisenberg picture of a time independent (in the Schrödinger picture) observable:  $\frac{d\hat{A}_H}{dt} = -i[\hat{A}_H, \hat{H}]$ .

Consider now a quantum system consisting in two parts, 1 and 2. Let  $\theta_1$  be an observable for part 1; then its average value in the state of the total system described by the density operator  $\hat{\rho}_{12}$  is given by

$$\begin{aligned} \langle \hat{\theta}_1 \otimes I_2 \rangle_{12} &= \text{tr}_{12}(\hat{\theta}_1 \otimes I_2 \hat{\rho}_{12}) = \text{tr}_1 \text{tr}_2(\hat{\theta}_1 \otimes I_2 \hat{\rho}_{12}) = \\ &= \text{tr}_1(\hat{\theta}_1 \text{tr}_2(I_2 \hat{\rho}_{12})) = \text{tr}_1(\hat{\theta}_1 \hat{\rho}_1^{\text{red}}) \end{aligned} \quad (2.16)$$

where

$$\hat{\rho}_1^{\text{red}} = \text{tr}_2(\hat{\rho}_{12}). \quad (2.16a)$$

is *the reduced density operator* for the subsystem 1; similarly

$$\langle I_1 \otimes \hat{\theta}_2 \rangle = \text{tr}_2(\hat{\theta}_2 \hat{\rho}_1^{\text{red}}) \quad (2.17)$$

with

$$\hat{\rho}_2^{\text{red}} = \text{tr}_1(\hat{\rho}_{12}). \quad (2.17a)$$

( $I_k$ ,  $k = 1, 2$  are the identity operators on the corresponding Hilbert spaces.) In particular, for the pure state (2.1),

$$\begin{aligned} \hat{\rho}_2^{\text{red}} &= \langle \uparrow_2 | \psi \rangle \langle \psi | \uparrow_2 \rangle + \langle \downarrow_2 | \psi \rangle \langle \psi | \downarrow_2 \rangle \\ &= \frac{1}{2} \langle \uparrow_2 | (|\uparrow_1\rangle \otimes |\downarrow_2\rangle - |\downarrow_1\rangle \otimes |\uparrow_2\rangle) \langle \uparrow_1 | \otimes \langle \downarrow_2 | - \langle \downarrow_1 | \otimes \langle \uparrow_2 | \rangle \langle \uparrow_2 | \rangle \\ &\quad + \frac{1}{2} \langle \downarrow_2 | (|\uparrow_1\rangle \otimes |\downarrow_2\rangle - |\downarrow_1\rangle \otimes |\uparrow_2\rangle) \langle \uparrow_1 | \otimes \langle \downarrow_2 | - \langle \downarrow_1 | \otimes \langle \uparrow_2 | \rangle \langle \downarrow_2 | \rangle \\ &= \frac{1}{2} (|\uparrow_1\rangle \langle \uparrow_1 | + |\downarrow_1\rangle \langle \downarrow_1 |), \end{aligned} \quad (2.18a)$$

and similarly,

$$\hat{\rho}_e^{\text{red}} = \frac{1}{2} (|\uparrow_2\rangle \langle \uparrow_2 | + |\downarrow_2\rangle \langle \downarrow_2 |). \quad (2.18b)$$

Therefore, in a matrix representation,

$$\rho_1^{\text{red}} = \rho_2^{\text{red}} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2} I. \quad (2.19)$$

So,

$$(\rho_k^{\text{red}})^2 = \frac{1}{4} I \neq \frac{1}{2} I = \rho_k^{\text{red}}, \quad k = 1, 2$$

and

$$\text{tr}((\rho_k^{\text{red}})^2) = \frac{1}{2} < 1 \quad (2.20)$$

i.e. each of the entangled spins in the pure state  $\psi$  is not in a pure state, but in a mixed state: the whole system is described by a wave function, but not each of its parts.

An immediate consequence of this fact is that the average value of any spin component of each particle vanishes:

$$\langle \sigma_k \rangle_l = \text{tr}(\sigma_k \rho_l^{\text{red}}) = 0, \quad k = 1, 2, 3, \quad l = 1, 2. \quad (2.21)$$

( $\sigma_k$  are the Pauli matrices  $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$

and  $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ , which means that no spin component of any of the two particles has a definite or precise value (and therefore a reality) that can be predicted with certainty without perturbing the corresponding particle (subsystem)<sup>8</sup>.

2.2.g. Impossibility of an instantaneous (and therefore superluminal) sending of information. Non locality.

In the situation described by the state (2.1), assume that observer 1 finds  $\uparrow$ , then observer 2 finds  $\downarrow$ . If this would imply a transmission of information from 1 to 2 and not only a correlation between 1 and 2, observer 2 should be able to distinguish between the following alternatives:

- i) 1 found 1,
- ii) 1 did not measure,
- iii) 1 measured in an arbitrary direction  $\hat{n}$ .

It is clear that 2 can not determine what actually occurred at 1, but only if the two observers come together and compare their results. However, though 1 can not send an information to 2 and therefore causality (2.1.b.) is not violated, as a consequence of the individual measurement at 1, the probability distribution at 2 has changed instantaneously since, from (2.14),

$$p(\uparrow_1) = \langle \uparrow_1 | \hat{\rho}_1^{\text{red}} | \uparrow_1 \rangle = \langle \uparrow_1 | \frac{1}{2} I | \uparrow_1 \rangle = \frac{1}{2} \langle \uparrow_1 | + \uparrow_1 | \rangle = \frac{1}{2}.$$

and from (2.5),

$$p(\downarrow_2 | \uparrow_1) = \frac{p(\uparrow_1 | \downarrow_2)}{p(\uparrow_1)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1 \quad (\text{certainty}). \quad (2.22)$$

If 1 would not have measured, then

$$p(\downarrow_2) = \langle \downarrow_2 | \hat{\rho}_2^{\text{red}} | \downarrow_2 \rangle = \langle \downarrow_2 | \frac{1}{2} I | \downarrow_2 \rangle = \frac{1}{2} \langle \downarrow_2 | \downarrow_2 \rangle = \frac{1}{2}.$$

This amounts to an *instantaneous change of state* of 2 (*collapse*), from a mixed state to a pure state i.e. to an instantaneous (and therefore superluminal) change of the density operator:

$$\hat{\rho}_2^{\text{red}} \rightarrow |\downarrow_2\rangle$$

and therefore to a violation of locality a' la EPR. Since this is a pure quantum mechanical result, we conclude that QM is a *non local theory*.

In the following we shall see that, in *the average*, the above instantaneous change of probability distribution does not occur.

2.2.h. Impossibility, in the average, of an instantaneous (and therefore superluminal) modification of a remote probability distribution<sup>9,10</sup>

Let U and V denote the two non interacting and spatially separated parts of a composite quantum system "U + V" (we can assume that in the past the two parts interacted with each other), and A and B corresponding measuring apparatuses; the total Hilbert is given by

$$\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_U \otimes \mathcal{H}_V \otimes \mathcal{H}_B \quad (2.23)$$

The interaction hamiltonians apparatuses-subsystems, respectively H(U,A) and H(V,B), lead to unitary evolution operators U(U,A) = U<sub>UA</sub> and U(V,B) = U<sub>VB</sub> which by causality commute with each other:

$$[U_{UA}, U_{VB}] = 0. \quad (2.24)$$

U<sub>UA</sub>(U<sub>VB</sub>) is the identity on  $\mathcal{H}_{VB} = \mathcal{H}_V \otimes \mathcal{H}_B$  ( $\mathcal{H}_{UA} = \mathcal{H}_U \otimes \mathcal{H}_A$ ). Before any measurement, the density operator of the total system is

$$\hat{\rho}^0 = \hat{\rho}_A^0 \otimes \hat{\rho}_{UV}^0 \otimes \hat{\rho}_B^0, \quad (2.25)$$

and in general  $\hat{\rho}_{UV}^0 \neq \hat{\rho}_U^0 \otimes \hat{\rho}_V^0$ . If A and B perform successive measurements, the density operator evolves to

$$\begin{aligned} \hat{\rho}_{BA} &= (U_{VB} U_{UA}) \hat{\rho}^0 (U_{VB} U_{UA})^\dagger = \\ &= U_{VB} U_{UA} \hat{\rho}_A^0 \otimes \hat{\rho}_{UV}^0 \otimes \hat{\rho}_B^0 U_{UA}^\dagger U_{VB}^\dagger. \end{aligned} \quad (2.26)$$

On the other hand, if the apparatus A is switched off, i. e. A makes no measurement, then U<sub>UA</sub> is the identity (since H(U, A) = 0): U<sub>UA</sub> = I<sub>U</sub> ⊗ I<sub>A</sub>.

Let  $\theta_V$  be an observable of the subsystem V; for its expectation value one has

$$\begin{aligned} \langle \hat{\theta}_V \rangle_{\hat{\rho}_{BA}} &= \text{tr}(\hat{\theta}_V \hat{\rho}_{BA}) = \text{tr}(\hat{\theta}_V U_{VB} U_{UA} \hat{\rho}_A^0 \otimes \hat{\rho}_{UV}^0 \otimes \\ &\hat{\rho}_B^0 U_{UA}^\dagger U_{VB}^\dagger) = \text{tr}(U_{UA}^\dagger U_{VB}^\dagger \hat{\theta}_V U_{VB} U_{UA} \hat{\rho}_A^0 \otimes \hat{\rho}_{UV}^0 \otimes \hat{\rho}_B^0); \end{aligned}$$

since by locality,  $[\hat{\theta}_V, U_{UA}] = 0$ , one has

$$\begin{aligned} \langle \hat{\theta}_V \rangle_{\hat{\rho}_{BA}} &= \text{tr}(U_{VB}^\dagger \hat{\theta}_V U_{VB} \hat{\rho}_A^0 \otimes \hat{\rho}_{UV}^0 \otimes \hat{\rho}_B^0) = \\ &= \text{tr}(\hat{\theta}_V (U_{VB} \hat{\rho}_A^0 \otimes \hat{\rho}_{UV}^0 \otimes \hat{\rho}_B^0 U_{VB}^\dagger)) = \text{tr}(\hat{\theta}_V \hat{\rho}_B) = \langle \hat{\theta}_V \rangle_{\hat{\rho}_B}. \end{aligned} \quad (2.27)$$

That is, the mean value of the observable  $\theta_V$  of the subsystem V does not depend whether a measurement with the apparatus A is performed or not on the subsystem U.

Consider in particular  $\hat{\theta}_V = |\beta\rangle \otimes \langle \beta|$ , the projector associated with the eigenvalue,  $\beta$  of an observable  $\hat{\beta}$  of the subsystem V; from (2.13) its mean value is given by

$$\text{tr}(|\beta\rangle \otimes \langle \beta| \hat{\rho}_{BA}) = \text{tr}(|\beta\rangle \otimes \langle \beta| \hat{\rho}_B)$$

and therefore

$$\Sigma_{\beta'} \langle \beta' | \beta \rangle \langle \beta | \hat{\rho}_{BA} \rangle \beta' = \Sigma_{\beta'} \langle \beta' | \beta \rangle \langle \beta | \hat{\rho}_B \rangle \beta'$$

i.e.

$$\langle \beta | \hat{\rho}_{BA} \rangle \beta' = \langle \beta | \hat{\rho}_B \rangle \beta' \quad (2.28)$$

Then, from (2.14),

$$p_{BA}(\beta) = p_B(\beta). \quad (2.29)$$

So, in contradistinction with the result for a measurement performed on an individual quantum system (2.2.g.), (2.28) expresses the fact that *at the statistical level*, that is, at the *ensemble level*, the probability distributions remain unaltered by measurements on a distant subsystem.

2.2.i. Formulae analogous to (2.6) and (2.7), but for photons, can be easily derived<sup>11,12</sup>; these formulae are crucial for the comparison of the predictions of QM and the EPR and Bell theorems.

Let  $|H_l\rangle$  and  $|V_l\rangle$  represent respectively the states of horizontal and vertical photon polarizations in an EPR type experiment ( $l = 1, 2$ ), and

$$\begin{pmatrix} |H_l\rangle \\ |H_l^\perp\rangle \end{pmatrix} = \begin{pmatrix} \cos \theta_l & \sin \theta_l \\ -\sin \theta_l & \cos \theta_l \end{pmatrix} \begin{pmatrix} |H_l\rangle \\ |V_l\rangle \end{pmatrix} \quad (2.30)$$

with  $\theta_l^\perp = \theta_l + \frac{\pi}{2}$ , the polarization states corresponding to the ordinary ( $\theta_l$ ) and extraordinary ( $\theta_l^\perp$ ) photon "paths" in a calcite crystal detector. Then, the two-photon states

$$|\psi_I\rangle = \frac{1}{\sqrt{2}} (|V_1\rangle \otimes |H_2\rangle - |H_1\rangle \otimes |V_2\rangle), \quad (2.31)$$

decay product of the ground state of positronium (negative parity), and

$$|\psi_{II}\rangle = \frac{1}{\sqrt{2}} (|H_1\rangle \otimes |H_2\rangle + |V_1\rangle \otimes |V_2\rangle), \quad (2.32)$$

coming from a cascade process  $J = 0 \rightarrow 1 \rightarrow 0$  in calcium atoms (positive parity), with the help of (2.30) can be written as follows:

$$|\psi_I\rangle = \sum_{a,b \in \{1,\perp\}} A_{12}^{ab}(I) |\theta_1^a\rangle \otimes |\theta_2^b\rangle, \quad (2.31')$$

$$|\psi_{II}\rangle = \sum_{a,b \in \{1,\perp\}} A_{12}^{ab}(II) |\theta_1^a\rangle \otimes |\theta_2^b\rangle, \quad (2.32')$$

where  $\theta_1^1 \equiv \theta_1$ ,  $l = 1, 2$ , and

$$A_{12}^{11}(I) = A_{12}^{\perp\perp}(I) = \frac{1}{\sqrt{2}} \sin(\theta_1 - \theta_2), \quad (2.33a)$$

$$A_{12}^{1\perp}(I) = -A_{12}^{\perp 1}(I) = \frac{1}{\sqrt{2}} \cos(\theta_1 - \theta_2), \quad (2.33b)$$

$$A_{12}^{11}(II) = A_{12}^{\perp\perp}(II) = \frac{1}{\sqrt{2}} \cos(\theta_1 - \theta_2), \quad (2.34a)$$

$$A_{12}^{1\perp}(II) = -A_{12}^{\perp 1}(II) = -\frac{1}{\sqrt{2}} \sin(\theta_1 - \theta_2). \quad (2.34b)$$

Then, the corresponding *joint probabilities*, again the *unique* prediction of QM, are:

$$\begin{aligned} p_I(|\theta_1^a\rangle, |\theta_2^b\rangle) &= p_I(|\theta_1^{\perp a}\rangle, |\theta_2^{\perp b}\rangle) = \\ &= p_{II}(|\theta_1^a\rangle, |\theta_2^b\rangle) = p_{II}(|\theta_1^{\perp a}\rangle, |\theta_2^{\perp b}\rangle) = \frac{1}{2} \cos^2(\theta_1, \theta_2). \end{aligned} \quad (2.35a)$$

and

$$\begin{aligned} p_I(|\theta_1^{\perp a}\rangle, |\theta_2^b\rangle) &= p_I(|\theta_1^a\rangle, |\theta_2^{\perp b}\rangle) = \\ p_{II}(|\theta_1^a\rangle, |\theta_2^b\rangle) &= p_{II}(|\theta_1^{\perp a}\rangle, |\theta_2^{\perp b}\rangle) = \frac{1}{2} \cos^2(\theta_1, \theta_2). \end{aligned} \quad (2.35b)$$

Clearly,

$$\sum_{a,b \in \{1,\perp\}} |A_{12}^{ab}(I)|^2 = \sum_{a,b \in \{1,\perp\}} |A_{12}^{ab}(II)|^2 = 1 \quad (2.36)$$

i.e.

$$\sum_{a,b \in \{1,\perp\}} p_I(|\theta_1^a\rangle, |\theta_2^b\rangle) = \sum_{a,b \in \{1,\perp\}} p_{II}(|\theta_1^a\rangle, |\theta_2^b\rangle) = 1. \quad (2.36')$$

### 3. BELL'S THEOREM<sup>13</sup> (1964)

The simplest statement of the theorem is the following: *Quantum mechanics violates local realism.*

The incompleteness of QM as claimed by EPR for the description of an individual quantum system, led to the idea that the incorporation of *additional variables*  $\lambda$ , called *hidden variables*, could complete

the theory. They can be arbitrary in number (in general finite) and should explain realism.

#### Proof of Bell's theorem

The hypothesis are the same as for the EPR theorem, *plus* the assumption of the existence of hidden variables  $\lambda$ , such that the pair

$$(\psi, \lambda)$$

gives a complete description of the system. As in EPR, take the system described by the wave function (2.1) and let the SG which measures the projection of spin 1 (2) be in the direction specified by the unit vector  $\hat{a}$  ( $\hat{b}$ ) of  $\mathbb{R}^3$ . One assumes that in the laboratory both measurements are simultaneous and that the choice of  $\hat{a}$  and  $\hat{b}$  are random. By locality, the choice of  $\hat{a}$  does not affect that of  $\hat{b}$  and viceversa. Let  $A = A(\hat{a}, \hat{b}; \psi, \lambda)$  and  $B = B(\hat{b}, \hat{a}; \psi, \lambda)$  be functions which give the results of the measurements on the spins 1 and 2 respectively; for simplicity we normalize their values to  $\pm 1$ . Locality is taken into account if  $A = A(\hat{a}, \psi, \lambda)$  and  $B = B(\hat{b}, \psi, \lambda)$ . Bell also assumed that there is a classical statistical distribution of the variables  $\lambda$  in the set  $\Lambda$ , given by a function  $\mu(\lambda)$ , with  $\int_{\Lambda} d\lambda \mu(\lambda) = 1$ . Then the *average value of the product of the projections of the spins (correlation)* is given by

$$P(\hat{a}, \hat{b}; \psi) = \int_{\Lambda} d\lambda \mu(\lambda) A(\hat{a}; \psi, \lambda) B(\hat{b}; \psi, \lambda). \quad (3.1)$$

This is *not* a consequence of only QM, but of QM and the hypothesis ii) and iii) (locality and realism) of EPR, where realism is represented by the set of variables  $\lambda$ . For  $\hat{b} = \hat{a}$  one has the situation of EPR: perfect correlation, and  $PP(\hat{a}, \hat{a}; \psi) = -1$ ; so

$$0 = \int_{\Lambda} d\lambda \mu(\lambda) (A(\hat{a}; \psi, \lambda) B(\hat{a}; \psi, \lambda) + 1);$$

then (except possibly for a set of measure zero),  $A(\hat{a}; \psi, \lambda) B(\hat{a}; \psi, \lambda) = -1$  i.e.  $B(\hat{a}; \psi, \lambda) = -A(\hat{a}; \psi, \lambda)$ . Then

$$P(\hat{a}, \hat{b}; \psi) = -\int_{\Lambda} d\lambda \mu(\lambda) A(\hat{a}; \psi, \lambda) A(\hat{b}; \psi, \lambda).$$

Consider a third direction  $\hat{c}$ ; then

$$\begin{aligned} P(\hat{a}, \hat{b}; \psi) - P(\hat{a}, \hat{c}; \psi) &= \\ &= -\int_{\Lambda} d\lambda \mu(\lambda) A(\hat{a}; \psi, \lambda) A(\hat{b}; \psi, \lambda) - A(\hat{a}; \psi, \lambda) A(\hat{c}; \psi, \lambda). \\ &= -\int_{\Lambda} d\lambda \mu(\lambda) A(\hat{a}; \psi, \lambda) A(\hat{b}; \psi, \lambda) (1 - A(\hat{b}; \psi, \lambda) A(\hat{c}; \psi, \lambda)) \end{aligned}$$

and so

$$\begin{aligned}
|P(\hat{a}, \hat{b}; \psi) - P(\hat{a}, \hat{c}; \psi)| &\leq \int_{\Lambda} d\lambda \mu(\lambda) |1 - A(\hat{b}; \psi, \lambda)A(\hat{c}; \psi, \lambda)| = \\
&= \int_{\Lambda} d\lambda \mu(\lambda) (1 - A(\hat{b}; \psi, \lambda)A(\hat{c}; \psi, \lambda)) = \\
&= 1 - \int_{\Lambda} d\lambda \mu(\lambda) A(\hat{b}; \psi, \lambda)A(\hat{c}; \psi, \lambda)
\end{aligned}$$

i.e.

$$|P(\hat{a}, \hat{b}; \psi) - P(\hat{a}, \hat{c}; \psi)| \leq 1 + P(\hat{b}, \hat{c}; \psi). \quad (3.2)$$

This is the simplest Bell's inequality (Bell, 1964). Notice that it does not depend on  $\lambda$ 's since these variables have been integrated.

The quantum prediction for  $P(\hat{a}, \hat{b}; \psi)$ , which we denote by  $P_q(\hat{a}, \hat{b}; \psi)$ , is given by

$$P_q(\hat{a}, \hat{b}; \psi) = -\hat{a} \cdot \hat{b} \quad (3.3)$$

(see Appendix). We prove now that (3.2) and (3.3) are contradictory: (3.2) amounts to

$$-1 - P(\hat{b}, \hat{c}; \psi) \leq P(\hat{a}, \hat{b}; \psi) - P(\hat{a}, \hat{c}; \psi) \leq 1 + P(\hat{b}, \hat{c}; \psi);$$

taking  $\hat{a}, \hat{b}$  and  $\hat{c}$  in a plane, with  $(\hat{a}, \hat{b}) = (\hat{b}, \hat{c}) = \frac{1}{2}(\hat{a}, \hat{c}) = \theta \in [0, \frac{\pi}{2}]$ , one obtains  $P_q(\hat{a}, \hat{b}; \psi) = P_q(\hat{b}, \hat{c}; \psi) = -\cos\theta$ ,  $P_q(\hat{a}, \hat{c}; \psi) = -\cos 2\theta$ . If QM reproduces Bell-EPR, then one should have

$$-1 + \cos\theta \leq -\cos\theta + \cos 2\theta \leq 1 - \cos\theta.$$

The second inequality holds since  $\cos 2\theta \leq 1$ ; however, for the first inequality,  $2\cos\theta \leq 1 + \cos 2\theta \leq 1 + \cos^2\theta - \sin^2\theta = 2\cos^2\theta$  i.e.  $\cos\theta \leq \cos^2\theta$ . If  $\theta = \frac{\pi}{2}$  (0) then  $0 \leq 0$  ( $1 \leq 1$ ), but if  $\theta \in (0, \frac{\pi}{2})$  then

$$1 \leq \cos\theta \quad (3.4)$$

which is *false*. Then, *the Bell-EPR hypothesis can not reproduce the predictions of QM*. QED

Notice that for the case of *perfect correlation*,  $\hat{a} = \hat{b}$  and then

$$P(\hat{a}, \hat{a}; \psi) = P_q(\hat{a}, \hat{a}; \psi) = -1 \quad (3.5)$$

i.e. EPR-Bell reproduces the QM result.

In 1969, Clauser **et al.**<sup>14</sup> derived another inequality which proved to be extremely useful for experimental purposes. If  $x, y, x'$  and  $y'$  take values in  $\{+1, -1\}$ , then the following equality holds:

$$xy - xy' + x'y + x'y' = \pm 2. \quad (3.6)$$

In fact, the left hand side of (3.6) can be written as

$$x(y - y') + x'(y + y')$$

and  $y - y'$  or  $y + y'$  equals  $\pm 2$ . Identifying:

$$x = A(\hat{a}; \psi, \lambda), y = B(\hat{b}; \psi, \lambda),$$

$$x' = A(\hat{a}'; \psi, \lambda) \text{ and } y' = B(\hat{b}'; \psi, \lambda),$$

multiplying by  $\mu(\lambda)$ , and integrating over  $\lambda$ , one has the CNSH inequality:

$$-2 \leq P(\hat{a}, \hat{b}; \psi) - P(\hat{a}, \hat{b}'; \psi) + P(\hat{a}', \hat{b}; \psi) - P(\hat{a}', \hat{b}'; \psi) \leq 2 \quad (3.7)$$

(since  $|\int \mu(xy - xy' + x'y + x'y')| \leq \int |\mu| |xy - xy' + x'y + x'y'| = 2 \int \mu = 2$ ).

It is easy to see how (3.7) is violated by QM<sup>15</sup>: choose the four unit vectors  $\hat{b}', \hat{a}', \hat{b}$  and  $\hat{a}$  in a plane with (say counterclockwise) angles given by  $\hat{b}', \hat{a}' = (\hat{a}', \hat{b}') = (\hat{b}, \hat{a}) = \frac{\pi}{4}$ ; then, replacing  $P$ 's by  $P_q$ 's, for the middle term in (3.7) one obtains

$$-\cos \frac{\pi}{4} + \cos \frac{3}{4}\pi - \cos \frac{\pi}{4} - \cos \frac{\pi}{4} = -2\sqrt{2},$$

which violates the inequality.

Almost all the experimental results confirm the violation by QM of the above Bell's inequalities and all other inequalities which have been obtained afterwards. A summary of the experimental situation can be found in Aspect<sup>16</sup>, Zeilinger<sup>17</sup>, and Weinfurter<sup>18</sup>.

In (2.2.g.) we have shown that QM is non local (a' la EPR), in agreement with the theoretical and experimental violation of the Bell's inequalities. The question of its completeness or not is solved in section 4.

#### 4. BELL'S THEOREM WITHOUT INEQUALITIES: GREENBERGER-HORNE-ZEILINGER<sup>19,20</sup> (1989)

The GHZ's theorem says that the *EPR hypothesis contradict quantum mechanics*.

We shall follow the presentation of Mermin<sup>21,22</sup>.

## Proof of the GHZ's theorem

Consider three spin  $\frac{1}{2}$  particles, 1, 2 and 3, which are the products of the decay of an initial particle (source). The Hilbert space  $\mathcal{H}$  of the spin part of the decaying particles is isomorphic to  $(\mathbb{C}^2)^3 \cong \mathbb{C}^8$ . After the decay, the three particles move freely (and therefore without interaction) along straight lines at 120° from each other, in a horizontal plane yz (one approximates the orbital motion by classical trajectories). For  $k = 1, 2, 3$ , we call  $z_k$  the direction of motion of particle  $k$ ,  $x_1 = x_2 = x_3 = x$  the direction normal to the plane, and  $y_k$  the direction in the plane normal to  $z_k$ . Along each path  $z_k$ , one sets two SG apparatuses to measure the spin projections in the directions  $x$  and  $y_k$ . Let  $\uparrow_k$  ( $\downarrow_k$ ) be the eigenstates of the spin operator  $\sigma_{z_k}$  with eigenvalue  $+1$  ( $-1$ ), and let the normalized state vector of the system of the decaying three particles be given by

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow_1\uparrow_2\uparrow_3\rangle - |\downarrow_1\downarrow_2\downarrow_3\rangle) \quad (4.1)$$

$|\psi\rangle$  is symmetric under the interchange of the particles.

Consider the spin operators

$$A = \sigma_{x_1}\sigma_{y_2}\sigma_{y_3}, \quad B = \sigma_{y_1}\sigma_{x_2}\sigma_{y_3}, \quad C = \sigma_{y_1}\sigma_{y_2}\sigma_{x_3} \quad (4.2)$$

They have the following properties:

i) Hermiticity:

$$A^\dagger = (\sigma_{x_1}\sigma_{y_2}\sigma_{y_3})^\dagger = \sigma_{y_3}^\dagger\sigma_{y_2}^\dagger\sigma_{x_1}^\dagger = \sigma_{y_3}\sigma_{y_2}\sigma_{x_1}, \quad \sigma_{x_3}\sigma_{y_2}\sigma_{y_3} = A, \quad (4.3)$$

since spin operators for different particles commute with each other; then also

$$B^\dagger = B \text{ and } C^\dagger = C. \quad (4.3')$$

ii) Commutativity:

$$\begin{aligned} AB &= \sigma_{x_1}\sigma_{y_2}\sigma_{y_3}\sigma_{y_1}\sigma_{x_2}\sigma_{y_3} = (-1)\sigma_{y_1}\sigma_{x_1}\sigma_{y_2}\sigma_{y_3}\sigma_{x_2}\sigma_{y_3} = \\ &= (-1)^2\sigma_{y_1}\sigma_{x_2}\sigma_{y_3}\sigma_{x_1}\sigma_{y_2}\sigma_{y_3} = BA, \end{aligned} \quad (4.4)$$

etc.

iii) Eigenvalues  $\pm 1$ :

$$A^2 = B^2 = C^2 = 1 \quad (4.5)$$

since e.g.

$$A^2 = \sigma_{x_1}\sigma_{y_2}\sigma_{y_3}\sigma_{x_1}\sigma_{y_2}\sigma_{y_3} = \sigma_{x_1}^2\sigma_{y_2}^2\sigma_{y_3}^2 = 1.$$

Then the set of eight eigenstates of A, B and C,

$$\begin{aligned} \{ &|1,1,1\rangle, |-1,1,1\rangle, |1,-1,1\rangle, |1,1,-1\rangle, |1,-1,-1\rangle, \\ &|-1,1,-1\rangle, |-1,-1,1\rangle, |-1,-1,-1\rangle \} \end{aligned} \quad (4.6)$$

form a basis of  $\mathcal{H}$  and therefore A, B and C is a complete set of commuting observables of the particles 1, 2 and 3. (The entries in  $1a, b, c$ , with  $a, b, c = \pm 1$ , respectively are the eigenvalues of A, B and C.)

iv)  $|\psi\rangle$  is eigenstate of A, B and C with eigenvalue equal to 1 i.e.

$$A|\psi\rangle = B|\psi\rangle = C|\psi\rangle = |\psi\rangle. \quad (4.7)$$

This can be easily verified taking into account that for any of the three particles,

$$\begin{aligned} \sigma_{x_k}|\uparrow_k\rangle &= |\downarrow_k\rangle, \quad \sigma_{x_k}|\downarrow_k\rangle = |\uparrow_k\rangle, \quad \sigma_{y_k}|\uparrow_k\rangle = i|\downarrow_k\rangle, \\ \sigma_{y_k}|\downarrow_k\rangle &= -i|\uparrow_k\rangle \text{ (in fact } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{ etc.)} \end{aligned}$$

$$\begin{aligned} \text{Then, } A|\psi\rangle &= \frac{1}{\sqrt{2}} (\sigma_{x_1}\sigma_{y_2}\sigma_{y_3}|\uparrow_1\uparrow_2\uparrow_3\rangle - \\ &- \sigma_{x_1}\sigma_{y_2}\sigma_{y_3}|\downarrow_1\downarrow_2\downarrow_3\rangle) = \\ &= \frac{1}{\sqrt{2}} (i^2)|\uparrow_1\uparrow_2\uparrow_3\rangle - (-i^2)|\uparrow_1\uparrow_2\uparrow_3\rangle) = |\psi\rangle, \end{aligned}$$

etc.

Suppose that on particles 3 and 2 one measures the spin projections with SG's in directions  $y_3$  and  $y_2$  and obtain, say, the values  $+1$  and  $-1$  respectively. Then one can predict with certainty (perfect correlation) that the spin projection of particle 1 in the  $x_1$  direction is  $-1$  since the state  $|\psi\rangle$  is an eigenstate of  $A = \sigma_{x_1}\sigma_{x_2}\sigma_{x_3}$  with eigenvalue  $+1$ , and so the product of the three spin projections must be  $+1$ . Assuming *locality* (EPR hypothesis ii) in section 2), the measurements on particles 2 and 3 do not disturb particle 1 since the particles are far enough so that they do not interact with each other. Then according to the criterion iii) of EPR (section 2), the eigenvalue  $-1$  of  $\sigma_{x_1}$  is an *element of physical reality*. The same analysis can be repeated for other measurements and one concludes with EPR that the six eigenvalues of  $\sigma_{x_1}, \sigma_{y_1}, \sigma_{x_2}, \sigma_{y_2}, \sigma_{x_3}, \sigma_{y_3}$ , respectively  $m_{x_1}, m_{y_1}, m_{x_2}, m_{y_2}, m_{x_3}, m_{y_3} \in \{1, -1\}$  are elements of the physical reality. Since  $\sigma_{x_k}$  and  $\sigma_{y_k}$  for  $k = 1, 2, 3$  can not have a common eigenvector, this leads to the conclusions that, as in the case of EPR, QM is not a complete theory, or, if it is complete, then it is in conflict with local realism. Since locality is violated (2.2.g), there remain the two possibilities: QM is complete or realist.

Assuming however the existence of the above elements of reality, they obey

$$m_{x_1} m_{x_2} m_{x_3} = 1 \quad (4.8)$$

since  $1 = m_{x_1} m_{y_2} m_{y_3} m_{y_1} m_{x_2} m_{y_3} m_{y_1} m_{y_2} m_{x_3} = m_{x_1} m_{x_2} m_{x_3}$ . But then consider the operator

$$D = \sigma_{x_1} \sigma_{x_2} \sigma_{x_3}. \quad (4.9)$$

D commutes with A, B and C: in fact

$AD = \sigma_{x_1} \sigma_{y_2} \sigma_{y_3} \sigma_{x_1} \sigma_{x_2} \sigma_{x_3} = \sigma_{x_1} \sigma_{x_1} \sigma_{y_2} \sigma_{y_3} \sigma_{x_2} \sigma_{x_3} = (-1) \sigma_{x_1} \sigma_{x_2} \sigma_{x_1} \sigma_{y_2} \sigma_{y_3} \sigma_{x_3} = (-1)^2 \sigma_{x_1} \sigma_{x_2} \sigma_{x_3} \sigma_{x_1} \sigma_{y_2} \sigma_{y_3} = DA$ , etc. Since A, B and C are a complete set of observables then  $|\psi\rangle$  is also an eigenstate of D with eigenvalue, according to EPR, equal to  $m_{x_1} m_{x_2} m_{x_3} = 1$ . However, as it can be easily verified,

$$D = -ABC, \quad (4.10)$$

(in fact  $-\sigma_{x_1} \sigma_{y_2} \sigma_{y_3} \sigma_{y_1} \sigma_{x_2} \sigma_{y_3} \sigma_{y_1} \sigma_{y_2} \sigma_{x_3} = -(-1)(\sigma_{y_2})^2 (\sigma_{y_3})^2 (\sigma_{y_1})^2 \sigma_{x_1} \sigma_{x_2} \sigma_{x_3} = \sigma_{x_1} \sigma_{x_2} \sigma_{x_3}$ ) and then according to QM,

$$D|\psi\rangle = -ABC|\psi\rangle \quad (4.11)$$

i.e.  $|\psi\rangle$  is an eigenvector of D with eigenvalue -1, contrary to the prediction of EPR i.e. QM contradicts realism. QED

**Corollary:** QM is a complete theory.

There is an argument in favor of the statement: *either QM is incomplete or it is non local*, which in particular implies that if *QM is complete then it is non local*. The argument is based on the so called "Einstein boxes" <sup>11,23</sup>. If a box initially containing a quantum particle is separated into two boxes A and B, then the wave function of the particle becomes

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|A\rangle + |B\rangle) \text{ where } \langle A|\psi\rangle = \langle B|\psi\rangle$$

gives the probability amplitude to find the particle in box A (B). If at B an observer finds (does not find) the particle, he can not however distinguish between the following two alternatives:

- i) A did not measure
- ii) A measured but did not find (found) the particle.

This is the analogous situation to that discussed in 2.2.g. Then the observer at A can not send an instantaneous (and therefore superluminal) information (or signal) to the observer at B through this mechanism. A density matrix analysis analogous to that in 2.2.g. can be made here which shows that, even if no instantaneous signal can be transmitted, there is a collapse  $\hat{\rho}_B^{\text{red}} \rightarrow |B\rangle$  (or  $\hat{\rho}_A^{\text{red}} \rightarrow |A\rangle$ ). In fact,

$$\hat{\rho} = |\psi\rangle \otimes \langle\psi| = \frac{1}{2} (|A\rangle \otimes \langle A| + |B\rangle \otimes \langle B| + |A\rangle \otimes \langle B| + |B\rangle \otimes \langle A|);$$

$$\hat{\rho}_A^{\text{red}} = \hat{\rho}_B^{\text{red}} = \text{tr}_{\hat{A}p} = \text{tr}_{\hat{B}p} = \frac{1}{2} I$$

which imply

$$(\hat{\rho}_A^{\text{red}})^2 = (\hat{\rho}_B^{\text{red}})^2 = \frac{1}{4} I \neq \frac{1}{2} I,$$

$$p(A) = p(B) = \langle A|\hat{\rho}_A^{\text{red}}|A\rangle = \langle B|\hat{\rho}_B^{\text{red}}|B\rangle = \frac{1}{2}.$$

It is clear that p(A) and p(B) respectively are the joint probabilities p(A, - B) and p(- A, B) where - B (- A) means that the particle is not found at B (A). Then for the conditional probabilities:

$$p(- A | B) = \frac{p(-A, B)}{p(B)} = \frac{p(B)}{p(B)} = 1,$$

$$p(A | - B) = \frac{p(A, -B)}{p(-B)} = \frac{p(A)}{p(A)} = 1,$$

(certainty).

## 5. HARDY'S THEOREM<sup>24</sup> (1992)

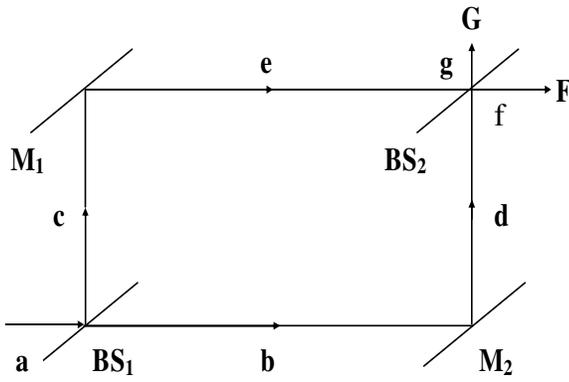
The Hardy's theorem has two parts:

- a) QM contradicts local realism.
- b) Elements of reality a' La EPR, corresponding to Lorentz invariant observables, are not Lorentz invariant.

The main interest for the 2<sup>nd</sup> part of the theorem, was the possibility of the existence of a preferred reference frame -like the microwave background radiation- to avoid paradoxes like going backwards in time if superluminal signals or instantaneous actions at a distance are allowed. However, as discussed in subsections 2.2.g. and 2.2.h., even if there are non local instantaneous effects, they do not consist in sending signals or information at velocities greater than c, and so paradoxes associated to that possibility are absent.

### 5.1. Mach-Zehnder interferometer

The scheme of the MZ apparatus is given in Figure 1. We describe it here in terms of photons, but the treatment for electrons and positrons in Hardy's theorem (subsection 5.2.) is similar.



**Figure 1.** Scheme of the Mach-Zehnder interferometer.

$BS_k$ ,  $k = 1, 2$ , are beam-splitters (half-silvered mirrors in the case of photons) and  $M_l$ ,  $l = 1, 2$  are totally reflecting mirrors; G and F are detectors. a, b, c, d, e, f and g are the quantum states (vectors in a Hilbert space) of a particle taking the corresponding "path". (A complete treatment should use a path integral.) As usual, a reflection on a splitter or a mirror introduces a 90 phase (factors  $i$ ). Consider the following two possibilities:

i) With  $BS_2$  present

The successive quantum states of the incident particle are given by the following chain:

$$\begin{aligned}
 a &\xrightarrow{BS_1} \frac{1}{\sqrt{2}}(b + ic) \xrightarrow{M_2, M_1} \frac{1}{\sqrt{2}}(id + i(ie)) \\
 &\xrightarrow{BS_2} \frac{1}{\sqrt{2}}\left(i\frac{1}{\sqrt{2}}(g + if) - \frac{1}{\sqrt{2}}(f + ig)\right) = -f \quad (5.1)
 \end{aligned}$$

which means that only the detector F clicks i.e. one has constructive interference at F and destructive interference at G: the particle "goes through" two paths (zwei Wegs); in other words, the presence of  $BS_2$  makes the particle to exhibit a wave-like nature. Notice that the detector which clicks is the one in the direction of the incident particle. In the analogy with the two slits Young experiment<sup>25</sup>, the situation is equivalent to that with the two slits opened.

ii) With  $BS_2$  removed

$$a \rightarrow \dots \frac{1}{\sqrt{2}}(id - e) \rightarrow \frac{1}{\sqrt{2}}(ig - f) \quad (5.2)$$

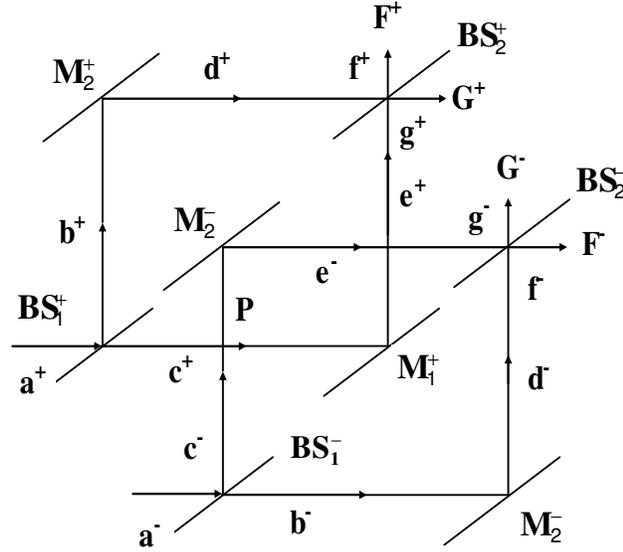
which means that both detectors G and F click, each with probability  $\frac{1}{2}$ . If G(F) clicks, one knows that

the particle "went through" the path b - d - g (c - e - f) (*welches Weg*) exhibiting a particle-like nature. This is analogous to block one of the two slits in the Young double slit experiment (no interference).

As Wheeler<sup>26</sup> has realized,  $BS_2$  can be inserted or removed at the last instants of the experiment i.e. much later than the moment in which the particle enters the interferometer; then one can decide whether *the particle goes through one path (particle-like behavior) or through two paths (wave-like behavior) after the particle has gone through one of these two alternatives!* But this means that one is acting on the past!?

## 5.2. Proof of Hardy's theorem

Consider the scheme in Figure 2, where there are two MZ interferometers, one for positrons (+) and one for electrons (-).



**Figure 2.** Scheme of two MZ interferometers, one for positrons (+) and one for electrons (-).

5.2.1. The possibility of electron-positron annihilation at P, assumed with probability 1 if the electron and positron travel through the intersecting paths, allows the detection of particles at  $G^+$  and  $G^-$  (interference destroying alternative) with or without the beam splitters  $BS_2^\pm$  installed. It is clear that if the detection at  $G^+$  and  $G^-$  occurs, then annihilation at P has not taken place; this is an example of a counterfactual event. Then, from a semiclassical point of view, if the positron path was  $a^+ \rightarrow c^+ \rightarrow e^+$ , then the electron path was  $a^- \rightarrow b^- \rightarrow d^-$ , and if the electron path was  $a^- \rightarrow c^- \rightarrow e^-$ , then the positron path was  $a^+ \rightarrow b^+ \rightarrow d^+$ .

As in 5.1., we follow the successive quantum states of the incident particles:

$$a^\pm \xrightarrow{BS_1^\pm} \frac{1}{\sqrt{2}}(b^\pm + ic^\pm);$$

the initial state is  $a^+a^- = |a^+, a^- \rangle = |a^+ \rangle \otimes |a^- \rangle$  and therefore

$$\begin{aligned} a^+a^- &\xrightarrow{BS_1^+} \frac{1}{2}(b^+ + ic^+)(b^- + ic^-) = \\ &= \frac{1}{2}(b^+b^- + ib^+c^- + ic^+b^- - c^+c^-); \end{aligned}$$

since  $c^+c^-$  annihilate at P i. e.

$$c^+c^- \xrightarrow{P} \gamma's \text{ (photons)}$$

one has

$$\begin{aligned} a^+a^- &\xrightarrow{BS_1^+} \frac{1}{2}(-\gamma's + b^+b^- + ib^+c^- + ic^+b^-) \xrightarrow{M_1^+, M_2^+} \\ &\frac{1}{2}(-\gamma's - d^+d^- - id^+e^- + ie^+d^-) \equiv \psi. \end{aligned} \quad (5.3)$$

For the beam splitters  $BS_2^\pm$  we have four possibilities:

i) Both  $BS_2^\pm$  active:

$$d^\pm \xrightarrow{BS_2^\pm} \frac{1}{\sqrt{2}}(if^\pm + g^\pm), e^\pm \rightarrow \frac{1}{\sqrt{2}}(f^\pm + ig^\pm);$$

then

$$a^+a^- \rightarrow \psi \rightarrow \frac{1}{4}(-2\gamma's + 3f^+f^- - if^+g^- - ig^+f^- + g^+g^-). \quad (5.4)$$

Probabilities check:  $(\frac{2}{4})^2 + (\frac{3}{4})^2 + 3 \times (\frac{1}{4})^2 = 1$ .

ii)  $BS_2^-$  active and  $BS_2^+$  removed:

$$\begin{aligned} d^+ \rightarrow g^+, e^+ \rightarrow f^+, d^- &\xrightarrow{BS_2^-} \frac{1}{\sqrt{2}}(if^- + g^-), \\ e^- &\xrightarrow{BS_2^-} \frac{1}{\sqrt{2}}(f^- + ig^-); \end{aligned}$$

then

$$a \rightarrow \psi \rightarrow \frac{1}{2\sqrt{2}}(-\sqrt{2}\gamma's + f^+f^- - i(2g^+f^- + f^+g^-)). \quad (5.5)$$

Probabilities check:  $(\frac{1}{2})^2 + 2 \times (\frac{1}{2\sqrt{2}})^2 + (\frac{1}{2\sqrt{2}})^2 = 1$ .

iii)  $BS_2^+$  active and  $BS_2^-$  removed:

$$\begin{aligned} d^- \rightarrow g^-, e^- \rightarrow f^-, d^+ &\xrightarrow{BS_2^+} \frac{1}{\sqrt{2}}(if^+ + g^+), \\ e^+ &\xrightarrow{BS_2^+} \frac{1}{\sqrt{2}}(ig^+ + f^+); \end{aligned}$$

then

$$a^+a^- \rightarrow \psi \rightarrow \frac{1}{2\sqrt{2}}(\sqrt{2}\gamma's + f^+f^- - i(2f^+g^- + g^+f^-)). \quad (5.6)$$

Probabilities check:  $(\frac{1}{2})^2 + 2 \times (\frac{1}{2\sqrt{2}})^2 + (\frac{1}{2\sqrt{2}})^2 = 1$ .

iv) Both  $BS_2^\pm$  removed:

$$d^\pm \rightarrow g^\pm, e^\pm \rightarrow f^\pm;$$

then

$$a^+a^- \rightarrow \psi \rightarrow \frac{1}{2}(\gamma's + g^+g^- + i(g^+f^- + f^+g^-)). \quad (5.7)$$

Probabilities check:  $4 \times (\frac{1}{2})^2 = 1$ .

The results (5.3)-(5.7) are a prediction of QM.

5.2.2. One assumes locality and realism by introducing the following eight functions:

$$F^\pm(0, \lambda), G^\pm(0, \lambda) = \begin{cases} 1, & \text{if a positron (electron)} \\ & \text{is detected at } F^\pm, G^\pm \\ & \text{with } BS_2^+ \text{ installed} \\ 0, & \text{if a positron (electron)} \\ & \text{is not detected at } F^\pm, G^\pm \\ & \text{with } BS_2^+ \text{ installed} \end{cases}$$

$$F^\pm(\infty, \lambda), G^\pm(\infty, \lambda) = \begin{cases} 1, & \text{if a positron (electron)} \\ & \text{is detected at } F^\pm, G^\pm \\ & \text{with } BS_2^+ \text{ removed} \\ 0, & \text{if a positron (electron)} \\ & \text{is not detected at } F^\pm, G^\pm \\ & \text{with } BS_2^+ \text{ removed} \end{cases}$$

Realism consists in the assumption that these functions of the hidden variable  $\lambda$  exist, while locality consists in the assumption that the functions for positrons (electrons) depend only of  $BS_2^+(BS_2^-)$ . Clearly, the F and G functions are the analogous of the A and B functions of Bell, while the beam splitters  $BS_2^\pm$  play the rôle of the SG's represented by  $\hat{a}$  and  $\hat{b}$ , in section 3.

From case iv), no  $f^+f^-$  term appears in (5.6), then

$$F^+(\infty, \lambda)F^-(\infty, \lambda) = 0 \quad (5.8)$$

in all experiments with both  $BS_2^\pm$  removed.

From case iii),

$$\text{if } G^+(0, \lambda) = 1 \text{ then } F^-(\infty, \lambda) = 1 \quad (5.9)$$

since the term  $g^+f^-$  appears in (5.5); this happens for  $(\frac{1}{2\sqrt{2}})^2 = \frac{1}{8}$  of the experiments with  $BS_2^+$  in place and

$BS_2^-$  removed.

From case ii),

$$\text{if } G^-(0,\lambda) = 1 \text{ then } F^+(\infty,\lambda) = 1 \quad (5.10)$$

since the term  $f^+g^-$  appears in (5.5); this happens for  $(\frac{1}{2\sqrt{2}})^2 = \frac{1}{8}$  of the experiments with  $BS_2^+$  removed and  $BS_2^-$  in place.

From case i),

$$G^+(0,\lambda)G^-(0,\lambda) = 1 \quad (5.11)$$

for  $(\frac{1}{4})^2 = \frac{1}{16}$  of the experiments with both  $BS_2^+$  present.

Since  $F^\pm(0, \lambda)$ ,  $F^\pm(\infty, \lambda)$ ,  $G^\pm(0, \lambda)$  and  $G^\pm(\infty, \lambda)$  are functions of  $\lambda$ , we can compare them independently of the settings  $BS_2^\pm$  ( $k = 1,2$ ) which determine their values.

From (5.11),  $G^+(0, \lambda) = G^-(0, \lambda) = 1$ , and from (5.9) and (5.10),  $F^+(\infty, \lambda) = F^-(\infty, \lambda) = 1$  which implies  $F^+(\infty, \lambda)F^-(\infty, \lambda) = 1$ , which is in contradiction with (5.8). Then, *QM contradicts local realism*. This is the Bell theorem without inequalities for two particles.

5.2.3. Let us assume with Hardy that *if an element of physical reality corresponds to a Lorentz invariant observable, then the numerical value of the element of physical reality is itself Lorentz invariant*. For example, if  $|a\rangle$  is an eigenstate of the observable  $A$  with eigenvalue  $a$ , i.e.  $A|a\rangle = a|a\rangle$  and  $A$  is Lorentz invariant, then  $[A] = a$  is Lorentz invariant, where  $A$  is the element of physical reality corresponding to  $A$  and  $[A]$  is its value (we have used the EPR sufficient criterion of reality iii) of section 2).

Define the operators

$$\hat{E}^\pm = |e^\pm\rangle \otimes \langle e^\pm|, \quad \hat{E} = \hat{E}^+\hat{E}^- \quad (5.12)$$

$\hat{E}^\pm$  are Lorentz invariant since they are the projectors onto the corresponding "arms" of the MZ's, and  $\hat{E}$  is Lorentz invariant since it is the product of Lorentz invariant operators. Since  $(\hat{E}^\pm)^2 = \hat{E}^\pm$ , and  $(\hat{E}^+\hat{E}^-)^2 = (\hat{E}^+\hat{E}^-)$  since  $\hat{E}^+$  and  $\hat{E}^-$  commute, their eigenvalues are 0 and 1. Clearly,

$$\hat{E}^\pm |e^\pm\rangle = |e^\pm\rangle, (\hat{E}^+\hat{E}^-)|e^+, e^-\rangle = |e^+, e^-\rangle \quad (5.13)$$

and therefore

$$[\varepsilon^+] = [\varepsilon^+\varepsilon^-] = 1. \quad (5.14)$$

Also, if  $|e^+, e^-\rangle_\perp$  is a state vector orthogonal to  $|e^+, e^-\rangle$ , then

$$(\hat{E}^+\hat{E}^-)|e^+, e^-\rangle_\perp = 0 \text{ implies } [\varepsilon^+, \varepsilon^-] = 0 \quad (5.15)$$

since  $\mathcal{H} \ni 0 = 0 |e^+, e^-\rangle_\perp$  where in the right hand side  $0 \in \mathbb{C}$ .

The other way around, if a system has an element of physical reality  $A$  corresponding to the observable  $A$  with value  $a$ , then the state vector of the system is an eigenvector of the operator  $A$  with eigenvalue  $a$ .

In particular,

$$\text{if } [\varepsilon^+][\varepsilon^-] = 1 \text{ then } [\varepsilon^+] = [\varepsilon^-] = 1 \text{ and therefore } [\varepsilon^+\varepsilon^-] = 1 \quad (5.16)$$

since the state is  $|e^+, e^-\rangle$  and  $(\hat{E}^+\hat{E}^-)|e^+, e^-\rangle =$

$$= (\hat{E}^+|e^+\rangle \otimes \hat{E}^-|e^-\rangle) = |e^+\rangle \otimes |e^-\rangle = |e^+, e^-\rangle.$$

Let  $K_+(K_-)$  be a reference frame in which the positron (electron) is observed before the electron (positron); if the positron (electron) has already passed through  $BS_2^+$  ( $BS_2^-$ ) but the electron (positron) has not yet passed through  $BS_2^+$  ( $BS_2^-$ ), then the process in the MZ's is given by

$$a^+a^- \rightarrow \frac{1}{2\sqrt{2}}(-\sqrt{2}\gamma s - 2if^+d^- + f^+e^- - ig^+e^-) \quad (5.17)$$

since in (5.3),  $d^+ \rightarrow \frac{1}{\sqrt{2}}(if^+ + g^+)$ ,  $e^+ \rightarrow \frac{1}{\sqrt{2}}(ig^+ + f^+)$  but  $d^- \rightarrow d^-$  and  $e^- \rightarrow e^-$

$$(a^+a^- \rightarrow \frac{1}{2\sqrt{2}}(-\sqrt{2}\gamma s - 2id^+f^- + e^+f^- - ie^+g^-) \quad 5.18)$$

since in (5.3),  $d^- \rightarrow \frac{1}{\sqrt{2}}(if^- + g^-)$ ,  $e^- \rightarrow \frac{1}{\sqrt{2}}(f^- + ig^-)$  but  $d^+ \rightarrow d^+$  and  $e^+ \rightarrow e^+$

Then:

In  $K_+$ , if the positron is detected in  $G^+$ , then the electron collapses to the state  $-ie^-$  and therefore

$$[\varepsilon^-] = 1. \quad (5.19)$$

In  $K_-$ , if the electron is detected in  $G^-$ , then the positron collapses to the state  $-ie^+$  and therefore

$$[\varepsilon^+] = 1. \quad (5.20)$$

In  $K_0$ , it is easy to verify by an explicit calculation that

$$(\hat{E}^+\hat{E}^-)\psi = 0 \quad (5.21)$$

with  $\psi$  given in (5.3); then

$$[\varepsilon^+ \varepsilon^-] = 0. \quad (5.22)$$

Clearly, this result contradicts (5.16) together with (5.19) and (5.20). The contradiction emerges because one is comparing the values of elements of physical reality in different reference frames, assuming that they were Lorentz invariant. QED

The contradiction can be seen in another way: In  $K_+$ ,  $[\varepsilon^-] = 1$  means that the electron went through  $e^-$ , then the positron went through  $b^+$  to avoid annihilation at P; in  $K_-$ ,  $[\varepsilon^+] = 1$  means that the positron went through  $e^+$

and therefore the electron through  $b^-$  again to avoid annihilation at P. So, in different frames the trajectories are not the same; moreover, the trajectories predicted in  $K_+$  and  $K_-$  are contradictory if we stay at  $K_0$  with both  $G^+$  and  $G^-$  detecting the particles.

#### ACKNOWLEDGMENT

The author thanks the graduate student Brenda Carballo for enlightened discussions. This work was partially supported by the project PAPIIT IN103505, DGAPA-UNAM, Mexico.

#### APPENDIX

##### Proof of (3.3)

Let  $|+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $|-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  be the eigenstates of  $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$  with eigenvalues +1 and -1. In a direction  $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ , the normalized eigenvector of  $\vec{\sigma} \cdot \hat{n}$  with eigenvalue + 1/2 is  $|\hat{n}^+\rangle = \cos\frac{\theta}{2}|+\rangle + e^{i\phi}\sin\frac{\theta}{2}|-\rangle$ . Prepare the singlet state of two particles  $|\psi\rangle = \frac{1}{\sqrt{2}}(|1, +\rangle \otimes |2, -\rangle - |1, -\rangle \otimes |2, +\rangle)$ . The probability of finding particle 1 in the

+direction of  $\sigma_3$  is  $\frac{1}{2}$  then particle 2 will be found with certainty along  $z$  with eigenvalue -1; then the probability of finding it in the direction  $+\hat{n}$  is  $|\langle 2, -|\hat{n}^+\rangle|^2 = \sin^2\frac{\theta}{2}$ . Then  $P_{++} = \frac{1}{2} \times \sin^2\frac{\theta}{2}$ . Similarly  $P_{--} = P_{++}$  while  $P_{+-} = P_{-+} = \frac{1}{2} \sin^2\frac{\pi-\theta}{2} = \frac{1}{2} \cos^2\frac{\theta}{2}$ . Then  $P_q(\hat{z}, \hat{n}) = (+1)(P_{++} + P_{--}) + (-1)(P_{+-} + P_{-+}) = \sin^2\frac{\theta}{2} - \cos^2\frac{\theta}{2} = -\cos\theta = -\hat{z} \cdot \hat{n} \cdot z$ . By spherical symmetry, this holds for arbitrary  $a$  and  $b$ . QED

#### REFERENCES

1. EINSTEIN, A.; B. PODOLSKY and N. ROSEN (1935): "Can Quantum-Mechanical Description of "Physical Reality be Considered Complete?", **Phys. Rev.** 47, 777-780.
2. SCHRÖDINGER, E. (1935): "Discussion of Probability Relations Between Separated Systems", **Proc. Camb. Phil. Soc.** 31, 555-563.
3. EINSTEIN, A. (1949): "Autobiographical Notes", in Albert Einstein, Philosopher-Scientist, The Library of Living Philosophers, VII, ed. by P. A. Schilpp, Northwestern University and Southern Illinois University, 81.
4. BOHM, D. (1989): "Quantum Theory", **Dover chapter** 22, 611.
5. BOHR, N. (1935): "Can quantum-mechanical description of physical reality be considered complete?", **Phys. Rev.** 48, 696-702.
6. GREENBERGER; D.M.; M.A. HORNE, A. SHIMONY and A. ZEILINGER (1990): "Bell's theorem without inequalities", **Am. J. Phys.** 58, 1131-1143, note 10.
7. CABELLO, A. and G. GARCÍA ALCAINE (1995): "La sorprendente incompatibilidad de la idea de realidad einsteniana con la mecánica cuántica", **Rev. Española de Física** 9, 11-17.
8. PAGE, D.N. (1982): "The Einstein-Podolsky-Rosen physical reality is completely described by quantum mechanics", **Phys. Lett.** 91A, 57-60.

9. GHIRARDI, G.C.; A. RIMINI and T. WEBER (1980): "A general argument against superluminal transmission through the Quantum mechanical measurement process", **Let. Nuov. Cim.** 27, (293-298).
10. GARCÍA ALCAINE, G. and G. ÁLVAREZ GALINDO (1987): "Las mediciones cuánticas no violan la causalidad relativista", **Rev. Española de Física** 1, 29-35.
11. HARDY, L. (1998): "Spooky action at a distance in quantum mechanics", **Contemporary Physics** 39 (419-429).
12. BALLENTINE, L.E. (2001): "Quantum Mechanics", **World Scientific**, 595-598.
13. BELL, J.S. (1964): "On the Einstein-Podolsky-Rosen paradox", **Physics** I, 195-200.
14. CLAUSER, J.F., M.A. HORNE; A. SHIMONY and R.A. HOLK (1969): "Proposed experiment to test local hidden variable theories", **Phys. Rev. Lett.** 23, 880-884.
15. CLAUSER, J.F. and A. Shimony (1978): "Bells's theorem: experimental tests and implications", **Rep. Prog. Phys.** 41, 1881-1926.
16. ASPECT, A. (2002): "Bell's theorem: the naive view of an experimentalist", in Quantum [Un]speakables-From Bell to Quantum Information, ed. by R. A. Bertlmann and A. Zeilinger, Springer.
17. ZEILINGER, A. (1999): "Experiment and the foundations of quantum physics", **Rev. Mod. Phys.** 71, S288-S297.
18. WEINFURTER, H. (2005): "The power of entanglement", **Physics World** 18, 47-51.
19. GREENBERGER, D.M.; M. HORNE and A. ZEILINGER (1989): "Going beyond Bell's theorem", in Bell's Theorem, Quantum Theory, and Conception of the Universe, ed. by M. Kafatos, Kluwer Academic, Dordrecht, 73-76.
20. GREENBERGER, D.M.; M. HORNE; A. SHIMONY and A. ZEILINGER (1990): "Bell's theorem without inequalities", **Am. J. Phys.** 58, 1131-1143.
21. MERMIN, N.D. (1990): "What's wrong with these elements of reality?"; **Physics Today** (june), 9-11.
22. \_\_\_\_\_. (1990): "Quantum mysteries revisited", **Am. J. Phys.** 58, 731-734.
23. NORSEN, T. (2005): "Einstein's boxes", **Am. J. Phys.** 73, 164-176.
24. HARDY, L. (1992): "Quantum Mechanics, Local Realistic Theories, and Lorentz Invariant Realistic Theories", **Phys. Rev. Lett.** 68, 2981-2984.
25. FEYNMAN, R.P.; R.B. LEIGHTON and M. SANDS (1965): "The Feynman Lectures on Physics", Vol III: **Quantum Mechanics**, Addison-Wesley, Reading, Massachusetts, chapter 1, 4-9.
26. WHEELER, J.A. (1983): "Law without Law", in **Quantum Theory and Measurement**, ed. by J. A. Wheeler and W. H. Zurek, Princeton University Press, Princeton, New Jersey, 182-213.
27. BOHM, D. (1952): "A suggested interpretation of the quantum theory in terms of "hidden" variables, I and II", **Phys. Rev.** 85, 166-193.

# EINSTEIN Y LAS TEORÍAS DE UNIFICACIÓN DE CAMPOS

H. Pérez Rojas<sup>1</sup> and E. Rodríguez Querts<sup>2</sup>

Instituto de Cibernética, Matemática y Física, Ciudad de La Habana, Cuba

## RESUMEN

En el presente trabajo se describen las fuerzas fundamentales de la naturaleza y se hace una breve exposición de los principales aspectos del Modelo Standard de las interacciones débil, electromagnética y fuerte. Además, se esbozan algunas ideas acerca de la llamada gran unificación de las interacciones.

## ABSTRACT

In this work the natural fundamental forces are described; and a brief explanation about the Standard Model regarding weak, electromagnetic and strong interactions is done. Besides; some ideas about the so called unified theory are presented.

## I. INTRODUCCIÓN

A Albert Einstein le cabe el mérito de haber formulado un nuevo principio de relatividad, que estableció un límite a la velocidad de propagación de toda señal que transporte información. El principio de relatividad einsteniano, además, trajo como consecuencia la unificación del espacio y el tiempo en una nueva entidad, el espacio-tiempo, en el cual la unificación de electricidad y magnetismo descubierta por James Clerk Maxwell se hace matemáticamente más evidente.

En sus últimos años Einstein buscó, aunque sin éxito, una teoría unificadora de las interacciones electromagnética y gravitacional. Inspirados en sus ideas, Kaluza y Klein propusieron un modelo de espacio tiempo de cinco dimensiones que unificaba en principio gravitación y electromagnetismo. La quinta dimensión se enrollaba como en un tubo, anticipando una idea básica para la moderna teoría de cuerdas.

Los trabajos de Weinberg, Salam y Glashow lograron, parcialmente, un objetivo similar, al unificar en una teoría las interacciones electromagnética y débil. Pero la vía seguida era muy distinta. Einstein había seguido un camino basado en las teorías clásicas (no cuánticas) de la relatividad general y la electrodinámica. La teoría del campo electrodébil se situaba, por el contrario, dentro del terreno de las modernas teorías cuánticas de campos con ruptura espontánea de la simetría.

En la actualidad se considera que las leyes que rigen el Universo, desde su expansión hasta las interacciones entre las llamadas partículas elementales que componen la materia ordinaria, deben formar

parte de una teoría unificada. La teoría de cuerdas, por ejemplo, pretende unificar las interacciones básicas a escala ultramicroscópica, del orden de la longitud de Planck  $l = 1.6 \times 10^{-35}$  m.

## II. INTERACCIONES FUNDAMENTALES

Hasta el momento se conocen cuatro tipos de fuerzas o interacciones básicas en la naturaleza. La más débil es la fuerza de gravitación, la cual es de largo alcance, al igual que la electromagnética. Sin embargo, debido a que las cargas gravitacionales no son apantallables (la masa es siempre positiva), esta interacción domina a escala cósmica. La teoría moderna de la gravitación, necesaria para describir con exactitud tanto el movimiento de los planetas (incluso de los satélites geoestacionarios), como la evolución del Universo, es la relatividad general; se basa en la equivalencia entre un campo gravitatorio y un sistema de referencia no-inercial (acelerado).

Entre sus implicaciones básicas está la expansión del Universo: en un pasado remoto se produjo una "gran explosión", a partir de la cual se formaron estrellas y galaxias, y a partir de tal explosión (siguiendo un proceso inflacionario: en sus comienzos el radio tuvo un crecimiento exponencial  $R(t) \sim e^{\lambda t}$ ) las galaxias se alejan unas de otras con velocidad creciente con la distancia. Actualmente todo indica que esta expansión ocurre de manera acelerada.

De acuerdo con la teoría no relativista de Newton, la interacción gravitatoria entre dos partículas cuya masa sea del orden de las masas atómicas, digamos de masa igual a la del protón  $m_p$ , está caracterizada por una constante adimensional extraordinariamente pequeña

E-mail: <sup>1</sup>hugo@icmf.inf.cu

<sup>2</sup>elizabeth@icmf.inf.cu

$$\frac{G_N m_p^2}{\hbar c} \approx 10^{-40}, \quad (1)$$

donde  $G_N$  es la constante gravitatoria. Por ello la gravitación se puede despreciar al estudiar las interacciones de las partículas a escala microscópica (excepto a muy altas energías, del orden de la masa de Planck  $E = M_p c^2 = 1.2 \times 10^{19} \text{ GeV}$ ).

En el micromundo se desarrollan procesos a energías comparables con la energía en reposo de las partículas. Por lo tanto, las teorías que describen las interacciones a esa escala deben ser consistentes con la relatividad especial y la mecánica cuántica. Esto es posible precisamente en el formalismo de la teoría cuántica de campos, que hace uso de una función Lagrangiana (análogamente al caso de la mecánica), a partir de la cual se obtienen las ecuaciones del movimiento). Sin embargo, estas ecuaciones del movimiento no describen partículas aisladas, sino campos. Este formalismo, capaz de describir los procesos de aniquilación y creación de partículas, constituye el lenguaje de la física de altas energías.

A la gravitación le sigue en intensidad la llamada fuerza débil, agente de ciertas desintegraciones radioactivas en que aparecen electrones y neutrinos, tales como la desintegración beta.

$$n \rightarrow p + e^- + \bar{\nu}. \quad (2)$$

Una característica importante de las interacciones débiles es que son de muy corto alcance. La distancia característica es  $r \sim 10^{-15} - 10^{-16} \text{ cm}$  y la probabilidad de interacción de los neutrinos con otras partículas es muy baja: por ello se les llama interacciones débiles.

Hasta 1967 los fenómenos de las interacciones débiles se describían satisfactoriamente mediante un modelo fenomenológico debido a Enrico Fermi, que utilizaba la corriente débil  $J_\lambda(x)$ , a la cual contribuye una parte hadrónica (se llaman hadrones a las partículas que interactúan fuertemente, como los bariones y los mesones) y otra parte leptónica (los leptones no interactúan fuertemente, son leptones, por ejemplo, el electrón y su neutrino asociado). Matemáticamente la interacción débil se describía mediante una función Lagrangiana,

$$L_F = \frac{G_F}{\sqrt{2}} J_\lambda(x) J_\lambda^\dagger(x), \quad (3)$$

donde  $J_\lambda$  es un cuatrivector,  $J_\lambda^\dagger$  es su conjugado hermítico y  $G_F$  la constante de acoplamiento débil.

Luego y en analogía con la electrodinámica cuántica, que ya había sido formulada anteriormente, se propuso una Lagrangiana de interacción de la forma

$$L_w = g J_\lambda(x) W_\lambda(x), \quad (4)$$

donde  $W_\lambda(x)$ , sería un campo bosónico. Si los bosones intermediarios de las interacciones tienen masa  $m$  diferente de cero, las fuerzas resultantes son de corto alcance, el factor de acortamiento de la interacción es proporcional a  $e^{-mr}$ . Como el fotón tiene masa nula, la fuerza de Coulomb es de largo alcance. Las interacciones débiles son de corto alcance y por lo tanto, los mesones intermediarios  $W$  deben ser masivos. Además, éstos aparecen en estados cargados.

Después de la débil, la más intensa es la fuerza electromagnética, entre cuyas múltiples acciones está la de atraer protones y electrones entre sí para formar el átomo. La existencia de cargas eléctricas de signos contrarios conduce al apantallamiento de la fuerza electrostática: una carga eléctrica tiende siempre a neutralizarse atrayendo cargas opuestas. Esto ocurre en el átomo, en la molécula, en el sólido y en cualquier cuerpo microscópico.

Los trabajos de Dirac, así como los de Born, Heisenberg y Jordan establecieron las bases para el desarrollo de la *electrodinámica cuántica*, es decir, una teoría relativista que describe la interacción del campo electromagnético cuantificado con el campo electrón-positrón. La electrodinámica cuántica toma en cuenta la estructura corpuscular de la luz, que se considera compuesta de partículas de spin 1 (los fotones), los cuales interactúan con las partículas cargadas (electrones, protones, mesones pi, etc.) a través de los procesos de emisión y absorción. La función Lagrangiana de la electrodinámica cuántica consta de tres partes: una parte que describe a los electrones y positrones, otra parte que describe al campo electromagnético en el vacío, y una tercera que contiene la interacción del campo electrón-positrón  $\Psi$  con el campo electromagnético  $A_\mu$ . Este término se puede escribir.

$$L = e \bar{\Psi} \gamma^\mu A_\mu \Psi, \quad (5)$$

donde  $e$  es la carga eléctrica,  $\bar{\Psi}$  es el conjugado del spinor  $\Psi$ , y  $\gamma^\mu$  es el conjunto de las cuatro matrices de Dirac. Una consecuencia fundamental de la teoría relativista de Dirac es la existencia de antipartículas. La antipartícula del electrón es el positrón, de carga positiva. Esta es una propiedad universal, a toda partícula, le corresponde una antipartícula.

Por último, la fuerza más poderosa es la nuclear (interacción fuerte), que atrae a protones y neutrones entre sí para formar el núcleo atómico. Para tener una idea comparativa, baste decir que entre dos protones la atracción a causa de esta fuerza es de 100 a 1 000 veces mayor que la repulsión electrostática, y esta última es mayor que la gravitatoria en una

cantidad del orden de  $10^{37}$ . Pero la fuerza nuclear se caracteriza por ser de muy corto alcance: su acción está limitada a distancias del orden de las dimensiones del núcleo atómico, es decir de  $10^{-13}$  cm.

En 1935, el físico japonés Hideki Yukawa propuso la existencia de los mesones masivos como partículas intermediarias en estas interacciones fuertes, por analogía con los fotones, intermediarios en las interacciones electromagnéticas. Los mesones de Yukawa fueron descubiertos en 1947 y designados como mesones  $\pi$ . En la actualidad, las interacciones fuertes se suponen determinadas por la interacción entre los quarks, partículas de carga fraccionaria componentes de los hadrones. Los quarks interactúan a través del campo gluónico, que sería la fuerza fundamental, y la fuerza nuclear a la que hemos hecho referencia sería como una fuerza residual, cuya relación con el campo gluónico sería parecida a la de las fuerzas moleculares con respecto a las fuerzas interatómicas.

Así, los componentes básicos de la materia son partículas de spin  $1/2$  (fermiones), que interactúan entre sí mediante el intercambio de partículas de spin 1 (bosones), mediadoras de las interacciones.

### III. MODELO STANDARD

Según el Modelo Standard, los constituyentes básicos de la materia son partículas de spin  $1/2$  (fermiones), los leptones y los quarks, que se comportan como partículas sin estructura interna, es decir, como verdaderos puntos, sin dimensiones perceptibles. Los leptones no participan en las interacciones fuertes. Se tienen seis leptones:

Lepton	Nombre	Carga eléctrica
$\nu_e$	neutrino electrónico	0
e	electrón	-1
$\nu_\mu$	neutrino muónico	0
$\mu$	muón	-1
$\nu_\tau$	neutrino $\tau$	0
$\tau$	leptón $\tau$	-1

Obviamente, a cada leptón, le corresponde un anti-leptón. Los quarks son partículas de carga fraccionaria  $2/3$  y  $1/3$  de la carga del electrón, entre los cuales se ejerce la interacción fuerte a través del campo gluónico. Hay dos propiedades fundamentales de los quarks: el confinamiento y la libertad asintótica. El confinamiento de los quarks consiste en que estas partículas no se encuentran libres; los quarks siempre aparecen en parejas o en tríos, como constituyentes de otras partículas. Los bariones están constituidos por tres quarks y los mesones por pares quark-antiquark. La libertad asintótica se refiere a que los quarks, cuando están

muy próximos (o su energía de interacción es muy grande, a muy alta energía), se comportan como si fuesen libres, o como si no interactuaran entre sí.

Los quarks son de diferentes tipos, a los que los físicos llaman humorísticamente *sabores*. Hasta el presente se suponen seis sabores o tipos de quarks: "up", "down", "strange", "charmed", "bottom", "top". La sustancia ordinaria requiere sólo de los quarks u, d, (a los cuales se les asocia el número cuántico I (spin isotópico), relacionado con las transformaciones en el espacio isotópico, que veremos más adelante) y de los leptones  $e^-$ ,  $\nu_e$ . Por ejemplo, un protón tiene la estructura:

$$p = \begin{cases} \text{quarks carga} \\ u & 2/3 \\ u & 2/3 \\ d & -1/3 \end{cases}$$

cuya carga eléctrica total es 1. Un neutrón, por el contrario, tiene la estructura:

$$n = \begin{cases} \text{quarks carga} \\ u & 2/3 \\ d & -1/3 \\ d & -1/3 \end{cases}$$

de carga total cero. Análogamente, un mesón  $\pi = (u\bar{d})$ , está compuesto de un quark u, de carga  $2/3$ , y de un antiquark  $\bar{d}$ , de carga  $1/3$ . Su carga es +1, etc. El mecanismo de desintegración beta podemos entenderlo ahora como que un quark d en el neutrón emite una partícula  $W^-$  de carga -1, y entonces pasa a ser un quark u, con lo que el neutrón original pasa de la estructura neutra  $udd$  a la  $uud$ , de carga +1 que corresponde al protón.

Hay otro conjunto de partículas como el barión  $\Lambda$  y los mesones  $K^\pm$ ,  $K^0$ ,  $\bar{K}^0$  que manifiestan propiedades caracterizadas por la "extrañeza", número cuántico que se designa por la letra S. La extrañeza está determinada por los quarks s. La existencia del *charmed* quark c fue propuesta por Glashow, Iliopoulos y Maiani. Actualmente se considera que existen seis quarks.

Quark	Sabor	I	$I_3$	S	C	$B^*$	T	Q/e
u	up	$1/2$	$1/2$	0	0	0	0	$2/3$
d	down	$1/2$	$-1/2$	0	0	0	0	$-1/3$
s	strange	0	0	-1	0	0	0	$-1/3$
c	charmed	0	0	0	1	0	0	$2/3$
b	bottom	0	0	0	0	-1	0	$-1/3$
t	top	0	0	0	0	0	1	$2/3$

Para los números cuánticos asociadas a los diferentes "sabores" se cumple

$$\frac{Q}{e} = I_3 + \frac{1}{2}(B + S + C + B^* + T), \quad (6)$$

donde B es el número bariónico, el cual es igual a  $\frac{1}{3}$  para todos los quarks. Entre los quarks se ejerce una fuerza, caracterizada por un número cuántico nuevo, llamado *color*. Así, cada quark de un *sabor* dado, por ejemplo u, puede aparecer en los *colores rojo, azul y amarillo*. La teoría que describe la interacción entre los quarks se llama, por esta razón, *cromodinámica cuántica*. Una consecuencia fundamental es que las partículas observables deben ser incoloras pues no se observan quarks libres en la naturaleza.

La primera teoría cuántica que se formuló, la electrodinámica cuántica, posee una invarianza fundamental, llamada de calibración (o de norma). Si al campo electromagnético  $A_\mu$  se le suma el cuatrigradiente de una función de las coordenadas y del tiempo  $f$ , y al spinor que define al campo electrón-positrón  $\Psi$  se le multiplica por un factor de fase

$e^{i\frac{e}{\hbar c}f}$  los resultados físicos no varían. Es conveniente destacar aquí que la Lagrangiana de la electrodinámica cuántica (sin el término de interacción (5)) es invariante ante una transformación global, es decir, donde  $f$  es independiente de las coordenadas. Tal transformación global no afecta al campo electromagnético. Ahora bien, si  $f$  depende de las coordenadas (la transformación es local), para mantener la invarianza de las consecuencias físicas de la teoría, es necesario introducir la interacción del campo electromagnético con el campo electrón-positrón; es decir, la interacción electromagnética es consecuencia de la simetría de calibración asociada al grupo U(1).

En 1954 los físicos C.N. Yang y R. L. Mills propusieron una generalización de la invarianza de calibración de la electrodinámica. En esta, cada componente espacio-temporal de los campos tendría, a su vez, componentes en un espacio abstracto llamado espacio isotópico, es decir, una componente del campo según cada dirección del espacio isotópico (por ejemplo, el protón y el neutrón se pueden considerar dos componentes isotópicas de una misma partícula, una componente en la *dirección* de carga +1, y la otra en la *dirección* 0). El campo propuesto por Yang y Mills sería un campo compuesto de modo tal que en cada dirección del espacio isotópico habría un cuatrivector potencial, semejante al cuatrivector  $A_\mu$  del campo electromagnético.

Las transformaciones de calibración varían así en cada punto, y son por ello esencialmente *locales*.

A diferencia de la electrodinámica, las transformaciones de norma de los campos de Yang-Mills son no-abelianas (no conmutativas); un ente fundamental, llamado tensor del campo, tendría propiedades no lineales, pues contendría un término con el producto de dos de sus componentes. Esto, da lugar a nuevas propiedades físicas para estos campos. Por ejemplo, cuando un fotón se propaga libremente no puede descomponerse en dos fotones, pues este proceso no lineal está prohibido en la electrodinámica. Pero en los campos de Yang-Mills, a causa de sus propiedades no lineales, es posible el proceso en que un bosón masivo Z se desintegra en un par de bosones pesados  $W^+$  y  $W^-$ .

Entonces, mediante un modelo de Yang-Mills se puede unificar en una teoría a la electrodinámica y las interacciones débiles como resultado de la simetría de calibración  $SU(2) \times U(1)$ . Las interacciones entre los quarks, por otra parte, se deben a la invarianza ante el grupo SU(3). Pero para que la teoría así obtenida sea renormalizable (es decir, las divergencias que aparecen en los cálculos específicos se puedan eliminar) los campos de calibración deben ser no masivos, como el fotón. Hace falta entonces un mecanismo mediante el cual se genere la masa de los bosones intermediarios en las interacciones débiles y preserve la renormalizabilidad.

¿Qué ocurre si el campo de Yang-Mills no abeliano interactúa con un campo escalar? En este caso, si hay ruptura espontánea de la simetría (es decir, si el estado básico no es simétrico ante la transformación de calibración, aunque la Lagrangiana sí lo es) aparecen componentes del campo escalar de masa nula que corresponden a partículas no físicas, que se pueden eliminar de la teoría y sólo queda una componente con masa. Esto significa que hay grados físicos de libertad que se pierden; pero, en su lugar, el campo de calibración adquiere masa; es decir, la ruptura de simetría produce partículas sin masa y es posible demostrar que estas son no físicas. Pero, por otra parte, el campo de calibración adquiere masa. Por cada partícula escalar sin masa, a causa de la ruptura de simetría, aparece ahora una partícula vectorial con masa. Este es el llamado mecanismo de Higgs.

En el modelo de unificación de las interacciones electromagnéticas y débiles, debido a Weinberg, Salam y Glashow, se usa de modo especial el mecanismo de Higgs, para que tres campos de norma, originalmente sin masa, adquieran masa a causa de su interacción con un campo escalar. Este campo escalar no está aún identificado en la naturaleza, sin embargo, las consecuencias físicas de la unificación han tenido ya comprobación experimental amplia. Este campo escalar aparece inicialmente con cuatro componentes independientes. A causa de la ruptura de simetría, sólo una de las

componentes del campo escalar es masiva y se le llama escalar de Higgs. Las otras tres componentes de masa nula pierden sentido físico y, en su lugar, aparecen las componentes longitudinales que adquieren masa. Un elemento básico del modelo de Weinberg y Salam es el de suponer inicialmente un campo de norma abeliano  $B_\mu$  y uno de calibración no abeliano de tres componentes:  $W_\mu^1$ ,  $W_\mu^2$  y  $W_\mu^3$ . De la combinación de  $B_\mu$  con  $W_\mu^3$  resultan dos campos neutros: uno sin masa, que es el campo electromagnético, y otro que adquiere masa por el mecanismo de Higgs, llamado campo  $Z_\mu$ . Por otra parte,  $W_\mu^1$  y  $W_\mu^2$ ; se combinan para dar dos campos (o partículas) cargados  $W_\mu^\pm$ , también masivos. Las partículas  $Z_\mu$  interactúan con los neutrinos y también con los electrones. Ellas son los agentes de la llamada corriente neutra, descubierta en 1975 en el CERN (Centro Europeo de Investigaciones Nucleares) y considerada como una de las primeras confirmaciones de la teoría de Weinberg-Salam.

El campo de color aparece mediado, según el Modelo Standard, por partículas sin masa, llamadas gluones (de *glue*, en inglés, pegamento) que serían los análogos de los fotones, como verdaderos intermediarios de las interacciones fuertes. La existencia de tres colores de quarks da una base para la teoría actual de las interacciones fuertes, la cromodinámica cuántica (en inglés, QCD). El conjunto de transformaciones unitarias que transforman los colores entre sí son el grupo de *calibración de color* SU(3), que tiene ocho generadores que determinan el grupo. El campo gluónico se describe así por un campo de Yang-Mills donde las componentes isotópicas corresponden a los distintos colores  $a = 1, 2, \dots, 8$ . En la QCD la simetría es exacta: por ello los gluones serían como el fotón, partículas de masa cero.

La cromodinámica cuántica presenta la propiedad de confinamiento, ya mencionada. Un quark de color dado atrae cargas de color de la misma polaridad. Como resultado de esto la carga de color dado disminuye a cortas distancias del quark, y aumenta, cuando se incrementa la distancia. Esto está asociado a la dependencia de la constante de acoplamiento con respecto al momentum  $g^2 = g(-p^2)^2$ , (se suele usar esta notación donde  $-p^2 > 0$  es el módulo del vector cuatri-momentum). Tomando un valor  $\mu^2$  como referencia, tiene la forma:

$$g(-p^2) = \frac{g(\mu^2)}{1 + (11 - \frac{2n_f}{3}) \ln \frac{-p^2}{\mu^2}}, \quad (7)$$

donde el número de sabores de quarks es  $n_f = 6$ . Como ejemplo, si uno tomase un mesón  $\pi$  y tratara

de separarlo en un quark y un antiquark, como la fuerza ejercida aumenta con la distancia ( $g(-p^2)$  aumenta), el aumento de energía potencial daría lugar a la formación de un nuevo par quark-antiquark, y el resultado final serían dos mesones  $\pi$ , pues solamente tienen una energía finita los estados ligados incoloros, es decir, carentes de un color neto, los cuales se llaman singletes de color.

#### IV. LA GRAN UNIFICACIÓN

Como ya hemos visto, los quarks interactúan entre sí mediante el campo gluónico, pero también interactúan entre sí y con los leptones mediante el campo electrodébil. Entonces, parece natural buscar modelos que unifiquen las interacciones fuertes, con el campo electrodébil. Un modelo que hace algunos años se tomó como prometedor fue el grupo SU(5). Para que sea una teoría unificadora de fuerzas débiles y fuertes, debe haber mecanismos que transformen hadrones en leptones. El modelo requiere para ello de unos bosones X super-masivos de energías del orden de  $10^{14}$ - $10^{15}$  GeV. Estos serían intermediarios en la descomposición del protón, por ejemplo, en un positrón y un mesón  $\pi$ . De acuerdo con esto, el protón debe tener una vida media de unos  $10^{33}$  años, y el neutrino debe tener una pequeñísima masa, probablemente del orden de algunas decenas de eV. La idea de la gran unificación se apoya también en el hecho de que las constantes de acoplamiento varían con la energía (o el momentum) de las partículas en interacción. Así, la constante de estructura fina, que caracteriza las interacciones electromagnéticas, es constante para bajas energías; para energías mayores aumenta su valor. También vimos que el campo gluónico se caracteriza por una constante de acoplamiento que disminuye al decrecer la distancia (incrementar el momentum y la energía) entre las partículas (libertad asintótica), mientras que la interacción débil disminuye, pero más lentamente. La teoría de la gran unificación prevé la unificación de las interacciones electrodébiles con las fuertes a energías gigantescas, del orden de  $10^{15}$  GeV, a las cuales corresponde una longitud del orden de  $10^{-29}$  cm. Es decir, a estas energías, las constantes de acoplamiento (que varían con el momentum) coinciden en valor. La interacción fuerte (descrita por el grupo SU(3)) decrece, y en cuanto a las otras dos, descritas por los grupos SU(2) y U(1), la débil decrece más lentamente y la electromagnética incrementa su valor hasta coincidir con las otras dos.

Esta energía prevista, de unificación de los campos fuertes y electrodébil, es aun  $10^4$  veces menor que la energía de Planck, característica de los efectos cuánticos gravitacionales. Para energías inferiores a  $10^{15}$  GeV, se produce una ruptura de la simetría, y quedan dos campos separados: el fuerte o de color

y el electrodébil. A su vez, a energías del orden de 100 GeV el campo electrodébil se separa en dos componentes. El campo electromagnético permanece hasta energías más bajas, pero la fuerza débil se manifiesta sólo mediante procesos virtuales de creación de las partículas W y Z. Es decir, entre 0 y 100 GeV, las fuerzas que prevalecen son la fuerte y la electromagnética. El hecho de que las energías determinantes de los cambios de simetría sean 100 y  $10^{15}$  GeV, plantea el problema de que en este modelo el rango de energías intermedias es muy grande: la relación entre ambos es de  $10^{13}$  y se piensa que esto es un defecto de la teoría y que puede haber muchos fenómenos notables en el fango de  $10^2$  a  $10^{15}$  GeV. La teoría de la gran unificación tiene una aplicación inmediata a escala

cosmológica. Se considera que al comenzar la expansión del universo (y probablemente la *inflación*), la temperatura media era del orden de  $10^{32}$  K. A esa temperatura todas las interacciones fundamentales tenían igual jerarquía. Al descender a menos de  $10^{28}$  K se produjo la separación entre las interacciones fuertes y el campo electrodébil. Entre esa temperatura y  $10^{15}$  K se produjo la transición de fase, en que se rompió la simetría electrodébil. A la temperatura media actual del universo visible, las fuerzas de la naturaleza tienen las características que detallamos anteriormente. De la igualdad inicial entre todas las interacciones fundamentales, alcanzada a temperaturas gigantescas, hemos llegado, al enfriarse el universo, a una escala jerárquica de las fuerzas de la naturaleza.

### REFERENCIAS

1. COLEMAN, S. (1985): **Aspects of Symmetry**, Cambridge University Press, Cambridge,
2. KAKU, M. (1993): **Quantum Field Theory**, Oxford University Press, Oxford,
3. WEINBERG, S. (1997): **The First Three Minutes**, Bantam Books, New York,
4. LINDE, A. D. (1990): **Inflation and Quantum Cosmology**, Academic Press, San Diego.
5. PERKINS, D.A. (1987): **Introduction to High Energy Physics**, Addison Wesley, Menlo Park, California.
6. LAWRIE, I. D. (1990): **A Unified Grand Tour of Theoretical Physics**, Adam Hilger, Bristol.

# ACERCA DE LA ENSEÑANZA DEL MODELO ATÓMICO DE BOHR

E. Marín, Facultad de Física, Universidad de La Habana, Ciudad de La Habana, Cuba

## RESUMEN

En el presente trabajo se presenta un enfoque alternativo para explicar el Modelo de Bohr, en el cual la condición de cuantización del momento angular es obtenida como consecuencia de aplicar el Principio de Correspondencia. Se pretende describir también cómo se puede extender dicho modelo, teniendo en cuenta consideraciones relativistas, para obtener las correcciones para el radio de las órbitas circulares y la energía de los electrones en ellas predichas por el modelo de Sommerfeld, e introducir el concepto de estructura fina. Presentaremos y discutiremos además, a nivel elemental, algunos ejemplos de fenómenos que ilustran cómo la teoría antigua de Bohr puede hacer una contribución importante a su comprensión.

## ABSTRACT

In this paper we present an alternative approach for obtaining the mean results of the Bohr's model of the atom. The starting point in our analysis will be the Bohr's correspondence principle, and we will discuss how it predicts the angular momentum quantization rule. It is also our purpose here to show how relativistic corrections to the Bohr model of the atom can be easily obtained from our approach as well as the concept of fine structure. We will also provide and discuss at elementary level some examples of phenomena which can be used to illustrate that the old Bohr's theory can still make an important contribution to their understanding.

## I. INTRODUCCIÓN

Pese a que su elegancia, relativa sencillez matemática y predicciones (principalmente con respecto a las líneas espectrales y las energías de ionización de átomos (ionizados) de un electrón o hidrogenoideos) hacen del Modelo atómico de Bohr un tema persistente en la mayoría de los cursos de Física Moderna, es bien conocido que el mismo es superado por la Mecánica Cuántica cuando se analizan sistemas multielectrónicos, incluyendo moléculas y sólidos. Por otra parte, el acelerado crecimiento del conocimiento científico que ha tenido lugar en los últimos años ha hecho necesaria la reformulación de los programas curriculares en las carreras de Ciencias e Ingeniería, donde se ha hecho necesaria la introducción de nuevos contenidos con la correspondiente reducción en el tratamiento de numerosos tópicos hasta ahora incluidos de manera reiterada. Los cursos de Física Moderna no son una excepción, y en particular, los temas correspondientes a los modelos clásicos y semiclásicos del átomo son candidatos potenciales a su eliminación o reducción debido a que sus principales resultados son explicados por las nuevas teorías. En este trabajo pretendemos mostrar algunos argumentos a favor de la necesidad de mantener el modelo de Bohr del átomo de un electrón en los cursos futuros de

Física y exponer algunos enfoques alternativos del tema que pueden ser utilizados en su enseñanza.

## II. EL MODELO ATÓMICO DE BOHR

### Lugar en los cursos de Física Moderna

En la mayoría de los cursos de Física Moderna [1-7] (precedidos invariablemente de los de Mecánica, Electromagnetismo, Óptica y Termodinámica), el Modelo Atómico de Bohr (N. Bohr, 1913 [8]) es generalmente presentado a continuación del Modelo Planetario debido a Rutherford, en el que tiene como precedentes fundamentales, en general, el estudio de los fundamentos de la Teoría Especial de la Relatividad (Einstein, 1905), el de la hipótesis de Planck (1901) acerca de la Cuantización de la Energía del Campo Electromagnético para describir el Espectro de Radiación Térmica del Cuerpo Negro, la presentación de los Postulados de Louis De Broglie (1924) y de las Relaciones de Indeterminación de Heisenberg (1926), pasando por los experimentos clásicos que demuestran la necesidad de nuevas teorías, como el del Efecto Fotoeléctrico, el Efecto Compton, los de Difracción de fotones y electrones por sólidos (Bragg, Laue, Davisson-Germer,..) y algunos más recientes relacionados con la interferencia de fotones, electrones y átomos. Después de presentar el Modelo de Bohr, la mayoría de los textos modernos pasan a presentar los fundamentos de

<sup>1</sup>Actualmente en Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada del Instituto Politécnico Nacional, Legaria 694, Colonia Irrigación 11500, México D.F

E-mail: [emarin@fisica.uh.cu](mailto:emarin@fisica.uh.cu), [emarin63@yahoo.es](mailto:emarin63@yahoo.es)

la mecánica cuántica ondulatoria de Schrödinger y su aplicación a la resolución de problemas sencillos como el movimiento de partículas sometidas a diversos tipos de campos de fuerza (como los que describen los pozos y barreras de potencial) con el objetivo de utilizar a continuación el modelo mecano-cuántico al estudio del átomo de un electrón, de sistemas más complejos como los átomos multielectrónicos, las moléculas y los sólidos, y de otros más sencillos, estructuralmente, como los núcleos atómicos. Algunos autores finalizan sus cursos con temas relacionados con la cosmología del universo en expansión [2-4, 6].

### Enfoque más utilizado en su presentación

Aunque son bien conocidas para la mayoría de los lectores esbozaremos a continuación las principales ideas relacionadas con el Modelo de Bohr.

### Postulados de Bohr

En los libros de texto citados arriba (presentamos solamente una selección de los más utilizados en la actualidad en diferentes universidades) se parte de algunos Postulados básicos para deducir las expresiones que permiten calcular la energía y el radio de las órbitas de Bohr. Estos postulados son:

- i. Los electrones solamente pueden encontrarse en órbitas discretas de energía  $E_n$ .
- ii. El movimiento de los electrones en estas órbitas tiene lugar sin emisión de energía. Solo se emite o absorbe energía al pasar el electrón de una órbita a otra y esta se emite o absorbe en forma de radiación electromagnética, es decir, de un fotón de energía  $h\nu$  igual a la diferencia energética entre las órbitas  $E_n - E_{n'}$ .
- iii. El momento angular del electrón en la órbita  $n$ ésima es

$$|\vec{L}| = n \frac{h}{2\pi}, \quad (1)$$

donde  $h$  es la constante de Planck y  $n$  un número entero.

Veamos algunos antecedentes experimentales que justifican la necesidad de su imposición, y cómo son utilizados para explicarlos.

### Series espectrales

La necesidad que tuvo Bohr de imponer las dos primeras condiciones puede ser justificada sobre la base de la Hipótesis de Planck en aras de evitar el "colapso" del Átomo de Rutherford (El modelo

nuclear del átomo desarrollado por Rutherford, y que explicaba los experimentos de dispersión de partículas cargadas por núcleos pesados, conducía a un átomo inestable debido a la radiación de energía de forma continua por el electrón acelerado en la órbita) y para justificar la existencia de espectros de energía discretos, para los cuales Balmer había encontrado hacia 1885 que las longitudes de onda,  $\lambda$ , de las líneas espectrales pueden expresarse mediante la fórmula:

$$\bar{\nu} = \frac{1}{\lambda} = R_H \left( \frac{1}{2^2} - \frac{1}{n^2} \right), \quad (2)$$

llamada **Fórmula de Balmer**, donde  $n > 2$  es un número entero y  $R_H = 109677.5810 \text{ cm}^{-1}$  es la llamada Constante de Rydberg en honor a quien generalizó en 1889 la expresión anterior a todo el espectro escribiéndola de la forma

$$\bar{\nu} = \frac{1}{\lambda} = R_H \left( \frac{1}{n'^2} - \frac{1}{n^2} \right), \quad (3)$$

con  $n' < n$ , conocida como Fórmula de Rydberg para las Series Espectrales del Átomo de H, donde  $R_H = 109677.5810 \text{ cm}^{-1}$ . Para  $n' = 2$  conduce a la Serie de Balmer, que se localiza en la región visible. Si  $n' = 1$  la Serie se denomina de Lyman, y está en la región ultravioleta.  $n' = 3$  se corresponde a la de Paschen y  $n' = 4$  a la de Brackett (Estas últimas pertenecientes al infrarrojo).

Los espectros atómicos son diferentes para cada elemento, hecho que al ser descubierto dio origen a la rama de la Física conocida como espectroscopia atómica.

### Deducción de la energía del electrón orbital en el átomo de Hidrógeno

En el Modelo de Bohr, atendiendo al primer postulado, se parte de igualar las fuerzas de Coulomb y Centrípeta que actúan sobre el electrón de masa  $m_0$  en el átomo de Rutherford, suponiendo una órbita de radio  $r$  constante y en la cual la velocidad angular del electrón es  $\omega$ , es decir,

$$\frac{e^2}{4\pi\epsilon_0 r^2} = m_0 r \omega^2, \quad (4)$$

donde  $e$  es la carga del electrón y  $\epsilon$  la permitividad eléctrica del vacío. Por otra parte, la ley de conservación de la energía plantea que:

$$E = E_c + E_p = \frac{1}{2} m r^2 \omega^2 - \frac{e^2}{4\pi\epsilon_0 r^2} \quad (5)$$

Hasta aquí todo se corresponde con el modelo de Rutherford. Sustituyendo (4) en (5) se obtiene

$$E = -\frac{e^2}{2 \cdot 4\pi\epsilon_0 r} = -\frac{1}{2(4\pi\epsilon_0)^{2/3}} (\epsilon^4 m_0 \omega^2)^{1/3} \quad (6)$$

Utilizando el Postulado (iii), es decir, la expresión (1), una condición empírica introducida para poder demostrar la validez de la Fórmula de Balmer, se obtienen entonces las expresiones para el radio de las órbitas electrónicas y la energía del electrón. Para la primera magnitud se tiene:

$$r_n = -\frac{n^2 \hbar^2 4\pi\epsilon_0}{m_0 e^2} \quad (7)$$

donde  $\hbar = h/2\pi$ . Para  $n = 1$  se tiene  $r_1 = a_0 = 0.529\text{Å}$ , que es el radio de la primera órbita de Bohr, aquella que ocupa el electrón en su estado básico, y que es conocido como Radio de Bohr.

Por su parte, para la energía del electrón en la  $n$ -ésima órbita se obtiene

$$E_n = -\frac{1}{2(4\pi\epsilon_0)^{2/3}} \frac{(e^4 m_0 16\pi^2 R_\infty^2 c^2)^{1/3}}{n^2} \quad (8)$$

Para la órbita  $n = 1$ ,  $E_1 = -13.59\text{eV}$ , y se conoce como **Energía de ionización**, aquella que es necesario suministrarle al átomo para ionizarlo, es decir, extraer al único electrón (Figura 1).

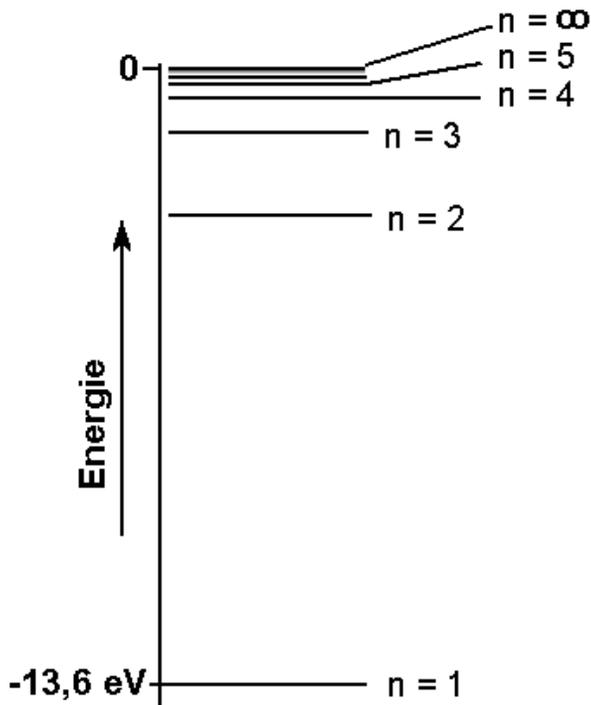


Figura 1. Niveles de energía del Hidrógeno según el modelo de Bohr.

Comparando esta ecuación con la fórmula de Rydberg (3), y utilizando el segundo Postulado (ii), se obtiene para la constante de Rydberg el valor

$$R_\infty = \frac{m_0 e^4}{8h^3 \epsilon_0^2 c} = 109737.318 \text{ cm}^{-1} \quad (9)$$

Este valor es ligeramente diferente del empírico para  $R_H$ , lo cual se debe a que en el razonamiento anterior se ha supuesto que el núcleo del átomo es mucho más pesado que el electrón orbital y que por ello está en reposo, es decir, no se ha tenido en cuenta el movimiento de ambas partículas alrededor del centro de masas del sistema que conduce a tener que sustituir  $m_0$  por la masa reducida del sistema. En la Mecánica se demuestra que el movimiento de dos partículas de masas  $m_0$  y  $M$ , separadas una distancia  $r$  tiene lugar alrededor de un centro de masas común. Si este centro está en reposo, la energía total de las dos partículas es igual a la de una partícula ficticia de masa reducida  $\mu$  que se mueve a la distancia  $r$  del centro de masas:

$$\mu = \frac{m_0 M}{m_0 + M} \quad (10)$$

donde  $M$  es la masa del núcleo, en este caso el protón. Luego, puede demostrarse que en todos los resultados anteriores hay que sustituir  $m_0$  por  $\mu$ , obteniéndose:

$$R_H = R_\infty = \frac{1}{1 + \frac{m_0}{M}} \quad (11)$$

Este resultado permitió diferenciar espectroscópicamente al H ordinario de su isótopo más pesado, el Deuterio (a ambos isótopos los diferencia la presencia de un neutrón en el núcleo de este último). Además permitió determinar la razón entre las masas del protón y del electrón,  $m_p/m_n \approx 1836.15$  toda vez que para el Hidrógeno  $M = m_p$ . Por supuesto que fue la demostración teórica de las fórmulas de Balmer y Rydberg.

Los resultados anteriores son extensibles a los llamados átomos hidrogenoideos, aquellos que solamente tienen un electrón, como los átomos ionizados de  $\text{He}^+$ ,  $\text{Li}^{++}$ , etc. Puede verse fácilmente que solamente es preciso hacer la sustitución

$$e^2 \rightarrow Ze^2 \quad (12)$$

en todas las expresiones anteriores.

### III. UN ENFOQUE ALTERNATIVO

Pese a los éxitos del modelo, el tercer postulado, que expresa la Cuantización del Momento Angular, aunque puede interpretarse sobre la base del postulado de De Broglie (Louis De Broglie propuso hacia 1924 la relación  $\lambda = h/mv$  para la longitud de onda asociada a una partícula de masa  $m$  moviéndose con velocidad  $v$ . Si la onda asociada a la partícula tiene que “entrar” en una órbita circular de radio  $r$  debe cumplirse que la longitud de la misma coincida con un múltiplo entero de longitudes de onda, es decir,  $2\pi rn = nh/mv$ , de donde se obtiene la condición de cuantización del momento angular,  $L = mvr_n = nh/2\pi$ ), no tiene un fundamento histórico precedente en el momento en que Bohr desarrolló su modelo, lo que dificulta a veces su aceptación por parte del estudiante. Nótese, como planteamos antes, que el trabajo de Bohr fue desarrollado alrededor del 1913 mientras que De Broglie publicó sus ideas aproximadamente 10 años después. El mismo Bohr utilizó la cuantización del momento angular como condición **ad hoc** para reproducir la fórmula de Balmer en la primera parte de su famosa trilogía al respecto [8]. En los dos siguientes trabajos trató de buscar una justificación para la misma. Su tercera derivación, considerada la más profunda de todas [9], comienza estipulando la coincidencia asintótica, para grandes valores de  $n$ , entre la frecuencia clásica de revolución del electrón en la órbita y la frecuencia de la radiación emitida en el “salto” cuántico del electrón entre órbitas, lo que se conoce como su Principio de Correspondencia.

#### El principio de Correspondencia en el Modelo de Bohr

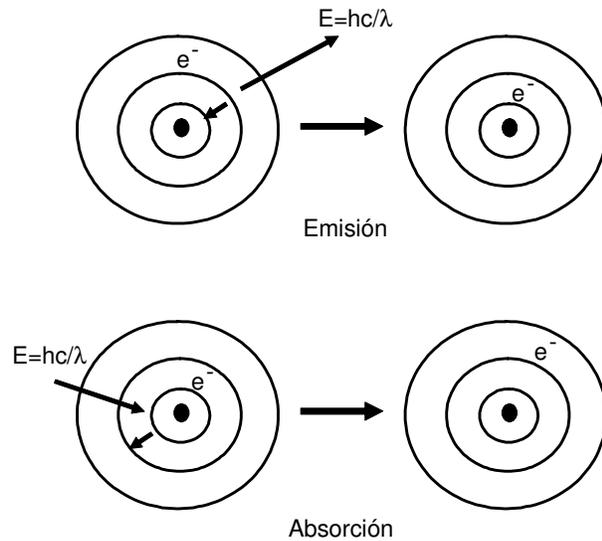
Este razonamiento es la base de un enfoque alternativo al presentado arriba para deducir las expresiones (6) y (7).

En esta formulación del problema se parte “solamente” de los dos primeros postulados, que escribiremos nuevamente:

- Los electrones solamente pueden encontrarse en órbitas discretas de energía  $E_n$ .
- El movimiento de los electrones en estas órbitas tiene lugar sin emisión de Energía. Solo se emite o absorbe energía al pasar el electrón de una órbita a otra  $n \rightarrow n'$ . La energía se emite o absorbe en forma de radiación electromagnética, es decir, de un fotón de energía  $h\nu$  igual a la diferencia energética entre las órbitas  $E_n - E_{n'}$ .

$$h\nu = E_n - E_{n'} \quad (13)$$

Si  $E_n > E_{n'}$  el fotón es emitido mientras que si  $E_n < E_{n'}$  el fotón es absorbido (Figura 2).



**Figura 2.** Procesos de emisión y absorción de la luz por un átomo.

Comparando esta expresión con la Fórmula de Rydberg (2) se obtiene:

$$E_n = - \frac{R_\infty hc}{n^2}, \quad E_{n'} = - \frac{R_\infty hc}{n'^2}, \quad (14)$$

Para calcular el valor de la constante  $R_\infty$  y compararla con  $R_H$ , se compara la frecuencia de revolución de los electrones en las órbitas con la de la radiación emitida o absorbida. En la física clásica estas frecuencias deben coincidir. La idea fundamental de Bohr fue que con el aumento de  $n$  las leyes de la física cuántica deben coincidir con las clásicas, que es, como hemos visto, la esencia de su Principio de Correspondencia.

Consideremos una transición electrónica entre órbitas vecinas ( $n - n' = 1$ ):

Para ella:

$$\begin{aligned} \nu &= R_\infty c \left( \frac{1}{n'^2} - \frac{1}{n^2} \right) = R_\infty c \left( \frac{1}{(n-1)^2} - \frac{1}{n^2} \right) = \\ &= \frac{R_\infty c}{n^2} \left( \frac{n^2}{(n-1)^2} - 1 \right) = \frac{R_\infty c}{n^2} \left( \frac{1}{(1-1/n)^2} - 1 \right) \quad (15) \end{aligned}$$

En el caso límite de  $n$  muy grande, o sea  $x = 1/n$  muy pequeño, haciendo un desarrollo en serie de Taylor alrededor de  $x = 0$  al término  $1/(1 - 1/n)^2 = 1/(1 - x)^2$  tenemos  $1/(1 - x)^2 = 1 + 2x + \dots$ , luego

$$v = \frac{R_{\infty}c}{h^2} \left[ 1 + \frac{2}{n} - 1 \right] = \frac{2R_{\infty}c}{n^3} \quad (16)$$

Sustituyendo  $\omega = 2\pi v = 4\pi R_{\infty}c/n^3$  en la expresión para la energía (5) se obtiene

$$E_n = -\frac{1}{2(4\pi\epsilon_0)^{2/3}} \frac{(e^4 m_0 16\pi^2 R_{\infty}^2 c^2)^{1/3}}{n^2}, \quad (17)$$

que coincide con la ecuación (8).

Igualando a (14) se llega a

$$R_{\infty} = \frac{m_0 e^4}{8h^3 \epsilon_0^2 c} = 109737.318 \text{ cm}^{-1} \quad (18)$$

Igualando (14) a (6) y sustituyendo (18) se obtiene el radio orbital (7), que junto con (16) conduce para el momento angular del electrón que se mueve con velocidad  $v_n$  y con frecuencia angular  $\omega_n$  sobre una órbita de radio  $r_n$ , a la Condición de Cuantización

$$|\vec{L}_n| = m_0 v_n r_n = m_0 r_n^2 \omega_n = n\hbar \quad (19)$$

La extensión a átomos hidrogenoideos se puede realizar de la misma manera expresada antes, es decir, haciendo los cambios expresados por (11) y (12).

Esta forma de presentar el problema constituye, en nuestro criterio, una variante útil para familiarizar a los estudiantes con las ideas fundamentales del principio de correspondencia, el cual es introducido en algunos cursos después de presentarse el modelo [1-7]. En la literatura consultada por el autor, un análisis similar al aquí presentado es realizado solamente por Haken y Wolf [10]. En las obras de Krane [2], Rohlf [4] y Eisberg & Resnick [1], estas ideas se discuten en ejemplos resueltos inmediatamente después de tratar el modelo, mientras que Halliday, Resnick & Walter [3], Beiser [5], Stierstadt [6] y Alonso & Finn [7] presentan el Principio de Correspondencia al discutir la aplicación del formalismo de la Mecánica Cuántica a casos sencillos como el movimiento de partículas en pozos de potencial infinitos. Una excelente discusión acerca de la historia de las primeras teorías de Bohr sobre la estructura atómica puede encontrarse en el artículo de Heilbron [10].

#### IV. APLICACIONES DEL MODELO DE BOHR

La mayoría de los cursos, aparte de ofrecer con el Modelo de Bohr una imagen intuitiva de cómo los electrones se mueven en el átomo, muestran varios ejemplos de aplicación y confirmación de la

Teoría. Entre otros podemos mencionar, sin seguir el orden lógico en que son o pueden ser enseñados:

- i. La Ley descubierta por Henry G. J. Moseley para describir los espectros característicos de Rayos X, es decir, el descubrimiento de que las líneas emitidas de mayor frecuencia satisfacen la ecuación  $v = (3R/4)(Z-1)^2$ , que puede ser derivada a partir del modelo de Bohr [1].
- ii. La descripción de los espectros de los metales alcalinos suponiendo que el movimiento del electrón óptico es "apantallado" por la carga del resto de los electrones e introduciendo el término llamado "defecto cuántico" o la "carga efectiva" en la expresión de la energía del electrón [10].
- iii. Los experimentos de J. Franck y G. Hertz, quienes en 1913, y aparentemente antes de los trabajos de Bohr, comprobaron la existencia de estados discretos de energía en el átomo en experimentos de colisión de electrones con átomos [1,7, 10].
- iv. La descripción de los niveles energéticos de los átomos muónicos [1, 10], descubiertos en la década del 50 del pasado siglo. El  $\mu^-$  es una partícula elemental de carga  $-e$  y masa 207 veces mayor que la del electrón. Este átomo se forma cuando un protón, u otro núcleo, capturan un muón antes de su posterior desintegración, tras haber sido producido, por ejemplo, mediante colisiones de neutrones con protones de alta energía (400MeV) (según esquemas como  $p+n \rightarrow p+p+\pi^-$  y  $\pi^- \rightarrow \mu^- + \tilde{\nu}_{\mu}$ ).
- v. Algunas evidencias experimentales relacionadas con la interacción de los átomos con campos externos [1-7, 15], aunque en estos casos hay que considerar además algunos resultados de la teoría mecano-cuántica del átomo.

#### Limitaciones del modelo

Cuando se discuten estas cuestiones, no deben dejar de mencionarse las limitaciones de la teoría estudiada. El modelo de Bohr es incompleto. Solamente es aplicable, como hemos visto, en el caso de átomos que contienen un solo electrón, y no en el de sistemas multieletrónicos, pues no considera la fuerza que ejercen ellos entre sí. Si observamos con detenimiento el espectro de emisión del Hidrógeno vemos que muchas líneas están formadas por varias componentes. El modelo no es capaz de explicar la existencia de las mismas. Tampoco dice nada respecto de las intensidades de las líneas espectrales, es decir, acerca de la probabilidad de que una transición electrónica predomine sobre otra. Quizás la más

sería deficiencia del modelo radica en la violación que hace de las relaciones de indeterminación al determinar exactamente el valor del momento lineal y la posición del electrón en el átomo. Se puede argumentar en defensa de Bohr que su teoría precedió en una década la introducción de la mecánica ondulatoria y de las ideas relacionadas con la indeterminación, aunque estas son introducidas en la mayoría de los cursos antes de abordarse el tema de los modelos atómicos. Sin embargo, esto no puede ser causa de que deje de tratarse el modelo completamente en un curso de Física Moderna. Las aplicaciones que discutiremos a continuación son ejemplos que permiten afirmar lo anterior.

### Aplicaciones:

#### i. Correcciones relativistas

Discutamos primeramente otro tópico interesante que no se aborda en los cursos tradicionales. Se trata de una manera relativamente simple de incluir correcciones relativistas en el modelo de Bohr [11] que conduce a resultados consistentes con los que se obtienen en el Modelo de Bohr-Sommerfeld para órbitas circulares. La brevedad de la deducción que presentaremos hace posible su inclusión en un curso estándar de Física Moderna a nivel universitario, teniendo en cuenta que el procedimiento seguido por Sommerfeld es largo y tedioso, por lo que muchos autores [1-8] simplemente se limitan a discutir sus resultados sin derivarlos. Hasta donde conocemos, solamente Alonso y Finn [7] parten de la expansión en serie de la relación energía-momento relativista para estimar el orden de magnitud de la corrección al valor de energía. Es nuestro propósito entonces, en lo que sigue, mostrar que es posible hacer correcciones relativistas en el modelo de Bohr sin tener que acudir al modelo de Sommerfeld o a la teoría cuántica relativista en un curso más avanzado de Mecánica Cuántica.

El punto de partida es la expresión relativista para la energía cinética que aparece en la expresión (5):

$$E_c = mc^2 - m_0c^2 \quad (20)$$

donde

$m = m_0[1 - \beta^2]^{-1/2}$  y  $\beta = v/c$ ,  $v$  es la velocidad del electrón en la órbita y  $c$  la velocidad de la luz en el vacío.

El procedimiento es similar que el seguido en el caso no relativista en cualquiera de las dos formulaciones mostradas arriba. Comencemos eliminando, primero  $r$ , y después  $v$ , de las

ecuaciones (4) y (20) respectivamente. De esta manera se obtiene

$$v = \frac{Ze^2}{n\hbar} = \alpha \frac{Zc}{n}, \quad (21)$$

donde

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx 1/137 \quad (22)$$

es la Constante de Estructura Fina.

y

$$r = \frac{\hbar n^2}{m_0\alpha Zc} \sqrt{1 - \beta^2} = \frac{n^2 a_0}{Z} \sqrt{1 - \left(\frac{\alpha Z}{n}\right)^2}, \quad (23)$$

con

$$r_1 = a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_0e^2} \quad (24)$$

es decir, el Radio de Bohr (Nótese que  $r = r_n(1 - \beta^2)$  no expresa otra cosa que la Contracción Relativista de Lorentz –Fitzgerald del “radio propio” de la órbita. Todo estudiante que haya recibido un curso básico de relatividad especial conoce la expresión para la contracción de la longitud propia de un objeto moviéndose a velocidades cercanas a  $c$ ).

En estas expresiones hemos introducido el número atómico  $Z$  para hacer extensible los resultados a otros átomos hidrogenoideos (estamos suponiendo que la masa nuclear es mucho mayor que la del electrón, aproximación que se hace más importante en la medida que  $Z$  aumenta, es decir, que los átomos de un electrón se hacen más pesados, pero introduce solamente pequeñas desviaciones en el caso de los más ligeros como el Hidrógeno. Tener en cuenta la masa reducida, lo que complica los cálculos, es innecesario para cumplir con el objetivo de nuestro análisis).

Sustituyendo las expresiones (20) a (24) en (5), y tras un procedimiento algebraico, obtenemos para la energía

$$E = m_0c^2 \left[ \sqrt{1 - \frac{\alpha^2 Z^2}{n^2}} - 1 \right] \quad (25)$$

El parámetro  $(\alpha Z/n)$  es mucho menor que 1, lo que permite hacer un desarrollo en serie de Taylor hasta el cuarto orden en la expresión anterior, lo cual conduce a:

$$E_n \approx -\frac{m_0 c^2}{2} \left(\frac{\alpha Z}{n}\right)^2 \left[ 1 + \frac{1}{4} \left(\frac{\alpha Z}{n}\right)^2 \right] \quad (26)$$

La teoría de Sommerfeld [12] conduce, por su parte, a

$$E_n \approx -\frac{m_0 c^2}{2} \left(\frac{\alpha Z}{n}\right)^2 \left[ 1 + \left(\frac{\alpha Z}{n}\right)^2 \left(\frac{n}{n_\phi} - \frac{3}{4}\right) \right] \quad (27)$$

con los números cuánticos radial y azimutal,  $n_r$  y  $n_\phi$  relacionados mediante el número cuántico principal,  $n$ , a través de  $n = n_r + n_\phi$ . En el caso de órbitas circulares  $n_r = 0$  y  $n = n_\phi$ , reduciéndose la Ecuación (27) a la (26). Similarmente puede demostrarse que la ecuación de Sommerfeld para la órbita elíptica

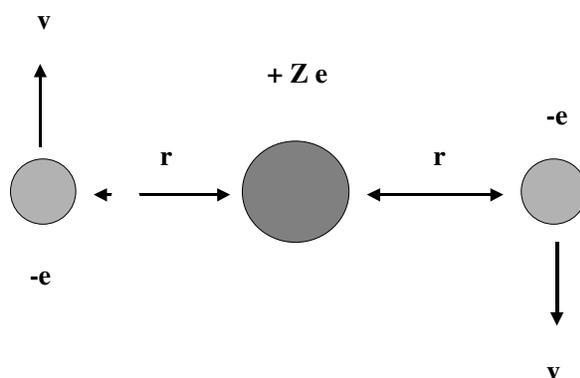
$$\frac{1}{r} = \frac{Z}{n_\phi^2 \Gamma^2 a_0} \left[ 1 + \frac{E}{m_0 c^2} \right] + A \cos(\Gamma \phi) \quad (28)$$

se reduce a la que encontramos para los radios permitidos de las órbitas circulares. En la ecuación anterior  $\phi$  representa el ángulo azimutal,  $\Gamma^2 = 1 - (\alpha Z/n_\phi)^2$  y  $A$  es una constante dependiente de la excentricidad de la órbita. En efecto, vemos que para órbitas circulares  $A = 0$  y  $n = n_\phi$ , que sustituidos en (28) conduce a (24).

Aunque los cursos de Física General incluyen tradicionalmente algunas nociones básicas sobre Teoría Especial de la Relatividad, muy poco se discute sobre su combinación con el átomo de un electrón. Las razones de esto son evidentes: El análisis cuántico relativista no encuentra lugar en este tipo de cursos que, de acuerdo a [13], *...deben constituir una primera aproximación al contenido y aplicación de la Mecánica Cuántica, destacando primordialmente sus razones, postulados y operatividad para explicar autoconsistentemente estructura y procesos en el microcosmos, pero sin erigirse en un curso abreviado de dicha ciencia. ...y ...Los temas deben ser presentados de forma tal (en la Física General), que el aparato matemático utilizado no esconda el análisis físico correspondiente, brindando la base conceptual y fenomenológica que requiere la Física Teórica para la posterior matematización rigurosa de las mencionadas teorías.* Por otra parte, el Modelo de Sommerfeld, aparte de las cuestiones mencionadas anteriormente, presenta solamente un valor histórico-metodológico, por cuanto sus resultados más importantes se derivan de manera más natural y elegante resolviendo la ecuación de Schrödinger para el átomo de un electrón.

## ii. Energía de ionización en átomos de dos electrones.

En un artículo reciente [14] se muestra cómo la teoría de Bohr puede ser aplicada en un modelo muy simplificado de un átomo de dos electrones para estimar sus energías de ionización con un error menor del 5% en la mayoría de los casos respecto del valor experimental (los estimados que se hacen en la mayoría de los cursos a partir del producto de dos funciones de onda para los estados 1s difieren en casi un 2%). En el modelo se considera que los dos electrones de cargas  $-e$  se mueven en órbitas circulares y se encuentran en lugares opuestos, a ambos lados del núcleo de carga  $+Ze$ , ambos sobre una línea que incluye a este último (Figura 3).



**Figura 3.** Dos electrones orbitan alrededor de un núcleo fijo.

Consideremos uno de los electrones. La fuerza neta de Coulomb que experimenta es debida a la suma de la fuerza de atracción hacia el núcleo y la repulsiva debida al otro electrón, que es fácil de demostrar que viene dada por:

$$F = -\frac{e^2}{4\pi\epsilon_0 r^2} \left( Z - \frac{1}{4} \right) \quad (29)$$

Esta fuerza debe ser igual a la fuerza centrípeta  $m_0 v^2/r$ , pudiéndose determinar la dependencia de  $v^2$  del radio orbital,  $r$ , la cual será de utilidad más adelante para calcular la energía cinética total del sistema  $E_c = mv^2$  (estamos considerando el núcleo en reposo).

La energía potencial de la configuración mostrada en la Figura 3 agrupa dos contribuciones debidas a la interacción electrón-núcleo y una debida a la interacción electrón-electrón, y se expresa como:

$$E_p = -\frac{e^2}{8\pi\epsilon_0 r} - (4Z + 1) \quad (30)$$

Utilizando ahora la ecuación que describe la cuantización del momento angular para el estado básico  $n = 1$  (estamos suponiendo que el átomo está en ese estado) puede demostrarse fácilmente que la Energía total del sistema,  $E = E_c + E_p$  viene dada por:

$$E = -\frac{m_0 e^4}{8\epsilon_0^2 h^2} \left(2Z - \frac{1}{2}\right) \left(Z - \frac{1}{4}\right) \quad (31)$$

Nótese que utilizando esta expresión se obtiene, en el caso del He ( $Z = 2$ ) el valor  $-83.3\text{eV}$ , mientras que el valor experimental es de  $-79\text{eV}$ , para una discrepancia de 5.4%. Para el Li ( $Z = 3$ ) ionizado la discrepancia se reduce a 3.9% ( $-205.8\text{ eV}$  calculado por  $-198.1\text{ eV}$  experimental), disminuyendo para átomos más pesados [14], aunque para ellos la necesidad de introducir efectos relativistas se hace mayor. Bohr y sus contemporáneos deben haber realizado cálculos mucho más engorrosos tratando de hacer predicciones similares, teniendo en cuenta distribuciones de carga menos uniformes, movimiento del núcleo, efectos relativistas, etc, siendo para ellos el cálculo mostrado arriba un ejemplo trivial, pero que puede resultar un ejercicio interesante a realizar por los estudiantes y un recurso útil para hacer simples estimados de la energía de ionización de átomos más complejos (el tratamiento del ión molecular de H que se hace en la mayoría de los libros de texto no está lejos del análisis anterior).

### iii. Átomos de Rydberg

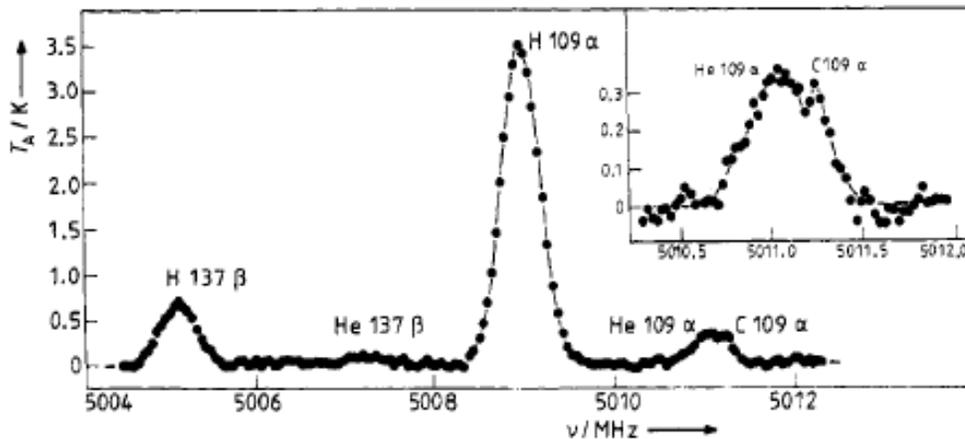
Hay una aplicación del modelo que puede resultar muy interesante, y que si bien es mencionado recientemente por algunos autores [10], no se describe en la literatura más utilizada [1-7]. Se trata de átomos en los cuales un electrón es excitado a un nivel energético muy elevado y que

constituyen un ejemplo de la continuidad existente entre la Mecánica cuántica y la Física clásica expresada en el Principio de Correspondencia.

Los denominados Átomos de Rydberg tienen propiedades asombrosas: son extremadamente grandes, pudiendo tener radios del orden de  $10^2\text{mm}$ , es decir, unas 100 mil veces el Radio de Bohr. Por otra parte, mientras el tiempo de vida de un electrón en un estado excitado de un átomo convencional es del orden de  $10^{-8}\text{ s}$ , los átomos de Rydberg pueden “vivir” algunos segundos. La causa radica en que para grandes valores de  $n$ , la separación energética entre dos estados consecutivos es muy pequeña, y según los resultados formulados primeramente por Einstein al analizar las características de la interacción de un campo de radiación electromagnética con un sistema atómico de dos niveles, la probabilidad de una transición espontánea es proporcional a la tercera potencia de la frecuencia de emisión [1]).

El electrón que se encuentra en el estado excitado se mueve prácticamente en el campo eléctrico del núcleo apantallado por el del resto de los electrones, es decir, en el campo debido a una carga  $e$ , exactamente igual a la del núcleo del átomo de Hidrógeno, de manera que se comporta como si perteneciera a un átomo de este elemento.

Semejantes átomos se han encontrado en nebulosas gaseosas, nubes de gases ionizados rarificados que rodean a algunas estrellas calientes, y cuyos espectros se han observado mediante radiotelescopios. La Figura 4 muestra el espectro medido en la Nebulosa de Orión A en el cual predomina la línea 109 $\alpha$  del Hidrógeno (transición  $n = 110 \rightarrow n = 109$ ,  $\Delta n = 1$ ). La emisión de radiación por otros átomos se observa también, por ejemplo en los picos correspondientes a las



**Figura 4.** Parte del espectro de radiofrecuencia de Orion A observado por Churchwell y Mezger [14, 15]. La temperatura de la antena,  $T_A$ , es proporcional a la intensidad de cada línea, de frecuencia  $\nu$ , relativa a la radiación de fondo. Transiciones donde  $\Delta n=1, 2, 3, \dots$  se clasifican con  $\alpha, \beta, \gamma, \dots$

transiciones He  $109\alpha$  y C  $109\alpha$ . La separación entre ellas se debe al efecto de la masa reducida sobre el valor de la constante de Rydberg [15]. De acuerdo con el Principio de Correspondencia, en la región de números cuánticos muy grandes, la frecuencia orbital del electrón viene determinada por la Ecuación (16) si se cumple la condición  $\Delta n = 1$ . Por lo tanto, existe una “regla de selección” que debe satisfacerse cuando  $n \rightarrow \infty$ . Esta es una buena oportunidad para introducir en el curso el concepto de regla de selección para las transiciones permitidas en los sistemas cuánticos (normalmente se comienza a mencionar este concepto al describirse la interacción de átomos con campos externos). Como  $n$  es grande, la radiación emitida aparece en la región de microondas.

En laboratorios terrestres se han “fabricado” Átomos de Rydberg con números cuánticos entre 10 y 290 [16]. Para ello se produce un haz de átomos cuyos electrones son excitados de un nivel a otro, en cascada, hasta llegar a los niveles energéticos mencionados. Esto se hace utilizando láseres sintonizables, sobre todo láseres de colorante. La detección se hace ionizando los átomos de Rydberg con ayuda del campo eléctrico de un condensador e identificando los mismos según su carga en un detector apropiado.

#### iv. Niveles energéticos en semiconductores extrínsecos

Muchos cursos de Física Moderna [1,2] pasan del estudio del átomo al de aglomeraciones de estos, como es el caso de moléculas y sólidos, discutiendo en muchos casos la estructura de bandas de estos últimos, y clasificándolos en metales, aislantes y semiconductores. Estos, a su vez, son clasificados en semiconductores intrínsecos y extrínsecos, al estudiarse cómo sus propiedades eléctricas pueden ser influenciadas y controladas por pequeñas cantidades de impurezas y dejando para otros cursos la explicación de cómo esto puede ser realizado en la práctica y aplicado en diferentes dispositivos.

Al semiconductor puro se le denomina intrínseco, ya que su comportamiento está determinado por sus propiedades intrínsecas, mientras que se clasifica como semiconductor extrínseco o dopado a aquel cuyas propiedades varían debido a interferencia externa, por ejemplo introduciendo en ellos átomos de impurezas. Hay dos tipos fundamentales de semiconductores intrínsecos. En un semiconductor tipo  $n$  (conductividad por electrones), los átomos de impureza tienen un electrón de valencia de más, comparado con el del semiconductor intrínseco al cual dopa. Un ejemplo es el Ge dopado con Arsénico ( $Z = 33$ ,

[Ge]4p<sup>3</sup>), un elemento del grupo V, cuyos átomos ocupan posiciones aleatorias en la red del Ge. El electrón que queda disponible después que se produce el enlace de un átomo de As con cuatro de Ge, está libre de moverse a través del cristal, determinando el tipo de conductividad. Dichos electrones ocupan niveles de energía ligeramente por debajo de la banda de conducción, pudiendo excitarse térmicamente en dicha banda sin dejar huecos en la de valencia. Como contribuyen con electrones, a estos átomos se les denomina donores.

El Modelo de Bohr permite estimar la energía de los niveles donores. Supongamos que el átomo de impureza está ligado al núcleo del átomo que lo cedió, y recordemos que la energía de un electrón en el estado básico del átomo de H es  $E = -m_0e^4/8\epsilon_0^2h^2$ . Esta debe ser la diferencia entre el nivel donador, ED y el fondo de la banda de conducción. Si sustituimos la constante dieléctrica por la del material y la masa del electrón por su masa efectiva (el electrón libre tiene una masa bien definida y cuando es acelerado en un campo eléctrico obedece las leyes de la mecánica newtoniana. Sin embargo, en el interior de la red cristalina, el electrón reacciona a la acción de un campo eléctrico con una masa efectiva que depende de la relación existente entre su energía y el número de onda en cada dirección cristalográfica, es decir, de la Relación de Dispersión) se obtiene una buena aproximación al valor experimental, que es 0.0127 eV para el ejemplo que nos ocupa [17]. A esta manera de encarar el problema se le denomina Aproximación de masa efectiva. Nótese que este valor es la mitad de la energía térmica a temperatura ambiente, lo cual quiere decir que a esa temperatura estos electrones son excitados a la banda de conducción.

De manera similar puede calcularse la energía de los llamados niveles aceptores en un semiconductor tipo  $p$ , en el cual los átomos de impureza tienen un electrón de valencia menos que los del material al cual dopan, como es el caso del Ge dopado con Galio ( $Z = 31$ , [Zn]4p<sup>1</sup>). Durante el enlace quedan vacancias de electrones (huecos) en los átomos de Ge responsabilizados con la conducción, que ocupan niveles energéticos dentro del Gap del Ge cercanos a la banda de valencia llamados aceptores, porque hacia ellos pueden excitarse electrones desde esta banda creando huecos, pero sin colocar electrones en la banda de conducción.

#### v. Estados excitónicos en semiconductores

Cuando sobre un semiconductor se hace incidir luz de energía  $E$  mayor que el ancho de su banda prohibida,  $E_g$ , se producen pares electrón-hueco. Si los electrones y los huecos no interaccionaran,

entonces solamente serían absorbidos fotones de energía  $E$ . Sin embargo, la interacción coulombiana entre el electrón y el hueco puede reducir la energía del par y hacer que aparezcan líneas con energías menores que  $E_g$  en los espectros de absorción. Los estados ligados de los pares electrón-hueco se denominan excitones. Pueden imaginarse, pictóricamente, como un electrón y un hueco orbitando uno alrededor del otro, es decir, como un sistema similar al átomo de un electrón, capaz de moverse a través del material bajo la acción de algún campo de fuerzas. El cálculo de los estados energéticos de los excitones puede realizarse entonces de manera similar, en muchos aspectos, al de los estados de impurezas discutidos arriba, solo que hay que considerar también la masa efectiva de los huecos.

#### IV. CONCLUSIONES

Aunque el modelo de Bohr del átomo es bien conocido desde hace casi un siglo, es difícil concluir a partir de la lectura de muchos libros de texto que la teoría todavía es capaz de explicar fenómenos descubiertos recientemente. La teoría cuántica antigua, que combina la Mecánica Newtoniana y las hipótesis de cuantización introducidas por Bohr en su modelo, aunque con serias limitaciones, puede ser utilizada todavía

para describir muchos fenómenos físicos interesantes –correcciones relativistas y sus consecuencias, energía de niveles energéticos en diferentes sistemas-, aunque los detalles específicos de los mismos solamente sean explicados por una teoría más rigurosa, aunque menos intuitiva, como la Mecánica Cuántica. El presente trabajo pretende ser un aporte para sugerir a los Profesores de Física una manera de cómo explicar a los estudiantes, a un nivel elemental, de qué forma la Teoría de Bohr juega aún un papel muy importante en nuestro entendimiento de los fenómenos del micromundo. Por otra parte, nuestra intención con los ejemplos aquí descritos es argumentar a favor de la importancia de mantener el tratamiento del Modelo de Bóhr en los cursos de Física General.

#### AGRADECIMIENTOS

La escritura de este trabajo estuvo motivada por la Clase Metodológica que tuvo que impartir el autor como parte de los ejercicios de oposición en opción a la Categoría Principal de Profesor Auxiliar en la Facultad de Física de la Universidad de La Habana en 2004. El autor agradece a los miembros del Tribunal de Categorización y a los demás colegas presentes en su exposición por sus comentarios y observaciones.

#### REFERENCIAS

- [1] EISBERG, R. y R. RESNICK (1985): **Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles** (John Wiley and Sons: New York).
- [2] KRANE, K. (1983): **Modern Physics** (John Wiley and Sons: New York).
- [3] HALIDAY, D.; R. RESNICK y J. WALKER (2001): **Fundamentals of Physics: Extended** (John Wiley and Sons: New York).
- [4] ROHLF, J.W. (1994): **Modern Physics from  $\alpha$  to  $Z^0$**  (John Wiley and Sons: New York).
- [5] BEISER, A. (1975): **Conceptos de Física Moderna** (Editorial Científico Técnica: La Habana).
- [6] STIERSTADT, K. (1989): **Physik der Materie** (VCH:Weinheim) (en alemán).
- [7] ALONSO, M. y E.J. FINN (1968): **Fundamental University Physics, III, Quantum and Statistical Physics** (Addison-Wesley Publishing Company).
- [8] Nos referimos a la famosa trilogía escrita por N. Bohr y publicada en 1913 “On the Constitution of Atoms and Molecules”, Partes I, II y III, en Phil. Mag. 26(151), 1-25; 153, 476-501 y 155, 857-875, respectivamente.
- [9] HEILBRON, J.L. (1985): **Physics Today**, october, 8-36.
- [10] HAKEN, H. y H.C. WOLF (2000): **The Physics of Atoms and Quanta: Introduction to Experiments and Theory** (Springer Verlag: Berlin Heidelberg).
- [11] KRAFT, D.W. (1974): **Am. J. of Phys.** 42, 837.

- [12] Aparece bien discutida, por ejemplo, en los textos: F. K. Richtmyer, E. H. Kennard y T. Lauritsen **Introduction of Modern Physics** (ER: Habana) 1967 y H. Semat **Física Atómica y Nuclear** (Aguilar: Madrid) 1971.
- [13] **Plan de Estudio C: Disciplina Física General**. Carrera: Licenciatura en Física. Curso regular diurno. Centro Rector: Universidad de La Habana. Ministerio de Educación Superior de la República de Cuba.
- [14] CAMERON REED, B. (2005): **Phys. Educ.** 40, 117.
- [15] CHURCHWELL, E. and P G. MEZGER (1970): **Astrophys. Lett.** 5, 227.
- [16] LATIMER, C.J. (1983): **Phys. Educ.**, 18, 86.
- [17] SOLYMAR, L. and D. WALSH (1995): **Lectures on the Electrical Properties of Materials** (Oxford University Press: NY).

# ANÁLISIS ALL-ELECTRON DE LA ESTRUCTURA ELECTRÓNICA DEL ThO<sub>2</sub> Y UO EN LA APROXIMACIÓN TB-LMTO-ASA

E. Chigo Anota<sup>1</sup>, Facultad de Ingeniería Química, Benemérita Universidad Autónoma de Puebla, Puebla, México

J. F. Rivas Silva<sup>2</sup>, Instituto de Física "Luis Rivera Terrazas"- Benemérita Universidad Autónoma de Puebla Puebla, México

## RESUMEN

Reportamos la estructura electrónica de los sólidos cristalinos UO y ThO<sub>2</sub> usando la aproximación TB-LMTO-ASA (Tight-Binding-Linear Muffin Tin Orbitals-Atomic Sphere Approximation) dentro de la teoría DFT (Density Functional Theory). Sus geometrías óptimas fueron obtenidas usando los pseudopotenciales ultrasuaves de Vanderbilt. Se reportan las densidades de estados total y parcial, así como la estructura de bandas. El UO muestra un comportamiento metálico y magnético mientras que el ThO<sub>2</sub> muestra un carácter metálico y no-magnético.

**Palabras clave:** aproximación TB-LMTO-ASA, teoría DFT.

## ABSTRACT

We report the electronic structure of the crystalline solids UO and ThO<sub>2</sub> using the TB-LMTO-ASA (Tight-Binding-Linear Muffin-Tin Orbitals-Atomic Sphere Approximation) into the framework of the density functionals (DFT) theory. Their optimal geometry were obtained using the ultrasoft pseudopotentials. Their **band structures** and *total and partial densities of states* were obtained too. The UO shows a metallic and magnetic behavior while the ThO<sub>2</sub> shows a metallic character and no-magnetic.

**Key words:** TB-LMTO-ASA approximation, DFT theory.

## 1. INTRODUCCIÓN

Como es bien sabido los cálculos electrónicos *ab-initio* sobre sólidos que contienen lantánidos y actínidos han ido incrementándose debido al comportamiento de fermion pesado [1] presentado por dichos compuestos, por eso la importancia de estudiar la estructura electrónica de los sólidos cristalinos ThO<sub>2</sub> y UO.

Previamente fueron estudiados dichos sólidos [2] usando el método de cluster atómico, y la teoría de los pseudopotenciales ERC [3] a nivel Hartree-Fock y CIS. Como resultado de este estudio se obtienen que el UO es covalente y el ThO<sub>2</sub> es iónico siendo una limitante el uso sencillo del modelo.

Ahora bien, la estructura cristalina del ThO<sub>2</sub> y UO es cúbica centrada en las caras (FCC) y con grupo espacial 225 y con parámetro de red experimental de 5.5997 Å y 4.92 Å respectivamente.

El ThO<sub>2</sub> se reporta en la literatura como **aislante** con un gap de 5 eV entre la banda de valencia 2p del oxígeno y la banda de conducción 6d-7s del torio [4].

Aquí se trabaja bajo la teoría de las funcionales de la densidad (DFT) [5] usando para esto la aproximación TB-LMTO-ASA (Tight Binding-Linear Muffin-Tin Orbital-Atomic Sphere Approximation) [6] implementado en el código del mismo nombre TB-LMTO-ASA [7] que realiza cálculos de todos los electrones (*all-electron*) y semi-relativistas (al incluir los términos de corrección de masa-velocidad y Darwin), usando como funcional de intercambio-correlación las parametrizaciones a nivel LSDA de von Barth-Hedin [8] y Ceperley-Alder [9] y a nivel GGS (LSDA no local con corrección del gradiente) las parametrizaciones de Langreth-Mehl-Hu [10] y Perdew-Wang [11]. Se obtiene la geometría óptima de estos compuestos usando teoría de los pseudo-potenciales vía el código CASTEP [12], así mismo, calculamos las densidades de estado total y parcial y la estructura de bandas.

## 2. METODOLOGÍA

La aproximación LMTO-ASA [6] divide a la celda unitaria del cristal en esferas de *Wigner-Seitz* (WS) traslapadas (con un traslape máximo de 20% **entre una esfera atómica y una intersticial y de 15% entre esferas atómicas**). Dentro de las esferas se

E-mail: <sup>1</sup>echigoa@sirio.ifuap.buap.mx

<sup>2</sup>rivas@sirio.ifuap.buap.mx

considera un potencial de simetría esférica, la cual representa la condición ASA (Atomic Sphere Approximation). Esto es aproximación sin zona de electrones libres en la estructura tipo muffin-tin. Para estructuras cristalinas de capa abierta, algunas esferas vacías son agregadas para cubrir el resto del espacio, detalle que en particular permite tomar en cuenta algo de la distribución electrónica en esa zona. El volumen total de las esferas WS iguala al volumen de la celda unitaria y así se elimina la región intersticial.

Por otro lado, los orbitales muffin-tin (MT) dependen en la energía, teniendo la siguiente forma linealizada

$$\Phi_{\Lambda}(\vec{r}) = i^l Y_L(r) \begin{cases} \phi_{\Lambda}(E, r) + p_{\Lambda}(r/S_R)^l & r < S_R \\ (S/r)^{l+1} & r > S_R \end{cases} \quad (1)$$

donde  $\phi_{\Lambda}(E, r)$  son halladas por solución numérica de la ecuación radial de Schrödinger (o cuasi-Dirac),  $\Lambda = \{RL\}$ , R representa el índice del sitio y la variable de momento angular  $L = \{lm\}$  es una combinación del momento angular l y su proyección m. Aquí,  $Y_L$  son los armónicos esféricos y  $S_R$  indica el radio MT. Los orbitales numéricos  $\phi_{\Lambda}(E, r)$  son aumentados dentro de la esfera por medio de una función de Bessel esférica renormalizada

$$J_{\kappa\Lambda}(r) = i^l Y_L \frac{(2l+1)!!}{(\kappa S_R)^l} j_l(\kappa r). \quad (2)$$

Fuera de la esfera, lo que se agrega es una función de Hankel esférica renormalizada:

$$H_{\kappa\Lambda}(r) = i^l Y_L \frac{(\kappa S_R)^{l+1}}{(2l-1)!!} h_l(\kappa r). \quad (3)$$

Aquí  $h_l = j_l - in_l$  representan una combinación lineal de funciones de Bessel y Neumann esféricas. **Las** colas de los orbitales MT en  $r > S_R$  son soluciones de la ecuación de Helmholtz con energía cinética igual a cero. Los parámetros de potencial  $p_{\Lambda}$  son elegidos de tal manera que hacen que la función de onda se haga continua y diferenciable en la frontera de la esfera.

Las características más atractivas del formalismo LMTO [6] son: (i) **que** emplea el mismo tipo de funciones base para todos los elementos, (ii) el conjunto base LMTO es mínimo, con únicamente 9 y 16 orbitales para elementos *spd* y *spdf* respectivamente.

Dicha metodología resuelve realmente las ecuaciones de Bloch para el sólido cristalino a través del esquema muffin-tin, tomando en cuenta el campo del cristal sobre cada uno de los iones de la celda unitaria, apareciendo un término correspondiente de energía en el Hamiltoniano del sistema.

Otra situación sobresaliente que se observa es que la solución numérica de la ecuación radial puede ser obtenida tomando en cuenta el Hamiltoniano no-relativista o una aproximación cuasi-relativista. Ambas opciones son posibles en el código TB-LMTO-ASA en su versión 47.

Las funcionales de intercambio-correlación  $V_{xc}$  utilizadas por dicho código son la de von Barth-Hedin (vBH) y Ceperly-Alder (CA) a nivel LSDA y las funcionales de Langreth-Mehl-Hu (LMH) y Perdew-Wang (PW) a nivel GGS.

### 3. RESULTADOS Y DISCUSIÓN

#### Optimización de la geometría vía CASTEP

La optimización de los sistemas (óxido de uranio y óxido de torio) fueron realizados **por** métodos de primeros principios, basados en la teoría de las funcionales de la densidad, en la aproximación de gradiente (GGS). En este método, la interacción core-electrón es representada usando los pseudo-potenciales ultrasuaves de Vanderbilt [13] en la representación de Kleinman-Bylander [14]. La parametrización de espín polarizado utilizada fue la de Perdew-Zunger [15] y la Perdew-Burke-**Ernzerhof** [16] para el potencial de intercambio-correlación. Los orbitales de los electrones de valencia son **expandidas** en ondas planas (mediante una combinación lineal de orbitales con cierto corte de energía y pseudopotenciales). El método de gradiente generalizado es empleado para relajar a los electrones a su mínima configuración energética fijando los átomos en las posiciones nucleares. El muestreo de la zona de Brillouin fue de  $4 \times 8 \times 8$  y  $5 \times 10 \times 10$  para el UO y ThO<sub>2</sub>, respectivamente, usando el esquema de Monkhorst-Pack [17]. El corte de energía para el conjunto de ondas planas fue de 400 eV aproximadamente. La autoconsistencia de los cálculos fueron realizados en función de que el cambio en energía total sea menor a 5 meV, garantizando así una buena convergencia [18], dado al tamaño finito de la función de onda.

Por otro lado, la configuración electrónica usada para los cálculos electrónicos en el programa TB-LMTO-ASA para el Th y U fue de [Hg] 6p7s5f6d y para el O de [Ne] 3s3p4s3d respectivamente, usándose 16 (10+6) y 20 (14+6) electrones de valencia respectivamente.

El volumen de la celda (usando el parámetro de celda optimizado) para el UO y ThO<sub>2</sub> fue de 206.451 y 292.959 u.a. respectivamente. Para esto se trabajó el espacio K con 29 puntos k en la zona irreducible de un total de 512 (de una malla de  $8 \times 8 \times 8$ ), permitiendo con esto un traslape entre las esferas atómicas de Wigner-Seitz de 14 %.

## Cálculos electrónicos en el UO

De acuerdo con los cálculos *ab-initio* realizados sobre los sólidos cristalinos UO y ThO<sub>2</sub> usando los parámetros de celda optimizados (ver Tabla I), se obtienen por un lado una buena concordancia con respecto a los valores experimentales.

**Tabla I.** Parámetros de celda.

Cristal/ Potencial Vxc	Parámetro de celda (Å)	Error %
<b>UO</b>		
UPP-Vanderbilt	4.9647 <sup>a</sup>	0.909
	4.88 <sup>b</sup>	0.813
	4.961 <sup>c</sup>	0.826
Experimental	4.92*	
<b>ThO<sub>2</sub></b>		
UPP-Vanderbilt	5.5790 <sup>a</sup>	0.369
Experimental	5.5997*	

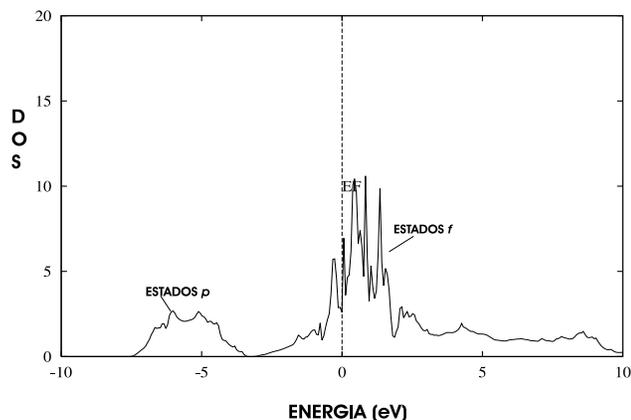
<sup>a</sup>Este Trabajo

<sup>b</sup>M. S. S. Brooks, *J. Phys. F: Met. Phys.* 14, 639 (1984).

<sup>c</sup>C. J. Pickard *et al. Phys. Rev. Letter* 85, 5122 (2000).

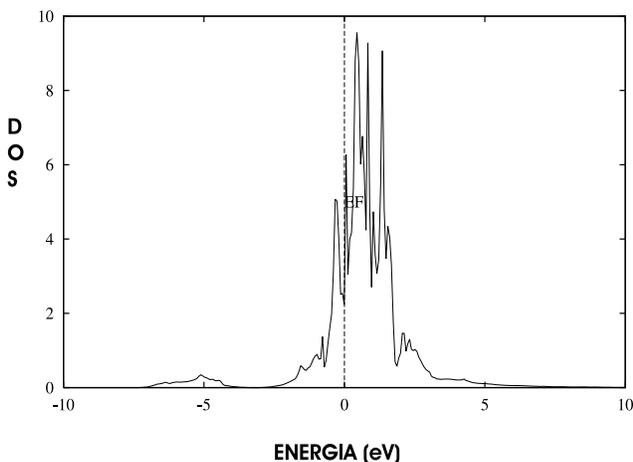
\*Wyckoff, R. W. G. *Crystal Structures* Vol. 1, John Wiley & Son, (1977).

Ahora bien, las gráficas de las densidades de estado total (Figura 1) y parcial (Figura 2) nos indican un comportamiento metálico por poseer bandas de energía centradas alrededor del nivel de Fermi con una gran contribución de los estados f del uranio, siendo los estados p que contribuyen a la banda de valencia principalmente (Figura 1). Por otro



**Figura 1.** Densidad de estados total para el UO.

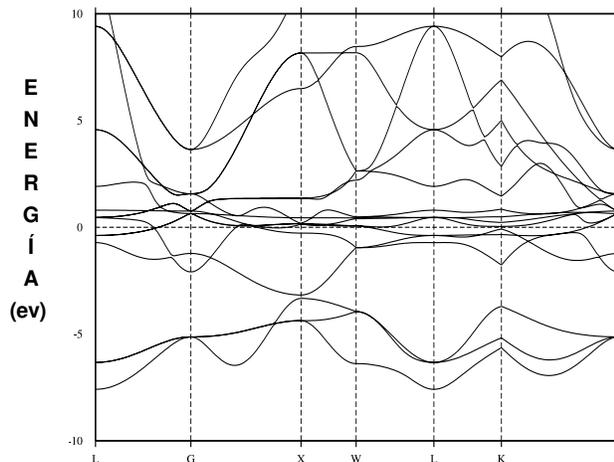
lado, la gráfica de la estructura de bandas (Figura 3) muestra una densa cantidad de bandas de energías alrededor del nivel de Fermi, apoyándonos en que



**Figura 2.** Densidad de estados parcial (estado f del uranio) para el UO.

también presentan bandas centradas sobre el nivel de Fermi con lo que podría indicarnos un posible comportamiento de fermion pesado [1,19]. El mismo sistema nos muestra un comportamiento ferromagnético cuyo momento magnético total de  $-1.8142 \mu_B$  (a nivel LSDA-vBH) variando muy leve en las otras parametrizaciones y donde la mayor contribución al material es por parte del uranio.

Además se observa gaps internos en la banda de valencia del orden de 0.5 eV indicado también en la gráfica de la estructura de bandas.



**Figura 3.** Estructura de bandas del UO a nivel LSDA-vBH.

Y según los valores de la carga en las esferas vacías (Tablas II y III) utilizadas para completar la condición ASA nos indican una pequeña covalencia entre el enlace molecular U-O en comparación por ejemplo de otro compuesto de uranio como el USb cuya carga de covalencia es de  $-0.621$  u. a. [20].

**Tabla II.** Cargas y momentos magnéticos a nivel LSDA.

ÁTOMOS	POSICIONES CRISTALOGRÁFICAS	EMT (u.a.)*	Mom. Mag. <sup>a</sup> ( $\mu_B$ )	Carga <sup>a</sup> (u.a.)	Mom. mag. <sup>b</sup> ( $\mu_B$ )	Carga <sup>b</sup> (u.a.)
U	(0.0, 0.0, 0.0)	3.2866	- 1.8496	0.86477	- 1.87833	0.86464
O	(0.5, 0.5, 0.5)	2.0686	0.0568	- 0.43499	0.061138	- 0.43491
*E	(0.25, 0.25, 0.25)	1.3512	- 0.01067	- 0.21489	- 0.011043	- 0.21487
*E	(-0.25, -0.25, -0.25)	1.3512	- 0.01074	- 0.21489	- 0.011043	- 0.21487
Th	(0.0, 0.0, 0.0)	3.226	0.000000	1.6047	0.000000	1.60505
O	(0.25, 0.25, 0.25)	1.9646	0.000013	- 0.3593	0.000000	- 0.35856
O	(-0.25, -0.25, -0.25)	1.9646	0.000013	- 0.3593	0.000000	- 0.35856
*E	(0.5, 0.5, 0.5)	2.7676	- 0.000010	-0.8861	0.000000	- 0.88793

\*Radio de la esfera muffin-tin.

<sup>a</sup> A nivel LSDA-vBH.

<sup>b</sup> A nivel LSDA-CA.

**Tabla III.** Cargas y momentos magnéticos a nivel GGS.

ÁTOMOS	POSICIONES CRISTALOGRÁFICAS	EMT (u.a.)*	Mom. mag. <sup>a</sup> ( $\mu_B$ )	Carga <sup>a,c</sup> (u.a.)	Mom. mag. <sup>b</sup> ( $\mu_B$ )	Carga <sup>b</sup> (u.a.)
U	(0.0, 0.0, 0.0)	3.2866	- 1.912486	0.88169	-1.942861	0.91850
O	(0.5, 0.5, 0.5)	2.0686	0.066432	- 0.44464	0.066804	- 0.48178
*E	(0.25, 0.25, 0.25)	1.3512	- 0.011860	- 0.21852	- 0.011196	- 0.21836
*E	(-0.25, -0.25, -0.25)	1.3512	- 0.011860	- 0.21852	- 0.011196	- 0.21836
Th	(0.0, 0.0, 0.0)	3.2260	0.000002	1.63500	0.000002	1.68219
O	(0.25, 0.25, 0.25)	1.9646	- 0.000001	- 0.37750	- 0.000001	- 0.42732
O	(-0.25, -0.25, -0.25)	1.9646	- 0.000001	- 0.8800	- 0.000001	- 0.82755
*E	(0.5, 0.5, 0.5)	2.7676	- 0.000001	- 0.8800	- 0.000001	- 0.82755

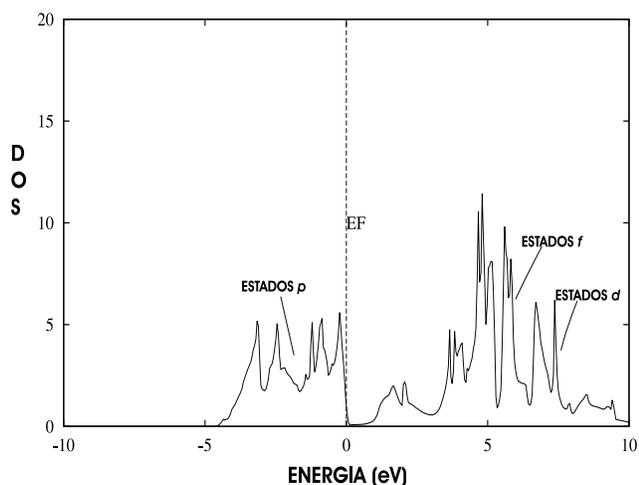
\*Radio de la esfera muffin-tin.

<sup>a</sup> A nivel GGS-LMH.

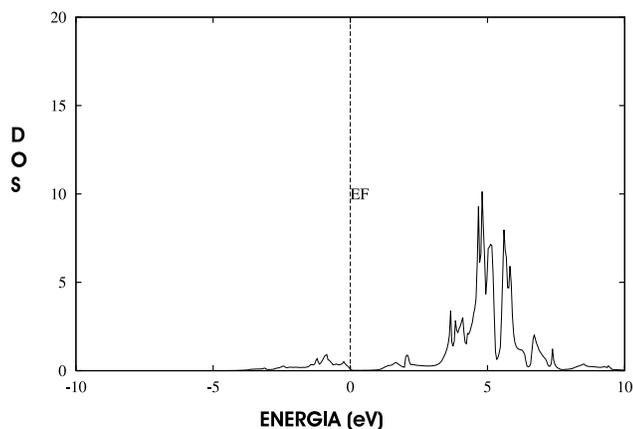
<sup>b</sup> A nivel GGS-PW.

### Cálculos electrónicos en el ThO<sub>2</sub>

Para el caso del ThO<sub>2</sub>, las gráficas de la densidad de estados total (Figura 4) y parcial (Figura 5) nos indican un pequeño gap de energía apenas por encima del nivel de Fermi de 0.5 eV aproximadamente correspondientes a una hibridización de estados spdf tanto del torio como del oxígeno esto nos indicaría un comportamiento más de aislante que metálico. En la misma gráfica se observa un gap interno en la banda de valencia del orden de 8 eV correspondiente a los estados p del oxígeno. De otra manera, tenemos estados 4f del torio abajo del nivel de Fermi y por encima a estados 5f del mismo torio. El pequeño gap alrededor del nivel de Fermi es resultado de la sobreestimación que hace la teoría DFT convencional a nivel LSDA.

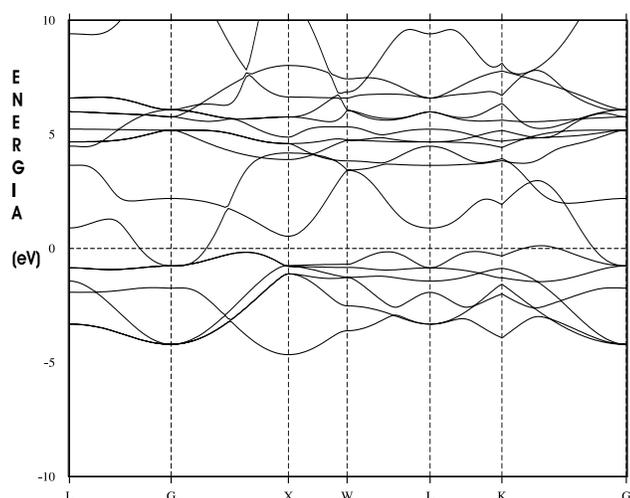


**Figura 4.** Densidad de estados total para el ThO<sub>2</sub>.



**Figura 5.** Densidad de estados parcial (estado f del torio) para el  $\text{ThO}_2$ .

Por su parte la estructura de bandas (Figura 6) nos muestra solo una banda alrededor del nivel de Fermi (observado en el punto G), con lo que señala una posible conducción. Como comparación con el UO no se tiene bandas horizontales alrededor de Fermi.



**Figura 6.** Estructura de bandas del  $\text{ThO}_2$  a nivel LSDA-vBH.

De acuerdo con los valores de la carga de las esferas vacías se observa una gran carga de covalencia y casi nulo momento magnético total del material ( $0.00001581 \mu_B$  a nivel LSDA-vBH). Este carácter covalente puede ser comparado con el ThS [21] cuyo valor es aproximadamente a mitad del  $\text{ThO}_2$ .

#### 4. CONCLUSIONES

Los cálculos obtenidos usando los parámetros de celda optimizados de acuerdo a la metodología de los pseudopotenciales, muestran un excelente acuerdo con el valor experimental y los valores teóricos reportados en la literatura.

Por otro lado, los cálculos electrónicos muestran que el  $\text{ThO}_2$  presenta un carácter más covalente que el UO, pero este último es el que presenta un carácter magnético. Además, el  $\text{ThO}_2$  tiene un comportamiento de aislante, siendo el UO metálico y una posible característica de fermion pesado.

Una descripción más adecuada de estos sistemas podría tenerse por ejemplo usando la aproximación LDA + U [22] la cual trabaja con un Hamiltoniano modelo que permite corregir los efectos de la autointeracción vía parámetros de correlación e intercambio, solucionando el problema central de la teoría DFT convencional [23] para el tratamiento de los efectos de correlación en sistemas que contienen elementos de lantánidos o actínidos.

#### AGRADECIMIENTOS

E. Chigo Anota agradece al Centro de Cómputo del Instituto de Física "Luis Rivera Terrazas" de la Benemérita Universidad Autónoma de Puebla las facilidades otorgadas para la realización de dicho trabajo.

Trabajo apoyado por la Vicerrectoría de Investigación y Estudios de Postgrado de la misma Universidad (Proyecto No. 28/G/EXC/05).

#### REFERENCIAS

- [1] STEWART, G. R. (1984): **Rev. Mod. Phys.** 56, 755;  
Peter Fulde, **J. Phys. F: Met. Phys.** 18, 601 (1988);  
Peter Fulde, **Ann. Phys.** 9, 871 (2000).
- [2] CHIGO ANOTA, E. y J. F. RIVAS SILVA (2005): **Rev. Col. Fís.** 37(1), 17.
- [3] FERNÁNDEZ PACIOS, L. and P. CHRISTIANSEN (1985): **J. Chem. Phys.** 82(6), 2664.  
E. Chigo Anota, Uso de los Pseudopotenciales de Ermler, Ross y Christiansen en algunos compuestos de actínidos. Tesis de Maestría. Benemérita Universidad Autónoma de Puebla, Puebla, México (2000); [http://es.geocities.com/echigoa/cee\\_solidos.html](http://es.geocities.com/echigoa/cee_solidos.html)
- [4] FREEMAN, A.J. and G. H. LANDER (1985): **Handbook on the Physics and Chemistry on the Actinides** 2, North-Holland.

- [5] HOHENBERG, P. y W. KOHN (1964): **Phys. Rev. B** 136, 864;  
W. Kohn y L. J. Sham, **Phys. Rev. A** 140, 1133 (1965);  
W. Kohn, A. D. Becke y R. G. Parr, **J. Phys. Chem.** 100, 12974 (1996);  
R. O. Jones y O. Gunnarsson, **Rev. Mod. Phys.** 61(3), 689 (1989);  
W. Kohn, **Rev. Modern Phys.** 71(5), 1253 (1999);  
Kieron Burke and friends (2003): **The ABC of DFT**, <http://dft.rutgers.edu/kieron/dft/beta>
- [6] SKRIVER, HANS L. (1984): **The LMTO Method**, Springer-Verlag;  
O. K. Andersen y O. Jepsen, *Phys. Rev. Lett.* 53, 2751 (1984);  
O. K. Andersen, Z. Pawłowska y O. Jepsen, **Phys. Rev. B** 34, 5253 (1986).
- [7] JEPSEN, O.; G. KRIER, A. BURKHARDT and O. K. ANDERSEN (1995): **The TB-LMTO-ASA program**, Max-Planck Institute, Stuttgart, Germany.
- [8] von BARTH, U. and L. HEDIN (1972): **J. Phys. C** 5, 1629.
- [9] CEPERLY, D.M. y B.J. ALDER (1980): **Phys. Rev. Lett.**, 566.
- [10] LANGRETH, D. C and M. J. MEHL: (1981): **Phys. Rev. Lett.** 47, 446;  
C. D. Hu y D. C. Langreth, **Physica Scripta.** 32, 391 (1985).
- [11] PERDEW, J.P. and Y. WANG (1986): **Phys Rev. B** 33, 8800.
- [12] CERIU<sup>2</sup> (2000): Versión 4.2 MatSci, **Manual de CASTEP version 3.1**, Molecular Simulations Inc.
- [13] VANDERBILT, D. (1990): **Phys. Rev. B** 41, 7892.
- [14] KLIENMAN, L. y D. M. BYLANDER (1982): **Phys. Rev. Lett.** 48, 1425.
- [15] PERDEW, J.P. and A. ZUNGER (1981): **Phys. Rev. B** 23, 5048.
- [16] PERDEW, J.P.; K. BURKE and M. ERNZERHOF (1996): **Phys. Rev. Lett.** 77, 3865.
- [17] MONKHORST, H.J. and J. D. PACK (1976): **Phys. Rev. B** 13, 5188.
- [18] BAUTISTA HERNÁNDEZ, A.; M. LÓPEZ-FUENTES, V. PACHECO-ESPEJEL and J.F. RIVAS SILVA (2005): **Rev. Mex. Fis.** 51, 157.
- [19] PETIT, L.; A. SVANE; W.M. TEMMERMAN and Z. SZOTEK (2002): **Phys. Rev. Lett.** 88(21), 216403-1.
- [20] CHIGO ANOTA, E.; A. FLORES RIVEROS y J. F. RIVAS SILVA (2006): Enviado a **Int. J. Quantum Chem.**
- [21] CHIGO ANOTA, E. y J. F. RIVAS SILVA (2005): Enviado a **Rev. Col. Fís.**
- [22] \_\_\_\_\_ (2004): **Rev. Mex. Fis. E** 50(2), 88;  
E. Chigo Anota y J. F. Rivas Silva, enviado a **Rev. Col. Fís.** (2005).
- [23] \_\_\_\_\_ (2005): **Rev. Col. Fís.** 37(2), 405.

# DISPERSIÓN INELÁSTICA DE LA LUZ POR EXCITACIONES ELECTRÓNICAS EN ÁTOMOS ARTIFICIALES

Alain Delgado<sup>1</sup>, Centro de Aplicaciones Tecnológicas y Desarrollo Nuclear, Ciudad de La Habana, Cuba  
Augusto González<sup>2</sup>, Instituto de Cibernética, Matemática y Física, Ciudad de La Habana, Cuba  
D.J. Lockwood<sup>3</sup>, Institute for Microstructural Sciences, National Research Council, Ottawa, Canada

## RESUMEN

En este artículo presentamos cálculos teóricos de la sección eficaz de dispersión inelástica de luz (Raman) por excitaciones electrónicas en un punto cuántico cargado con 42 electrones. Los estados multielectrónicos involucrados en el cálculo de la amplitud de transición Raman son obtenidos en la aproximación de fase aleatoria (RPA). La evaluación de las reglas de sumas pesadas por energía permitió clasificar las excitaciones electrónicas en uniparticulares (SPEs) y colectivas de carga (CDEs) y de espín (SDEs). Los espectros Raman calculados en geometría polarizada y no polarizada son mostrados en diferentes regiones de la energía del láser incidente. Se predicen teóricamente las ventajas asociadas al cálculo o medición de los espectros en régimen no resonante. Los valores calculados de las razones de polarización de las intensidades Raman demuestran el rompimiento de las reglas de selección del efecto Raman en presencia de campos magnéticos externos. La regla del salto de la intensidad Raman es propuesta como herramienta novedosa para la identificación de la naturaleza (carga o espín) de las excitaciones electrónicas. En la región de resonancia extrema, se obtienen espectros Raman dominados por picos asociados a estados excitados uniparticulares. Se argumenta cualitativamente la importancia de incorporar los tiempos de vida de los estados intermedios para la descripción de los espectros Raman en la región de energías del láser incidente 40-50 meV por encima de la brecha energética de la estructura semiconductor.

## ABSTRACT

In this article we present a theoretical investigation of the inelastic light scattering by electronic excitations in a quantum dot charged with 42 electrons. The energies and wave-functions of the multielectronic states involved in the Raman process, are obtained in the framework of Random Phase Approximations (RPA). Charge-Density (CDE) and Spin-Density (SDE) excitations are identified by evaluating the multipolar energy-weighted sum rules (EWSR). We compute Raman spectra for different values of the incident laser energy in both, polarized and depolarized geometry. Calculated Raman spectra with an excitation energy below the bandgap reveals several advantages for identifying and following individual peaks. The breakdown of Raman polarization selection rules under the influence of an external magnetic field is studied by calculating the polarization ratios. The breakdown of these selection rules under resonant excitation with the semiconductor bandgap, which we termed Raman intensity jump-rule, is found and proposed as a useful tool for identifying the character (charge or spin) of electronic excitations. We found that Raman spectra in the extreme resonance region are dominated by strong peaks associated to single-particle excitations (SPEs). The interference effects between the intermediate states in the Raman transition amplitude are evaluated. The main features of Raman spectra with excitation energy well-above the bandgap are qualitatively reproduced in terms of lifetimes of the intermediate states.

## I. INTRODUCCIÓN

En la actualidad un elevado porcentaje de las investigaciones de la física de la materia condensada, concentran sus esfuerzos en el estudio de las propiedades físicas de heteroestructuras semiconductoras cuyas dimensiones espaciales se encuentran en la escala nanométrica. Técnicas modernas sofisticadas como la Epitaxia de Haces Moleculares (MBE) han hecho posible la obtención de gases electrónicos cuasi-bidimensionales en materiales semiconductores. Tales sistemas constituyen un

punto de partida ideal para la fabricación de estructuras con dimensionalidad reducida como los hilos (cuasiunidimensionales) y los puntos (cuasicerodimensionales) cuánticos [1]. Estos últimos, conocidos también como átomos artificiales dada la cuantización completa de su espectro energético, son obtenidos por diversos métodos y con diversas geometrías [2]. El estudio de las propiedades ópticas y electrónicas de los puntos cuánticos ha tenido en los últimos años una gran fuerza [3] dado que los resultados de estas investigaciones tienen un impacto directo en el desarrollo de nuevas ramas de

E-mail: <sup>1</sup>gran@ceaden.edu.cu

<sup>2</sup>agonzale@icmf.inf.cu

<sup>3</sup>David.Lockwood@nrc-cnrc.gc.ca

la física aplicada. La espectroscopia de excitaciones electrónicas en átomos artificiales, o sea, puntos cuánticos que contienen un número  $N$  de electrones confinados en la banda de conducción, constituye la analogía más natural de la espectroscopia en sistemas atómicos. El espectro de estados excitados de átomos artificiales es una información básica para el desarrollo de dispositivos electrónicos cuyos principios de funcionamiento se basan en los efectos cuánticos presentes en estas nanoestructuras. La dispersión inelástica de luz (Raman) permite investigar diferentes sectores del espectro de excitaciones electrónicas de estos sistemas. En la década de los años 1990 aparecieron en la literatura los primeros trabajos experimentales sobre mediciones de espectros Raman asociados a excitaciones electrónicas en átomos artificiales que contienen cientos de electrones [4–7]. Una amplia descripción de los resultados obtenidos en estos trabajos aparece publicada en las referencias [8,9]. Recientemente se han reportado mediciones de espectros Raman de excitaciones electrónicas en puntos cuánticos notablemente pequeños que contienen como promedio entre 2 y 6 electrones confinados en las dimensiones de la estructura semiconductor [10,11].

En general las excitaciones electrónicas en puntos cuánticos pueden clasificarse en excitaciones uniparticulares (SPEs) y excitaciones colectivas de carga (CDEs) y de espín (SDEs). Los picos Raman asociados a las CDEs pueden ser medidos o calculados en la llamada configuración de geometría polarizada, en la cual los vectores de campo eléctrico del fotón dispersado e incidente son paralelos, mientras que los estados excitados del tipo SDEs son obtenidos en geometría no polarizada donde el vector de campo eléctrico del fotón dispersado es ortogonal con respecto a la misma magnitud del fotón incidente. Esta diferenciación está dada por las reglas de selección del efecto Raman [12]. Schüller y colaboradores [5,7] han mostrado experimentalmente que fuertes picos Raman asociados a SPEs aparecen en los espectros medidos como resultado de iluminar el punto con un láser cuya energía está muy cercana al valor de la brecha energética ( $E_{\text{gap}}$ ) del semiconductor mientras que para energías de incidencias muy por encima de  $E_{\text{gap}}$  el espectro Raman es dominado por excitaciones de naturaleza colectiva.

La obtención de estos resultados experimentales motivó un conjunto de estudios teóricos encaminados a la descripción de la fenomenología encontrada en los citados trabajos. Estas teorías [13–17] sólo válidas en régimen no resonante, han descrito las posiciones de las excitaciones colectivas para diferentes valores del vector de onda de la radiación incidente y de campos magnéticos externos. Sin embargo, los citados formalismos presentan limitaciones inherentes a las aproximaciones usadas

tales como: i) la incapacidad de describir las relaciones de intensidades entre los picos Raman asociados a SPEs y CDEs y/o SDEs obtenidas en los experimentos, ii) la incapacidad de predecir la presencia de picos Raman asociados a SPEs, iii) no consideran los estados electrónicos en la banda de valencia, los cuales juegan un papel crucial en el cálculo de la sección eficaz de dispersión Raman [18]. Este artículo tiene como objetivo fundamental enriquecer y complementar las interpretaciones discutidas en los trabajos experimentales publicados en esta temática aunque no sólo se limita a eso, dado que nuevas predicciones teóricas han emergido del análisis de los cálculos realizados.

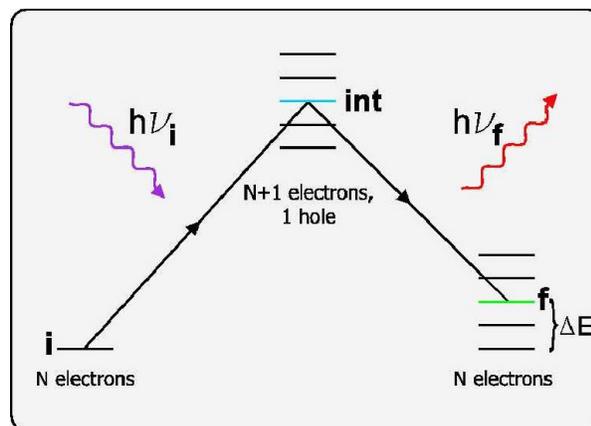
En la próxima sección describimos de manera general el esquema de cálculo implementado para la obtención de la sección eficaz diferencial Raman. La sección III contiene la discusión y análisis de los resultados. Finalmente las conclusiones generales son brindadas en la sección IV.

## II. MÉTODOS TEÓRICOS Y ESQUEMA DE CÁLCULO IMPLEMENTADO

La idea básica de la dispersión Raman es simple: un fotón con energía  $h\nu_i$  el cual es dispersado inelásticamente por la materia puede ganar (procesos anti-Stoke) o perder (Stoke) energía. La teoría de perturbaciones permite obtener la expresión para el cálculo de la amplitud de transición Raman [20],  $A_{fi}$ , definida por la siguiente ecuación:

$$A_{fi} \sim \sum_{\text{int}} \frac{\langle f | \hat{H}_{e-r}^+ | \text{int} \rangle \langle \text{int} | \hat{H}_{e-r}^- | i \rangle}{h\nu_i - (E_{\text{int}} - E_i) + i\Gamma_{\text{int}}} \quad (1)$$

Una representación esquemática del proceso de dispersión Raman es mostrado en la Figura 1.  $|i\rangle$  y  $|f\rangle$  son los estados electrónicos inicial y final respectivamente que describen a los  $N$ -electrones confinados en el punto cuántico. En principio, la suma en la



**Figura 1.** Amplitud de transición Raman y su interpretación en términos de transiciones virtuales.

Ecuación (1) debe incluir todos los estados electrónicos que no anulen los elementos matriciales en el numerador. En los trabajos experimentales mencionados la energía del láser incidente,  $h\nu_i$ , toma valores del orden de  $E_{\text{gap}}$ . Esto permite restringir la suma en la Ecuación (1) por estados intermedios que son excitaciones interbandas de la estructura semiconductor, o sea, por estados de un sistema en el cual existen  $(N + 1)$  - electrones confinados en la banda de conducción y 1 hueco confinado en la banda de valencia.  $\hat{H}_{e-r}$  denota el Hamiltoniano de interacción del sistema electrónico con el campo de fotones de la radiación incidente [12].  $E_i$  y  $E_{\text{int}}$  son las energías correspondientes a los estados iniciales e intermedios y  $\Gamma_{\text{int}}$  es un parámetro asociado a los tiempos de vida de estos últimos. La ley de conservación de la energía establece la ecuación,

$$h\nu_i - h\nu_f = E_f - E_i = \Delta E, \quad (2)$$

donde  $h\nu_f$  es la energía del fotón dispersado,  $E_f$  es la energía del estado excitado final y  $\Delta E$  es la energía de excitación electrónica o corrimiento Raman asociado al estado  $|f\rangle$ . Una vez obtenida la amplitud

de transición Raman podemos calcular la sección eficaz diferencial de dispersión Raman,  $d\sigma$ , a través de la siguiente expresión:

$$d\sigma \sim \sum_f |A_{fi}|^2 \delta(E_i + h\nu_i - E_f - h\nu_f). \quad (3)$$

La conservación de la energía es forzada por la función delta de Dirac en la Ecuación (3). En nuestros cálculos, ésta es simulada por la función de Lorentz:

$$\delta(x) \approx \frac{\Gamma_f / \pi}{x^2 + \Gamma_f^2} \quad (4)$$

### A. Esquema computacional para la obtención de la amplitud de transición Raman

Las ecuaciones y expresiones explícitas detalladas de las diferentes aproximaciones involucradas en el cálculo de la amplitud de transición Raman,  $A_{fi}$ , pueden ser encontradas por los lectores interesados en las referencias [18,19]. En esta sección brindaremos una descripción de la metodología de cálculo implementada auxiliándonos de la Figura 2.

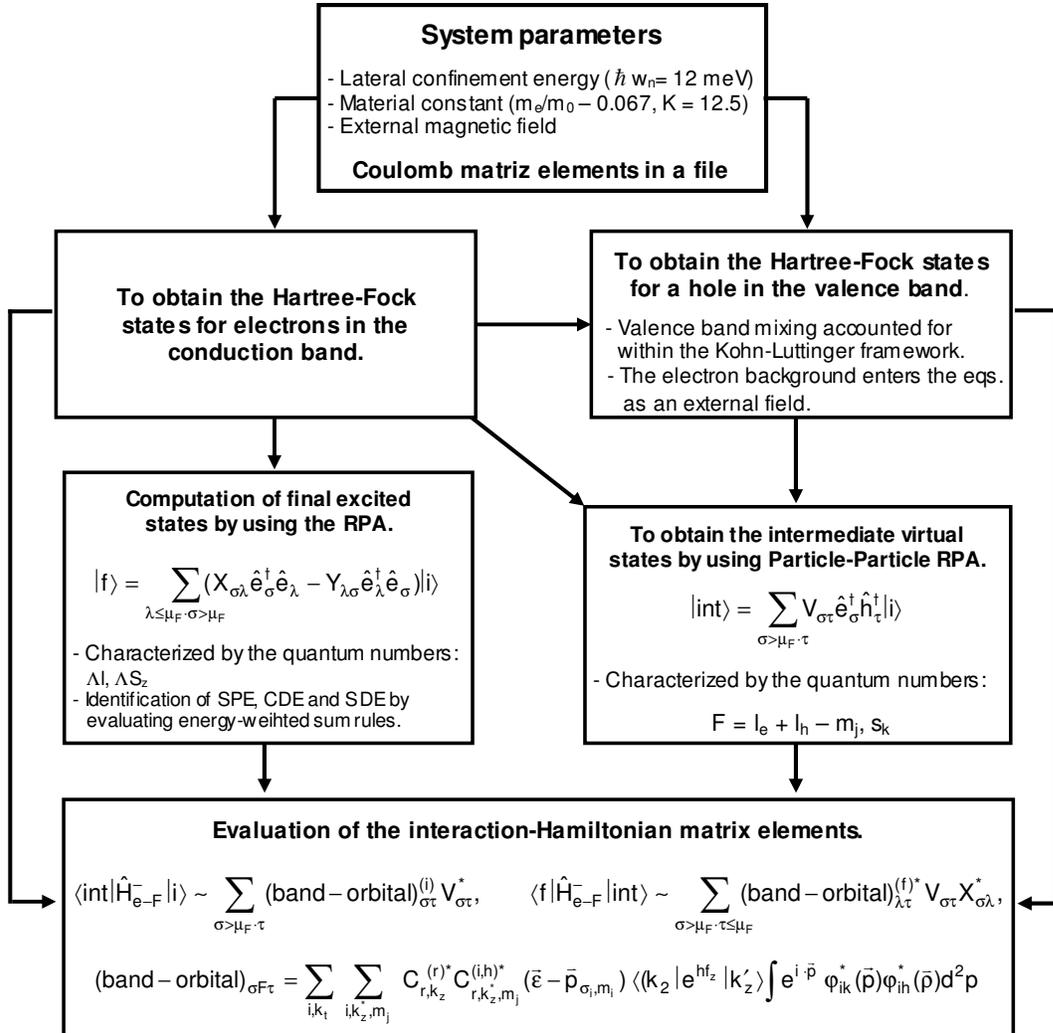


Figura 2. Esquema de cálculo implementado.

El punto cuántico en forma de disco es modelado con un potencial de paredes rígidas en la dirección  $z$  (dirección de crecimiento de la heteroestructura) y un confinamiento lateral (direcciones  $x$ ,  $y$ ) de tipo parabólico con una energía característica  $\hbar\omega_0 = 12$  meV. La base de funciones utilizada que caracteriza los estados de una partícula están construídas como el producto de las autofunciones del oscilador bidimensional, del pozo infinito y las funciones de espín [18].

Con el objetivo de hacer eficiente el cálculo de las energías y funciones de onda de los sistemas multielectrónicos involucrados, nosotros calculamos previamente todos los valores de los elementos matriciales de la interacción de Coulomb,  $\langle \alpha, \beta | 1/r | \gamma, \delta \rangle$ , donde  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\delta$  son estados arbitrarios de una partícula. Estos elementos matriciales son cargados en memoria al inicio de cada cálculo permitiéndonos resolver las ecuaciones no lineales íntegro-diferenciales de Hartree-Fock para 42 electrones en pocos minutos, o calcular todos los estados intermedios considerados (alrededor de 10,000) en la Ecuación (1) en pocos días. Las ecuaciones de Hartree-Fock para huecos incluye el campo electrostático generado por el background de electrones en el punto y considera los efectos de la mezcla de las bandas de valencia a través del Hamiltoniano de Kohn-Luttinger [18].

Los estados excitados finales e intermedios son descritos en el marco de la Aproximación de Fase Aleatoria [21] (RPA). Mediante la RPA es posible calcular las energías y funciones de onda de los estados finales en diferentes sectores del espectro de excitaciones electrónicas del punto, donde cada sector está definido por reglas de selección para la variación (respecto al valor de la magnitud en el estado base) del momentum angular ( $\Delta l = 0, 1, 2, \dots$ ) y la proyección del espín ( $\Delta S_z = 0, \pm 1$ ). La evaluación de reglas de sumas pesadas por energía [21] permite identificar, de manera independiente al cálculo de la amplitud de transición Raman, los estados colectivos de carga (CDEs), de espín (SDEs) y las excitaciones uniparticulares (SPEs). Por otro lado, los estados intermedios del proceso Raman, están caracterizados por la proyección del espín del electrón añadido  $S_z$  y el momento angular total del par,  $F = l_e + l_h - m_j$ .

Una vez calculados los estados finales e intermedios son evaluados los elementos matriciales del Hamiltoniano de interacción del punto cuántico con el campo electromagnético en términos de los coeficientes que caracterizan las funciones de onda de los estados multielectrónicos finales e intermedios y los coeficientes de la expansión de los estados uniparticulares de Hartree-Fock de electrones y huecos en la base de funciones utilizada.

## B. Aproximación para el cálculo de la amplitud de transición Raman en régimen no resonante (ORA)

La ORA (*off-resonant approximation*) es una aproximación en la cual la expresión para el cálculo de la amplitud de transición Raman,  $A_{fi}$ , involucra solamente a los estados final e inicial del punto cuántico. La suma por estados intermedios desaparece; esto significa que en los marcos de esta aproximación no es posible la explicación de efectos producto de resonancias con estados intermedios en el proceso de dispersión Raman.

La deducción de la expresión para la amplitud de transición Raman en esta aproximación puede ser consultada en la referencia [12]. Las suposiciones básicas para su derivación son: (i) la energía del láser incidente está lo suficientemente lejos de las energías de los estados intermedios, por lo que la dependencia en la Ecuación (1) con  $E_{int}$  en el denominador puede ser despreciada; (ii) existe una ventana energética de estados intermedios (40 meV por encima de  $E_{gap}$ ) donde las variaciones de  $\Gamma_{int}$  son despreciables y la relación de completitud  $\sum_{int} |int\rangle\langle int| \approx 1$  es casi satisfecha. Bajo estas suposiciones la amplitud de transición Raman,  $A_{fi}^{ORA}$ , puede escribirse como:

$$A_{fi}^{ORA} \sim \langle f | H_{e-r}^+ H_{e-r}^- | i \rangle. \quad (5)$$

Haciendo uso de las expresiones para  $H_{e-r}^+$  y  $H_{e-r}^-$  [12] es posible deducir la siguiente fórmula para el cálculo de la amplitud de transición Raman:

$$A_{fi}^{ORA} \sim \sum_{\alpha, \alpha'} \langle \alpha | e^{i(\bar{q}_i - \bar{q}_f) \cdot \bar{r}} | \alpha' \rangle \left\{ \frac{2}{3} (\bar{\epsilon}_i \cdot \bar{\epsilon}_f) \langle f | e_{\alpha\uparrow} e_{\alpha'\uparrow} + e_{\alpha\downarrow} e_{\alpha'\downarrow} | i \rangle \right. \\ \left. + \frac{1}{3} (\bar{\epsilon}_i \times \bar{\epsilon}_f) \cdot \langle f | \hat{z} (e_{\alpha\uparrow} e_{\alpha'\uparrow} - e_{\alpha\downarrow} e_{\alpha'\downarrow}) + (\hat{x} - i\hat{y}) e_{\alpha\uparrow} e_{\alpha'\downarrow} + (\hat{x} + i\hat{y}) e_{\alpha\downarrow} e_{\alpha'\uparrow} | i \rangle \right\}. \quad (6)$$

Del análisis de la Ecuación (6) podemos inferir algunas conclusiones. Primero, solamente los estados finales excitados colectivos tendrán una amplitud Raman asociada diferente de cero. No es posible describir picos Raman asociados a excitaciones uniparticulares en esta aproximación. Segundo, las reglas de selección del efecto Raman asociadas a las polarizaciones de los fotones incidente y dispersado pueden ser entendidas de la Ecuación (6): el primer término, el cual no altera el número cuántico de espín del estado inicial está

multiplicado por un factor  $\bar{\epsilon}_i \cdot \bar{\epsilon}_f$ . Esto significa que los estados finales colectivos de carga aparecen en geometría polarizada. Por otro lado, los otros términos que modifican el espín del estado inicial están multiplicados por un factor  $\bar{\epsilon}_i \times \bar{\epsilon}_f$  y consecuentemente los picos Raman asociados a estas excitaciones aparecen en geometría no polarizada.

### III. RESULTADOS

En este artículo reportamos los resultados para un punto cuántico con 42 electrones confinados en la banda de conducción. El ancho del pozo en la dirección  $z$  es  $L = 25$  nm y la energía de confinamiento característica es  $\hbar\omega_0 = 12$  meV. El número de electrones corresponde a un punto de capas cerradas. En nuestro estudio consideramos estados excitados con energías de excitación  $\Delta E \leq \hbar\omega_{LO}$ , siendo  $\hbar\omega_{LO} \approx 30$  meV el valor de energía umbral para la emisión de fonones ópticos longitudinales en el GaAs. Esto significa que los estados finales tienen pequeños anchos para los cuales hemos tomado un valor constante  $\Gamma_f = 0.1$  meV. De igual manera para los estados intermedios con energías de excitación menores que  $\hbar\omega_{LO}$  escogemos el parámetro  $\Gamma_{int} = 0.5$  meV. Para valores superiores en energías de excitación, como resultado de la emisión de fonones LO fijamos el parámetro  $\Gamma_{int} = 10$  meV. Todos los espectros Raman están calculados en la configuración de retrodispersión, donde la luz incidente y dispersada forman un ángulo de  $20^\circ$  respecto a la normal del punto.

#### A. Espectro de excitaciones electrónicas

En la Figura 3 mostramos el espectro de estados finales excitados del punto modelado, calculados en la aproximación de fase aleatoria. Las energías de

excitación de estos estados es precisamente la magnitud medida en los experimentos de dispersión inelástica de luz (corrimiento Raman). En la figura se identifican los estados colectivos de carga (CDE) y espín (SDE) en todos los sectores y la contribución de cada estado colectivo a la regla de suma pesada por energía es explícitamente señalada. El resto de los estados mostrados corresponden a excitaciones uniparticulares.

La evolución del espectro de estados excitados monopoles ( $\Delta l = 0, \Delta S_z = 0$ ) como función del campo magnético externo es mostrada en la Figura 4. Las excitaciones colectivas de carga y espín, cuya contribución a la regla de suma pesada por energía excede el 5%, son representadas con cuadrados y triángulos respectivamente. La contribución de cada modo es proporcional al tamaño de los símbolos usados. Las excitaciones uniparticulares de carga (SPEs(C)) y espín (SPEs(S)) son también mostradas con líneas horizontales largas y cortas respectivamente.

Con relación a las excitaciones colectivas de carga se pueden notar dos características fundamentales: (i) la energía de excitación asociada a este modo evoluciona de manera suave respecto al campo magnético externo y (ii) el tamaño de los símbolos no cambia para diferentes valores del campo, manifestándose la fuerte naturaleza colectiva de estos modos.

Los estados de bajas energías de excitación corresponden a excitaciones colectivas monopoles de espín. A diferencia de los modos colectivos de carga podemos observar, para valores de campos magnéticos  $B \geq 2$  T más de un estado colectivo contribuyendo a la regla de suma pesada por energía.

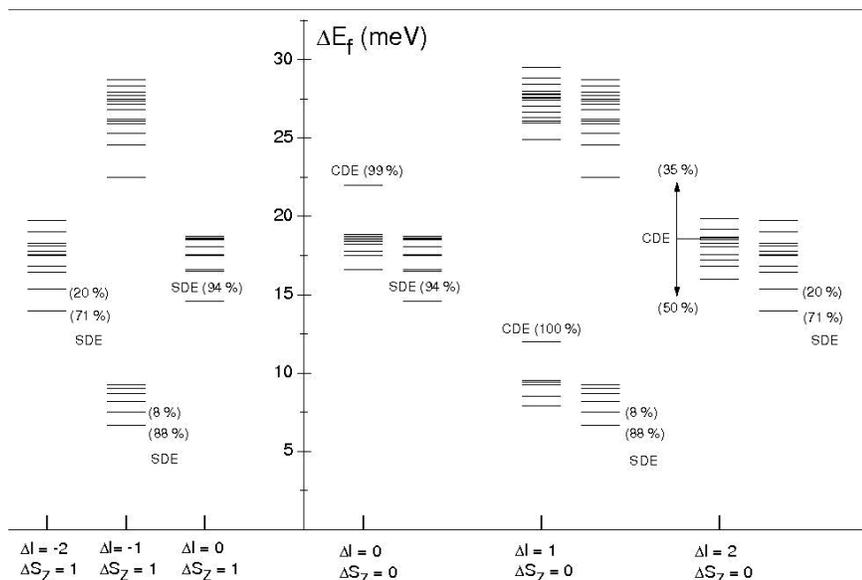
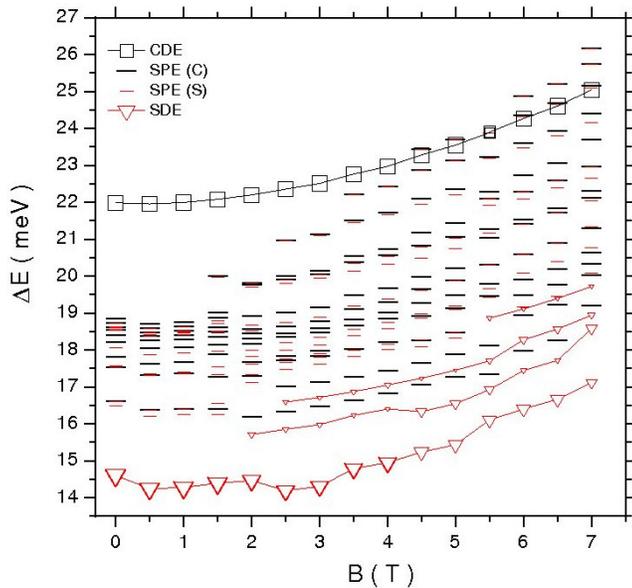


Figura 3. Espectro de excitaciones electrónicas multipolares del punto cuántico.  $B = 0$  T.



**Figura 4.** Espectro de excitaciones electrónicas monopolares ( $\Delta l = 0$ ,  $\Delta S_z = 0$ ) vs campo magnético externo.

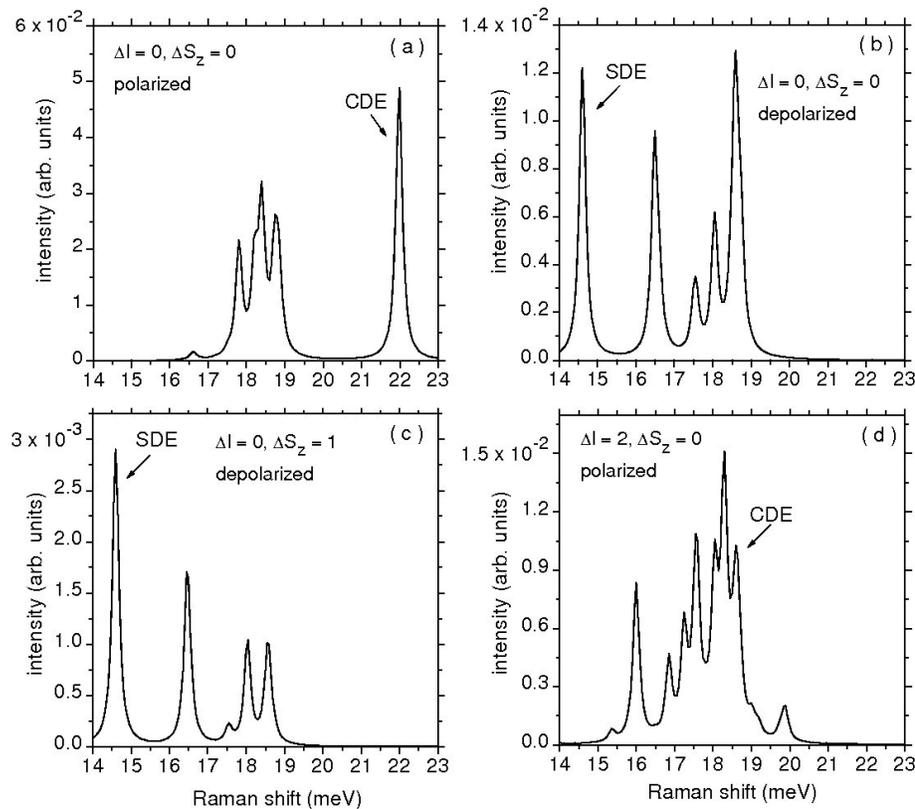
### B. Espectros Raman en régimen no resonante ( $h\nu_i < E_{\text{gap}}$ )

En este régimen no existen resonancias con estados intermedios. Sin embargo, esto no implica que la física de la espectroscopia Raman pueda ser descrita en su totalidad con una aproximación que desprece la contribución de dichos estados como la ORA. Las

principales características de los espectros Raman en esta región son expuestos de manera resumida a continuación. Un análisis más amplio de los mismos puede ser consultado en la referencia [22].

#### 1. Contribución de las excitaciones electrónicas multipolares en los espectros Raman

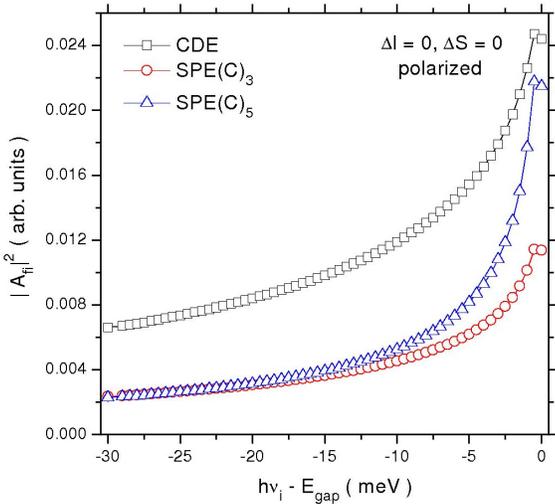
Espectros Raman asociados a excitaciones electrónicas en diferentes sectores de momento angular y espín son mostrados en la Figura 5. Las intensidades Raman correspondientes a excitaciones monopolares y cuadrupolares ( $\Delta l = 2$ ) exhiben magnitudes comparables aunque se observa un predominio de los picos Raman asociados a los modos monopolares. La contribución al espectro Raman en geometría no polarizada de las excitaciones de espín del tipo spin-flip ( $\Delta S_z = \pm 1$ ) es mucho más débil comparada con la contribución de los estados excitados de espín caracterizados por un cambio en el valor del espín total pero no en el valor de su proyección. Dado que ambos modos están muy cercanos en energías los picos Raman colectivos medidos en geometría no polarizada son adjudicables a estados excitados en los cuales la proyección total del espín del estado base no es modificada en el proceso de dispersión. Los espectros Raman asociados a excitaciones dipolares ( $\Delta l = 1$ ) [18] revelan intensidades uno o dos órdenes de magnitud menores en comparación con las mostradas en la Figura 5.



**Figura 5.** Espectros Raman para diferentes sectores de excitaciones electrónicas.  $h\nu_i = E_{\text{gap}} - 5$  meV,  $B = 0$  T.

## 2. Comportamiento de las intensidades Raman como función de la energía del láser incidente

Es posible, a partir del esquema de cálculo implementado, monitorear el valor de la intensidad Raman asociada a cualquier estado excitado del punto cuántico en la medida que se varía la energía del láser incidente. En esta región, donde la energía de la luz que incide sobre el punto es variada en el intervalo  $(E_{\text{gap}}-30 \text{ meV}, E_{\text{gap}})$ , las intensidades Raman asociadas tanto a los modos SPEs como a los modos colectivos (CDEs, SDEs) muestran un comportamiento monótono creciente cuando  $h\nu_i$  se incrementa. Este comportamiento se muestra en la Figura 6 donde hemos planteado los valores de  $|A_{fi}|^2$  para dos estados uniparticulares y el modo colectivo del sector monopolar de excitaciones electrónicas. Aunque no se han reportado mediciones de espectros Raman en esta región, desde el punto de vista teórico se predice un comportamiento de las intensidades Raman muy conveniente para la identificación de los picos en un experimento real. Otro aspecto a resaltar es que, efectivamente, existe un límite para el valor de energía del láser incidente, por debajo de la brecha energética del semiconductor, a partir del cual los picos más fuertes en el espectro están asociados a excitaciones colectivas. Dicho límite, a partir del cual la fenomenología del efecto Raman puede ser descrita con una aproximación del tipo de la ORA fue investigado en la referencia [12].

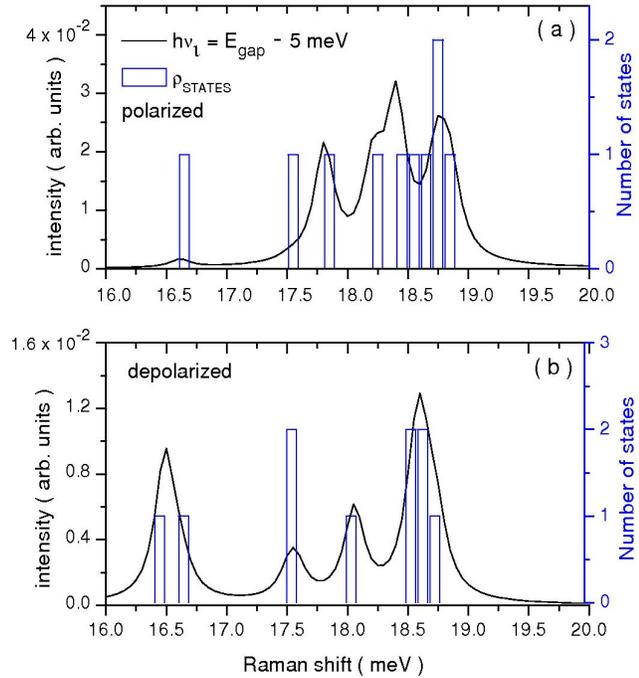


**Figura 6.** Comportamiento de  $|A_{fi}|^2$  para el modo colectivo monopolar y dos modos uniparticulares de carga como función de la energía del láser incidente.  $B = 0 \text{ T}$ .

## 3. Correlación entre las intensidades de los picos Raman y la densidad de estados finales

Es razonable esperar picos Raman intensos para valores de la energía de excitación donde existe una aglomeración de estados excitados finales de una partícula. Lo novedoso en este punto, mostrado para

el caso de excitaciones monopoles en la Figura 7, es la estrecha correlación obtenida entre la densidad de niveles de energía de los SPEs de carga con la intensidad de los picos Raman obtenidos en la configuración de geometría polarizada y entre los SPEs de espín con la intensidad de los picos Raman obtenidos en la configuración de geometría no polarizada en ausencia de campos magnéticos externos. O sea, la correlación existente entre la intensidad de los picos Raman y la densidad de estados excitados uniparticulares está mediada también, en ausencia de campos magnéticos externos, por las reglas de selección del efecto Raman deducibles sólo para los modos colectivos.

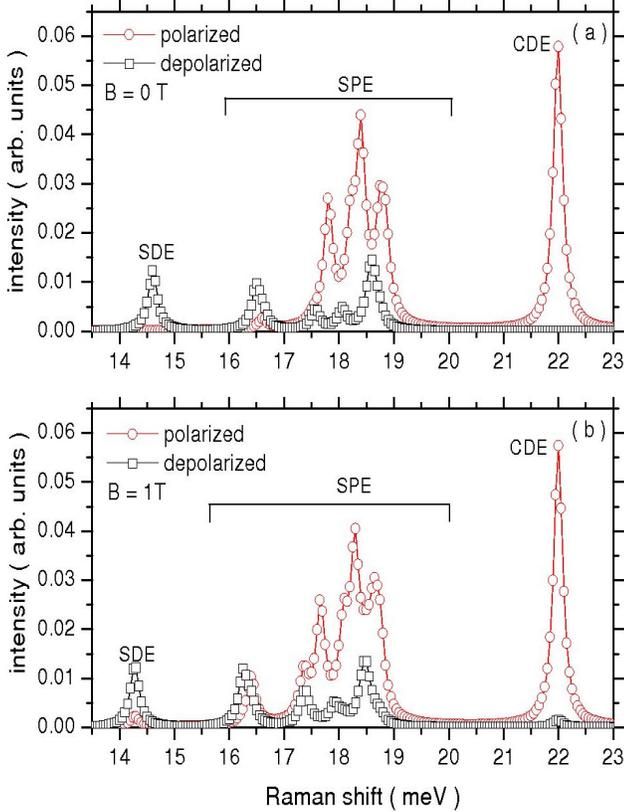


**Figura 7.** Espectros Raman de excitaciones monopoles en ambas geometrías. Comparación con la densidad de estados SPEs.  $B = 0 \text{ T}$ .

## 4. Rompimiento de las reglas de selección del efecto Raman en un campo magnético externo

En presencia de un campo magnético externo perpendicular al plano de movimiento de los electrones confinados en el punto, las reglas de selección del efecto Raman se rompen parcialmente para los estados colectivos y drásticamente para las excitaciones de una partícula. Este efecto es ilustrado en la Figura 8 donde hemos planteado los espectros Raman asociados a excitaciones monopoles para  $B = 0$  y  $1 \text{ T}$ . Mediante el cálculo de las razones de polarización [23] de los picos Raman monopoles para  $B = 0, 1 \text{ T}$  se evaluaron cuantitativamente los efectos del campo magnético en las reglas de selección de la dispersión Raman: (i) se chequearon para  $B = 0$  las reglas de selección para las intensidades Raman asociadas a los estados colectivos, (ii) el análisis de las razones de polarización de las

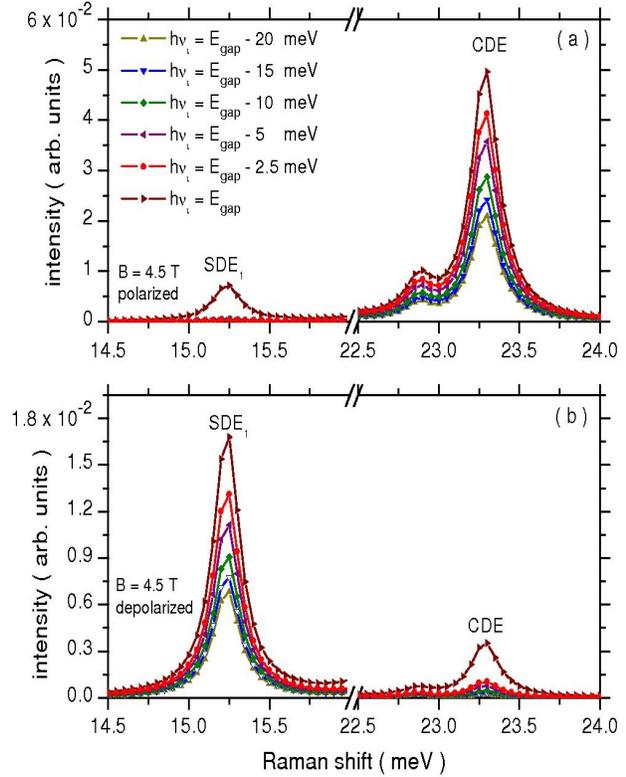
intensidades Raman para cada modo uniparticular mostró que dichas excitaciones, al igual que los modos colectivos, obedecen las reglas de selección del efecto Raman en ausencia de campos magnéticos externos, (iii) la presencia de un campo magnético externo, en este caso  $B = 1$  T, rompe las reglas de selección del efecto Raman parcialmente para los modos colectivos y drásticamente para los estados excitados uniparticulares.



**Figura 8.** Espectros Raman de excitaciones monopolares para  $B = 0$  y  $1$  T en ambas geometrías.

### 5. Regla del salto de la intensidad Raman cuando $h\nu_i = E_{\text{gap}}$

Este es un resultado bien interesante de esta investigación y está estrechamente ligado con lo explicado en la sección anterior. En presencia de un campo magnético externo si seguimos el comportamiento de la intensidad Raman correspondiente a un estado excitado de carga (espín) en geometría no polarizada (polarizada) como función de la energía del láser incidente, encontramos valores muy pequeños de las intensidades cuando  $h\nu_i < E_{\text{gap}}$ . Sin embargo, si  $h\nu_i = E_{\text{gap}}$  la amplitud de transición Raman aumenta de manera abrupta. Este comportamiento lo denominamos la regla del salto para la intensidad Raman y es mostrado gráficamente en la Figura 9 para los modos colectivos monopolares en un campo magnético externo  $B = 4.5$  T.

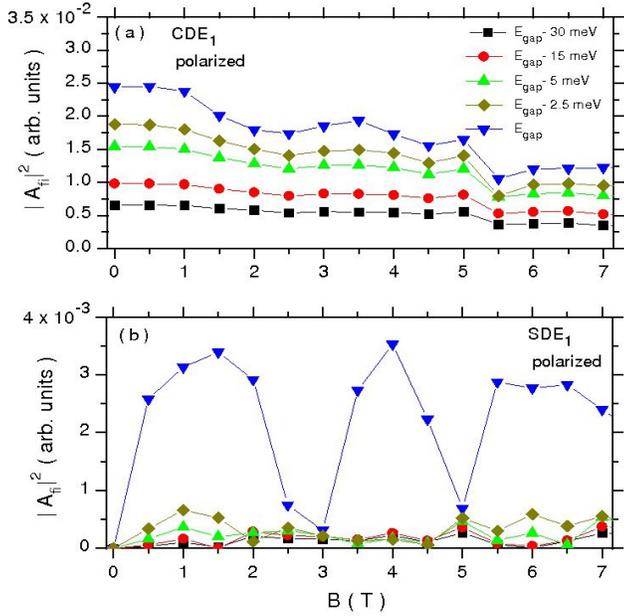


**Figura 9.** Espectros Raman de excitaciones monopolares en ambas geometrías para  $B = 4.5$  T. Los picos Raman asociados a las excitaciones uniparticulares han sido sustraídos de los gráficos.

El comportamiento de las intensidades Raman de estos mismos modos, en geometría polarizada para diferentes valores de  $h\nu_i$  y del campo magnético externo  $B$ , es mostrado en la Figura 10. La regla del salto de la intensidad Raman es apreciable en la Figura 10 (b) para el modo colectivo de espín en todos los valores de campo excepto en los puntos  $B = 0$  (regla de selección),  $2.5$ ,  $3.0$  y  $5.0$  T donde observamos un colapso de este comportamiento. Un análisis del mismo tipo al presentado en la Figura 10 se realizó para las excitaciones uniparticulares de bajas energías obteniéndose un resultado similar [23]. La regla del salto es obedecida por todos los modos monopolares para valores pequeños de  $B$ . Aunque esta regla no es universal respecto al campo magnético externo, constituye una herramienta útil para identificar la naturaleza de las excitaciones electrónicas en puntos cuánticos.

### C. Espectros Raman en régimen de resonancia extrema ( $h\nu_i \approx E_{\text{gap}}$ )

Algunas de las características discutidas en la sección anterior se manifiestan también en este régimen, sin embargo, aparecen también nuevos efectos en los espectros Raman, característicos de la espectroscopia Raman en régimen de excitación resonante.



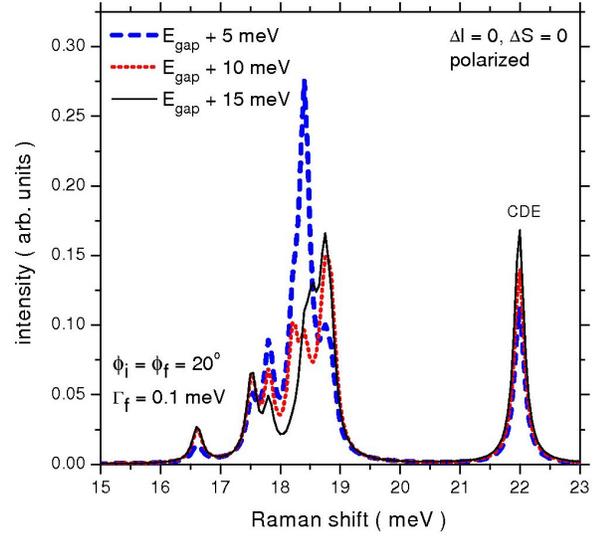
**Figura 10.** Comportamiento de las amplitudes de transición Raman correspondientes a CDE (a) y SDE1 (b) en geometría polarizada vs. B.

### 1. Predominio de los picos Raman correspondientes a excitaciones uniparticulares

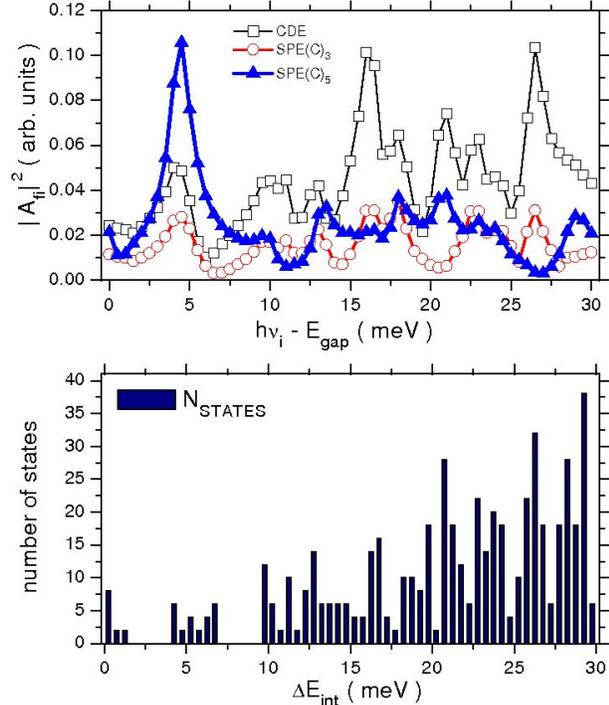
Las intensidades Raman asociadas a las excitaciones uniparticulares experimentan un incremento notable en condiciones de excitación resonante. En este régimen la mayor contribución a la amplitud de transición Raman, Ecuación. (1), está dada por el estado intermedio en resonancia con la energía del láser incidente. Este estado (virtual) decae indiscriminadamente a estados excitados uniparticulares o colectivos. Como fue mostrado en las Figuras 3 y 4, existe un número grande de estados finales uniparticulares compactados en regiones estrechas de la energía de excitación (picos en la densidad de estados finales) contribuyendo esto a que los picos Raman en esta región sean mayores que aquellos correspondientes a los modos colectivos. Este resultado está ilustrado en la Figura 11, la cual muestra los espectros Raman monopoles para diferentes valores de  $h\nu_i$ .

### 2. Correlación entre las intensidades de picos Raman individuales y la densidad de niveles de energía de los estados intermedios

En la Figura 12 planteamos los valores de  $|A_{fi}|^2$  como función de  $h\nu_i$  en el intervalo  $(E_{gap}, E_{gap} + 30 \text{ meV})$  en conjunto con la densidad de niveles de energías de los estados intermedios para los mismos estados finales usados en la construcción de la Figura 6. Como resultado de las resonancias con los estados intermedios podemos observar fluctuaciones abruptas en los valores de las amplitudes de transición Raman planteados. Los picos en los valores de  $|A_{fi}|^2$  reproducen, fundamentalmente en la región donde  $h\nu_i - E_{gap} \leq 15 \text{ meV}$ , los inicios de los conglomerados de estados intermedios.



**Figura 11.** Espectros Raman de excitaciones monopoles en régimen de resonancia extrema. B = 0 T.

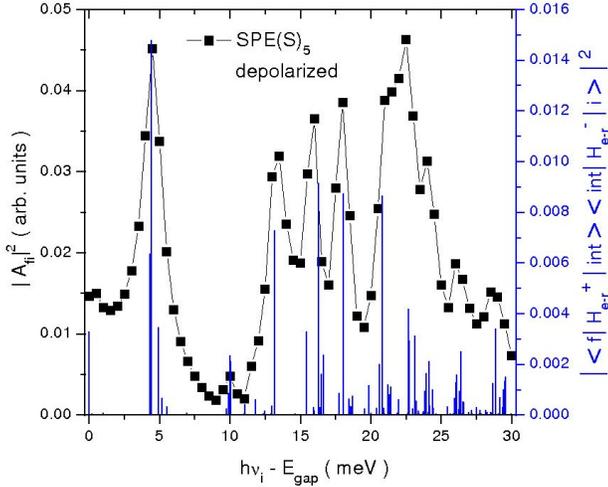


**Figura 12.** Comportamiento de  $|A_{fi}|^2$  asociados a excitaciones monopoles en régimen resonante en conjunto con la densidad de estados intermedios. B = 0 T.

### 3. Efectos de interferencia en el cálculo de la amplitud de transición Raman

En la Figura 13 evaluamos la contribución de los diferentes estados intermedios que entran en la suma de la Ecuación (1). En esta figura comparamos la magnitud  $|A_{fi}|^2$  asociada a un estado final (uniparticular) de espín con las contribuciones individuales del estado intermedio en resonancia con la energía del láser incidente. Como conclusión podemos extraer que

sólo aquellos estados intermedios cuyas energías están muy cercanas al valor de  $h\nu_i$  contribuirán a la amplitud de transición Raman, por tanto los efectos de interferencia en el cálculo de  $A_{fi}$  son débiles.

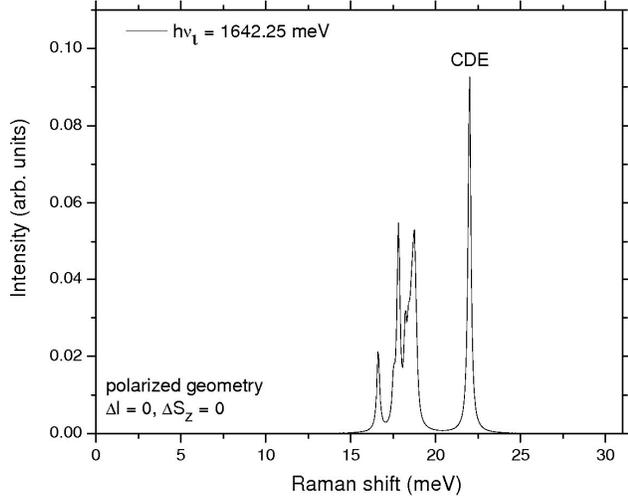


**Figura 13.**  $|A_{fi}|^2$  asociado a una excitación uniparticular de espín en conjunto con la contribución individual de cada estado intermedio en resonancia con  $h\nu_i$ .  $B = 0$  T.

#### D. Espectros Raman en régimen resonante en la región donde $h\nu_i \gg E_{\text{gap}}$

Como mencionamos al inicio de la sección III, para el cálculo de la amplitud de transición Raman en este régimen asumimos que el parámetro  $\Gamma_{\text{int}}$ , asociado al tiempo de vida del estado intermedio  $|int\rangle$ , experimenta un incremento abrupto cuando  $E_{\text{int}} > E_{\text{gap}} + \hbar\omega_{\text{LO}}$ . Como resultado, la contribución de estos estados a  $A_{fi}$  no es resonante, por lo que la intensidad de los picos Raman serán funciones suaves respecto a  $h\nu_i$ , en similitud a lo explicado en la sección III B2. Sin embargo, un incremento en la intensidad Raman dado un valor particular de  $h\nu_i$  en este régimen no sólo puede estar dado por un incremento en el valor del producto  $\langle f | \hat{H}_{e-r}^+ | X \rangle \langle X | \hat{H}_{e-r}^- | i \rangle$ , una disminución en el valor del parámetro  $\Gamma_{\text{int}}$  se manifiesta también en un aumento de la intensidad Raman. En el conjunto de estados intermedios incluidos en el cálculo de la amplitud de transición Raman pueden existir estados con propiedades físicas específicas [18] que se manifiestan en una disminución del  $\Gamma_{\text{int}}$  asociado. En la Figura 14 mostramos un espectro Raman calculado con  $h\nu_i = E_{\text{gap}} + 40$  meV. Esta energía de incidencia la hemos escogido en resonancia con un estado intermedio cuya anchura energética ha sido fijada a  $\Gamma_{\text{int}} = 2$  meV. En la figura se puede observar un espectro Raman donde el pico más intenso corresponde al modo colectivo monopolar de carga. La inclusión de los tiempos de vida de los estados intermedios [24] en el esquema de cálculo implemen-

tado podría complementar cuantitativamente la idea cualitativa expuesta en esta sección, lo cual explicaría la fenomenología observada en este régimen.



**Figura 14.** Espectro Raman monopolar en geometría polarizada.  $h\nu_i = E_{\text{gap}} + 40$  meV.  $B = 0$  T.

#### IV. CONCLUSIONES

En este trabajo hemos implementado un esquema de cálculo consistente de la amplitud de transición Raman asociada a excitaciones electrónicas en un punto cuántico cargado con 42 electrones. Los espectros Raman calculados han revelado características muy interesantes y una gran sensibilidad con respecto a la energía de la luz incidente y a la presencia o no de campos magnéticos externos. En la región de régimen no resonante, las intensidades de los picos Raman dependen suavemente de la energía del láser incidente. Para  $B = 0$ , las reglas de selección del efecto Raman, deducibles a partir de la ORA para los modos colectivos, es obedecida también por los estados excitados uniparticulares. Esto implica, que es posible obtener información del perfil de la densidad de estados finales excitados a partir de la estructura del espectro Raman medido o calculado. En presencia de campos magnéticos externos, se rompen las reglas de selección de la polarización del efecto Raman, particularmente cuando  $h\nu_i$  es incrementado hasta la brecha energética del semiconductor, efecto que hemos denominado regla del salto de la intensidad Raman. En régimen de excitación resonante las intensidades Raman muestran un comportamiento de fuertes fluctuaciones con la energía de la luz incidente como resultados de las resonancias con los estados intermedios. En este régimen los picos más intensos en los espectros están en correspondencia con estados excitados uniparticulares. Se determinó que los efectos de interferencia en el cálculo de la amplitud de transición Raman son débiles, quedando el valor de esta magnitud determinado por la contribución de los estados intermedios cuasi-resonantes con la energía del láser incidente. Una

interpretación cualitativa para los espectros Raman en la región donde la energía de la radiación incidente está muy por encima de la brecha energética del semiconductor es propuesta. Estos resultados complementan las interpretaciones que se han

brindado a los trabajos experimentales publicados en el tema y estimulan la realización de futuros experimentos que podrían ayudar a la comprensión de aspectos complejos de la espectroscopia Raman de excitaciones electrónicas en puntos cuánticos.

#### REFERENCIAS

- [1] HEITMANN, D. and J.P. KOTTHAUS (1993): **Phys. Today** 46(6), 56 y referencias en ese trabajo.
- [2] WOJS, A.; L. JACAK and P. HAWRYLAK (1998): **Quantum Dots**, Springer-Verlag.
- [3] Proceedings of the International Conference on Quantum Dots (2004): **Physica E** 26.
- [4] STRENZ, R. **et al.** (1994): **Lett.** 73, 3022.
- [5] SCHÜLLER, C. **et al.** (1996): **Phys. Rev. B** 54, 17304.
- [6] LOCKWOOD, D.J. **et al.** (1996): **Phys. Rev. Lett.** 77, 354.
- [7] SCHÜLLER, C. **et al.** (1998): **Phys. Rev. Lett.** 80, 2673.
- [8] SOTOMAYOR-TORRES, C.M.; D.J. LOCKWOOD, and P.D. WANG (2000): **J. Electron. Mater.** 29, 576.
- [9] SHULLER, C.; C. STEINEBACH and D. HEITMANN (2001): **Solid State Communications** 119 323.
- [10] BROCKE, T. **et al.** (2003): **Phys. Rev. Lett.** 91, 257401.
- [11] PASCUAL, C. **et al.**, cond-mat/0506143.
- [12] GONZÁLEZ, A. and A. DELGADO (2005): **Physica E** 27, 5.
- [13] REBOREDO, F.A. and C.R. PROETTO (1994): **Phys. Rev. B** 50 15174.
- [14] BARATAAS, A.; A.G. MALSHOKOV; V. GUDMUNDSSON and K.A. CHAO (1996): **J. Phys. Cond. Matt.** 8 L325.
- [15] STEINEBACH, C. **et al.** (1998): **Phys. Rev. B** 57 1703.
- [16] SERRA, LI. **et al.** (1999): **Phys. Rev. B** 59 15290.
- [17] LIPPARINI, E. (1999): **Phys. Rev. B** 60 8734.
- [18] DELGADO, A.; A. GONZÁLEZ and D.J. LOCKWOOD (2004): **Phys. Rev. B** 69, 155314.
- [19] DELGADO, A.; A. GONZÁLEZ and E. MENÉNDEZ-PROUPIN (2002): **Phys. Rev. B** 65 155306 .
- [20] LOUDON, R. (1964): **Adv. Phys.** 13, 423.
- [21] RING, P. and P. SCHUCK (1980): *The Nuclear Many-Body Problem* (Springer-Verlag, New-York).
- [22] DELGADO, A.; A. GONZÁLEZ and D.J. LOCKWOOD (2005): **Solid State Communications** 135, 554-562.
- [23] \_\_\_\_\_ (2005): **Phys. Rev. B** 71, 241311(R).
- [24] ODRIAZOLA, A.; A. DELGADO y A. GONZÁLEZ: Trabajo en curso.

# ESTIMACIÓN DE MODOS TEM<sub>mn</sub> LAGUERRO-GAUSSIANOS

Omel Mendoza Yero, Laboratorio de Tecnología Láser-ICTM, Universidad de La Habana,  
Ciudad de La Habana, Cuba

## RESUMEN

Mediante la modelación de la potencia/energía transmitida a través de una abertura circular de radio variable por haces láser laguerro-gaussianos, se propone un método para la estimación de sus modos transversales. El método es válido para haces láser compuestos por un único modo de oscilación transversal laguerro-gaussiano. Se muestra que las posiciones de los ceros radiales en la función de irradiancia de cada haz láser coinciden con puntos de pendiente nula en la curva teórica correspondiente a la potencia de la radiación transmitida en función del radio de la abertura circular. Los resultados son mostrados en forma de gráficos ilustrativos, con ayuda de herramientas de cálculo proporcionadas por el programa "MAPLE". Este trabajo encuentra aplicación en el campo de la caracterización de haces láser, particularmente útil para la obtención de información sobre haces láser con simetría cilíndrica y anchos de emisión fuera del espectro visible.

## ABSTRACT

By mapping the power/energy of Laguerre-Gaussian beams transmitted through a circular aperture of variable radius, a method to estimate their transversal electromagnetic modes is introduced. This method holds for laser beams made up by a unique transversal Laguerre-Gaussian mode. It is shown that positions of radial zeros in the irradiance function of each laser beam coincide with points of zero slopes in the theoretical curve corresponding to power transmitted vs. radius of the circular aperture. With the help of tools given by "Maple" software, results are showed by using several graphics. This work finds application in the field of laser beam characterization, particularly useful to collect information concerning laser beams with cylindrical symmetry and emission bandwidths out of the visible spectrum.

## 1. INTRODUCCIÓN

La distribución espacial de energía de un haz láser en un plano perpendicular a su dirección de propagación puede describirse a través de su composición de modos transversales TEM<sub>mn</sub>. Los subíndices m y n indican el número de ceros de la función de irradiancia en direcciones bien definidas.

La determinación de la estructura de modos de los haces láser influye en la selección de los mismos para una función específica. Esta tarea puede ser muy compleja, y en ocasiones, de difícil solución si se tiene en cuenta que los generadores láser pueden emitir haces láser complejos, compuestos por varios modos transversales TEM<sub>mn</sub> al mismo tiempo [1].

Usualmente, se emplean métodos para la determinación de modos transversales basados en: a) impresiones sobre un papel fotosensible del perfil de irradiancia de haces láser de bajas potencias (decenas de W) o bajas energías de pulsos (decenas de mJ) [2], b) la observación de distribuciones espaciales de energía debidas a interacciones de haces láser con materiales apropiados [3], etc. En estos casos, la clasificación de los modos TEM<sub>mn</sub> presentes en el haz láser depende de criterios

subjetivos y/o de las propiedades del material que se emplea en la visualización del modo transversal presente en el mismo.

En este trabajo se introduce un método sencillo para la estimación de modos transversales TEM<sub>mn</sub> laguerro-gaussianos (LGs) a partir de mediciones de potencia transmitida a través de un orificio circular de radio variable. El método es aplicable a haces láser que posean un único modo de oscilación transversal laguerro-gaussiano (LG). En el epígrafe 3 se discuten las condiciones de validez del método propuesto.

## 2. ESTIMACIÓN DE MODOS TEM<sub>mn</sub> LAGUERRO-GAUSSIANOS

Se considera un sistema de coordenadas cilíndricas  $r, \varphi, z$  cuyo eje  $z$  coincide con la dirección de propagación de cierta perturbación electromagnética generada por un láser. En general, soluciones de la ecuación de onda paraxial [2] en la forma  $u_{mn} = (r, \varphi, z)$  nos permiten modelar el campo eléctrico de dicha perturbación, representando al mismo tiempo, el modo de oscilación transversal TEM<sub>mn</sub> del haz láser correspondiente. Si se coloca una abertura circular de radio  $a$  en un plano perpendicular a la dirección de propagación de un haz láser, centrada respecto al eje  $z$  del sistema de

coordenadas, la potencia normalizada  $P_{mn} = (z, a)$  transmitida a través de la misma viene dada por la siguiente expresión

$$P_{mn}(z, a) = \int_0^{2\pi a} \int_0^a |u_{mn}(r, \varphi, z)|^2 r dr d\varphi \quad (1)$$

De acuerdo con la expresión (1), la fracción de potencia transmitida  $P_m(z, a)$ , depende del radio  $a$  del orificio circular y de los subíndices  $m$  y  $n$  correspondientes al modo  $TEM_{mn}$  del haz láser. Además, de las bases teóricas que describen la propagación de los haces láser [2] - [4], se conoce que si un haz láser posee un único modo  $TEM_{mn}$ , éste se mantiene invariante de forma durante su propagación por el espacio libre. Por tanto, las funciones  $P_m(z, a)$  son características de cada modo  $TEM_{mn}$  puro y, en principio, deben proporcionar información que permita determinar el modo  $TEM_{mn}$  presente en un haz láser.

En la Figura 1 se muestran distribuciones espaciales de energía de modos  $TEM_{mn}$  LGs sobre planos perpendiculares a sus direcciones de propagación. Note que el subíndice  $m$  indica el número de ceros radiales presentes en el perfil de irradiancia, mientras que el subíndice  $n$  cuantifica los ceros angulares.

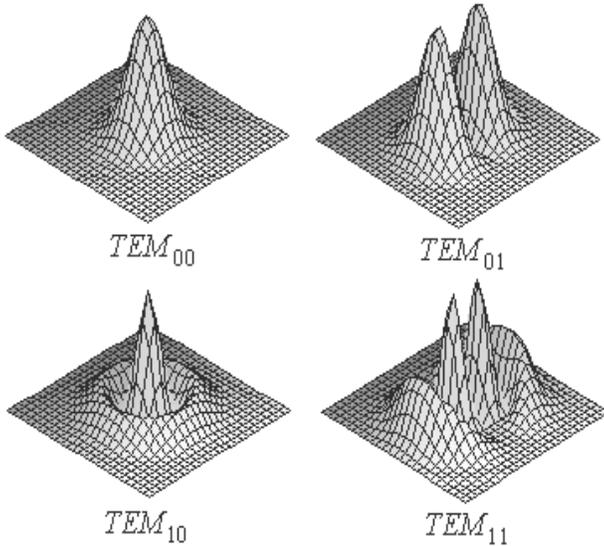


Figura 1. Perfiles de irradiancia LGs.

Las funciones  $u_{mn}(r, \varphi, z)$  (soluciones de la ecuación de onda paraxial) que representan el campo electromagnético de los modos  $TEM_{mn}$  LGs vienen expresadas a través de la multiplicación de funciones exponenciales (curvas gaussianas) por polinomios de laguerre generalizados  $L_n^m(\dots)$  [2], [3].

En particular, la irradiancia  $|u_{mn}(r, \varphi, z)|^2$  de un haz láser con modo  $TEM_{mn}$  LG, en un plano  $z$  perpendicular a su dirección de propagación, puede escribirse en la forma

$$|u_{mn}(r, \varphi, z)|^2 = \frac{m!}{\pi(1+\delta)(m+n)!} \left(\frac{2w_0}{w(z)}\right)^2 \left(\frac{2r^2}{w(z)^2}\right)^n \left(L_n^m\left(\frac{2r^2}{w(z)^2}\right)\right)^2 \exp\left(-2\frac{r^2}{w(z)^2}\right) \cos(n\varphi)^2 \quad (2)$$

En la expresión (2)  $w(z)$  y  $w_0$  representan los radios del haz láser según el criterio  $1/e^2$  [3], en un plano  $z = \text{constante}$  y en su parte más estrecha, respectivamente. A su vez,  $\delta$  toma valor uno si se trata del modo  $TEM_{00}$  y cero para cualquier otro modo.

Resulta difícil encontrar una solución analítica general para  $P_{mn}(z, a)$ . Sin embargo, con ayuda de un programa de cálculo adecuado es posible, asignando valores numéricos al radio del haz  $w(z)$ , simular el comportamiento de cada modo  $TEM_{mn}$  por separado. En particular, para el modo  $TEM_{20}$  LG y  $w(z) = 1$ , con el programa denominado "MAPLE", se puede obtener una curva como la mostrada en la Figura 2.

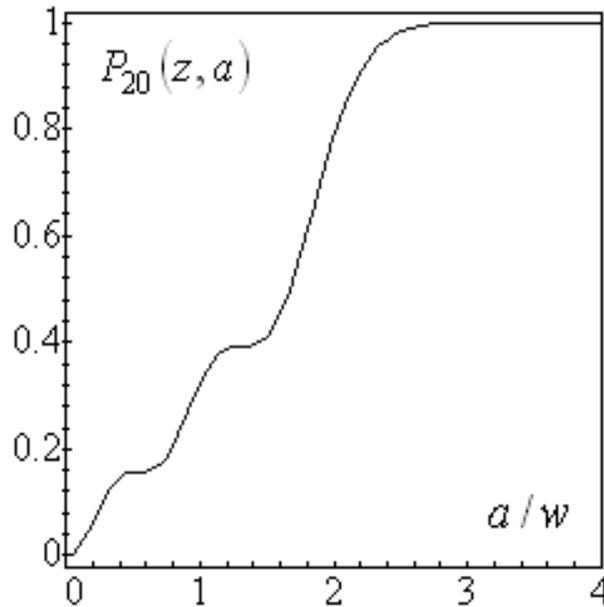
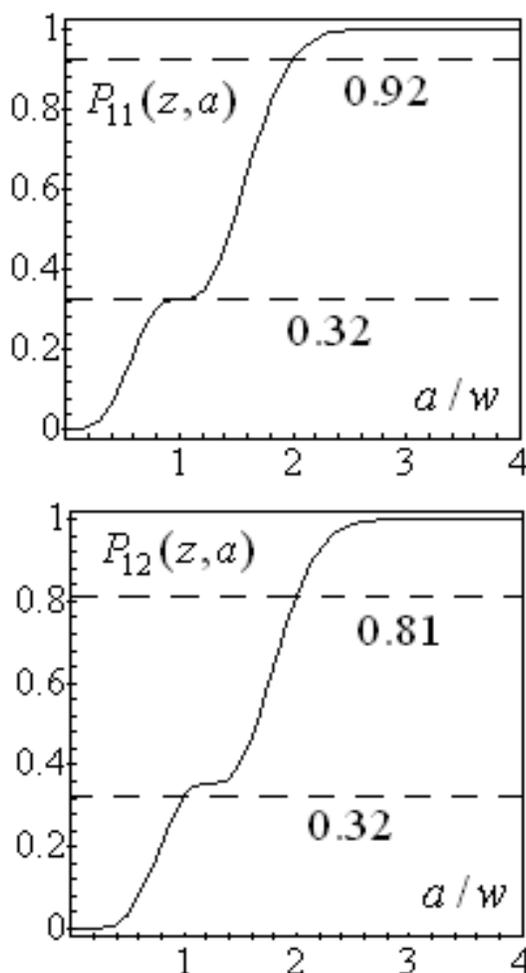


Figura 2. Potencia transmitida. Modo  $TEM_{20}$  LG.

Después de una simple inspección de la Figura 2, parece claro que el número de ceros radiales del perfil de irradiancia es igual al número de "pasos" (cambios bruscos de inflexión) presentes en esta curva. Puede comprobarse que, en general, este comportamiento se cumple para todos los modos  $TEM_{mn}$  LGs, permitiendo así la asignación a cada modo del subíndice  $m$  a partir de las curvas de  $P_{mn}(z, a)$  v.s  $a/w$ . Además, las posiciones  $r = \eta w$  ( $\eta$  denota un número positivo) de los ceros radiales de las funciones  $|u_{mn}(r, \varphi, z)|^2$  coinciden con las posiciones  $a = \eta w$  en las curvas de potencia transmitida donde  $dP_{mn}(z, a)/da = 0$ .

Con el objetivo de identificar un modo  $TEM_{mn}$  LG se necesita además del conocimiento del número de ceros angulares (subíndice  $n$ ) presentes en su perfil de irradiancia. Esta información puede obtenerse también de las curvas de potencia transmitida, ya que las mismas poseen características únicas de cada modo  $TEM_{mn}$  LG. Con este fin, se utilizan los valores de la fracción de potencia  $P_{mn}(z,a)$  para los radios de la abertura circular  $a = w$  y  $a = 2w$ , respectivamente. Dichas fracciones de potencia pueden localizarse fácilmente en las curvas de  $P_{mn}(z,a)$  v.s  $a/w$ . Concretamente, el método para estimar el valor del subíndice  $n$  consiste en determinar, a partir de dichas curvas, los pares de valores  $P_{mn}(z,a = w)$ ,  $P_{mn}(z,a = 2w)$  (típicos de cada modo  $TEM_{mn}$  LG) correspondientes a las razones  $a/w = 1$  y  $a/w = 2$ , respectivamente.

En la Figura 3 se representan, con líneas discontinuas, las fracciones de potencias  $[P_{11}(z,w), P_{11}(z, 2w)]$  y  $[P_{12}(z, w), P_{12}(z, 2w)]$  en el caso de los modos  $TEM_{11}$  y  $TEM_{12}$  LGs y  $w = 1$ . Note que para cada modo  $TEM_{mn}$  LG, existe un par diferente de fracciones de potencia transmitida, aún cuando el número de ceros radiales pueda ser el mismo.



**Figura 3.** Fracciones de potencia transmitida. Modos  $TEM_{11}$  y  $TEM_{12}$  y LGs.

Cuando una de estas fracciones de potencia coincida para diferentes modos  $TEM_{mn}$  LGs (por ejemplo:  $P_{11}(z, w) = P_{12}(z, w) \cong 0.32$  en la Figura 3), la otra permite discernir de qué subíndice  $n$  se trata (note en este caso que,  $P_{11}(z, 2w) \neq P_{12}(z, 2w)$ ). Además, siempre existe la posibilidad de tomar un tercer valor de potencia transmitida (por ejemplo:  $P_{mn}(z,a = 1/2w)$ ,  $P_{mn}(z,a = 3/2 w)$ , etc.) en caso necesario.

### 3. CONDICIONES DE VALIDEZ DEL MÉTODO PROPUESTO

El método presentado para la estimación de modos  $TEM_{mn}$  LGs se basa en un procedimiento similar al utilizado en el método de apertura variable, descrito por la norma ISO/TR 11146 [5], para medir el radio de haces láser. En correspondencia con lo anterior, el detector empleado debe satisfacer los requerimientos expuestos en la sección 4.2.2 de dicha norma. En particular, su área sensible tiene que escogerse de forma tal que más del 99 % de la potencia transmitida sea capturada por el detector. Por esta razón, los efectos de difracción debidos al truncado del haz por la abertura circular, si bien originan ondulaciones en los perfiles de irradiancia en campo cercano [3] no influyen en las mediciones de potencia transmitida, sí la selección del área sensible del detector es adecuada.

Adicionalmente, es recomendable que, al igual que en muchos de los sistemas comerciales analizadores de haces láser [6], se coloque el detector en la zona de campo cercano ( $z < \pi a^2/\lambda$ ) donde  $\lambda$  es la longitud de onda media en el vacío de la radiación láser [3]. Esto se debe a que los patrones de difracción de Fresnel para aberturas circulares se encuentran espacialmente más localizados que sus correspondientes patrones entendidos de Fraunhofer [3], [4], [7]-[9].

De esta forma, en la zona de campo cercano, el área sensible del detector podría estimarse a partir del área limitada por la sombra geométrica de la abertura circular sobre la superficie sensible del detector. Teniendo en cuenta la extensión de dicha sombra geométrica debida al radio de curvatura del frente de onda [2]-[4] de los modos  $TEM_{mn}$  LGs, el área del detector  $A_D$ , puede calcularse aproximadamente de la siguiente expresión [6]:

$$A_D \cong \pi a^2 \left( 1 + \frac{d}{z_0 + \frac{z_R^2}{z_0}} \right)^2 \quad (3)$$

En la expresión (3)  $z_R = \pi w_0^2/\lambda$ , es la distancia de Rayleigh del haz láser,  $d$  es la distancia entre el

plano de la abertura circular y el plano que localiza la superficie del detector y  $z_0$  es la distancia entre el plano donde se localiza el mínimo radio  $w_0$  del haz láser y el plano de la abertura circular.

#### 4. CONCLUSIONES

Se introdujo un método sencillo para estimar el valor de los subíndices  $m$ ,  $n$  correspondientes a los modos  $TEM_{mn}$  laguerro-gaussianos presentes en haces láser. La información necesaria para la determinación de un modo  $TEM_{mn}$  laguerro-gaussiano se obtuvo de las curvas de potencia de la radiación láser transmitida a través de un orificio circular de radio variable. De esta forma, se evitan errores de apreciación visual que suelen aparecer como conse-

cuencia del uso de otros procedimientos (por ejemplo: mediante la comparación de marcas dejadas por un haz láser sobre papel fotosensible con perfiles teóricos de irradiancia conocidos).

Las condiciones de validez del método propuesto son similares a las establecidas en la norma ISO/TR 11146 [5] para el método de abertura variable (ver epígrafe 4.2 de esta norma). Además, como la potencia total transmitida a través de una abertura circular de radio  $a$  depende de la distribución de energía del haz en el plano de la abertura, no es posible discernir de las curvas de  $P_{mn}(z, a)$  v.s  $a/w$  si se trata de un modo puro  $TEM_{mn}$  laguerro-gaussiano o de una combinación de modos que den lugar a la misma distribución espacial de energía.

#### REFERENCIAS

1. DISO, D.; M.R. PERRONE and M. L. PROTOPAPA (1999): **Opt. & Laser Technol.** 31, 411.
2. KOGELNIK, H. y T. LI (1966): **Appl. Opt.** 5, 1550.
3. SIEGMAN, A.E. (1986): **Lasers**, University Science Book, Mill Valley, California.
4. SVELTO, O. (1998): **Principles of Lasers**, London: Plenum Press, New York.
5. International Standard ISO/TR 11146-3 (2004): **Technical Report**, Switzerland, First edition, ISO.
6. JOHNSTON, T.F. and J.M. FLEISCHER (1996): **Appl. Opt.** 35, 1719.
7. GOODMAN, J. W. (1996): **Introduction to Fourier Optics**, The McGraw-Hill Companies, New York.
8. HECHT, E. (1987): **Optics**, Addison-Wesley, Menlo Park, California.
9. BORN, M. and E. WOLF (1980): **Principles of Optics**, Pergamon Press, Great Britain.

# OPTIMIZACIÓN DEL DETECTOR DE LATIDO FETAL: EMPLEO DEL MICROCONTROLADOR DS87C550

Antonio Jiménez Cañas, Departamento de Ultrasonica, Instituto de Cibernética, Matemática y Física (ICIMAF), Ciudad de La Habana, Cuba

## RESUMEN

Se exponen los criterios de adecuación y procesamiento de la señal Doppler ultrasónica empleados en el Detector de Latido Fetal, partiendo de la utilización del microcontrolador D87C51. Igualmente se muestran las características del algoritmo optimizado de autocorrelación utilizado y los resultados obtenidos. Adicionalmente y en aras de reducir consumo, volumen y recursos materiales, se analizan las ventajas que se derivan del empleo del microcontrolador DS87C550 de nueva generación, el cual es totalmente compatible con el anterior e introduce nuevas posibilidades en el diseño.

**Palabras clave:** simulación, autocorrelación, procesamiento de señales, microcontrolador.

## ABSTRACT

Adaptation criteria and processing of Ultrasonic Doppler signal employed in Fetal Heart Detector are presented, starting from the use of D87C51 microcontroller. In the same way, the characteristics of the autocorrelation algorithm optimized as well as the obtained results are shown. Additionally, for the sake of reducing consumption, volume and material resources, the advantages derived of the employment of the new generation DS87C550 microcontroller are considered. It's completely compatible with the previously used and introduces new possibilities in the design.

**Key words:** simulation, autocorrelation, signal processing, microcontroller

## INTRODUCCIÓN

La detección del latido fetal constituye una práctica de gran importancia en el diagnóstico obstétrico, ya que permite examinar una serie de anomalías y comportamientos atípicos que se presentan durante el proceso de gestación [2] [3], y en este sentido, la introducción de los Detectores Fetales electrónicos constituye una contribución importante. El principio de funcionamiento de estos equipos se basa en el empleo del efecto Doppler Ultrasónico [4] [1], a partir de la interacción del ultrasonido con los músculos, las paredes, las válvulas y la sangre en el corazón fetal, lo que permite finalmente evaluar el estado clínico del feto con la obtención del sonido y la frecuencia cardíaca fetal. El valor de la frecuencia fetal se expresa en latidos por minuto (LPM). Esta información puede obtenerse utilizando métodos de procesamiento de señal como la conocida Autocorrelación en Tiempo Real [10] [11] [8] [9], pudiendo ser implementados en una PC, o bien, empleando un microcontrolador.

El Detector Fetal se considera un equipo de diagnóstico médico por ultrasonido Doppler que emplea la técnica de transmisión continua CW no direccional. Esto significa que se transmite y se recibe el ultrasonido de forma continua y simultánea, pudiéndose detectar solamente cambios de amplitud, de sentido y de frecuencia. Para esto se

emplea dentro del sensor un elemento cerámico transmisor y otro elemento cerámico receptor, que operan de forma independiente.

La señal Doppler fetal está compuesta por componentes espectrales que se encuentran distribuidas en tres zonas del espectro respecto a una señal portadora, que en nuestro caso es de 2MHz: las producidas por el movimiento de las paredes del corazón o de sus músculos, comprendidas entre los  $\pm 200$  y los  $\pm 400$  Hz, las producidas por las válvulas del corazón, comprendidas entre los  $\pm 400$  y los  $\pm 1000$  Hz y las producidas por el flujo arterial sanguíneo, que están comprendidas entre los  $\pm 1000$  y los  $\pm 2000$  Hz.

No obstante, aparecen otras componentes espurias que se originan a partir del flujo venoso, del flujo amniótico, del cambio del ángulo de incidencia del haz ultrasónico, por el movimiento del feto, etc. Estas componentes moduladoras varían aleatoriamente en sentido general, pudiendo predominar en ocasiones la señal originada en las paredes, otras veces la originada en las válvulas, etc. Se afirma que esta señal Doppler fetal presenta un comportamiento de tipo ciclo estacionario [12]. Constituye una señal modulada en amplitud de 2 MHz. cuya componente moduladora posee una forma de onda pulsátil y periódica, estando constituida por las componentes

espectrales anteriormente mencionadas. Entonces, si se aplicara detección por envolvente, se trasladarían a banda base todas las componentes de frecuencia anteriormente mencionadas, ruidosas y no ruidosas, obteniéndose como resultado una señal con compo-tes espectrales comprendidas en el margen audible, conteniendo además la información de baja frecuencia correspondiente a la pulsación cardíaco fetal.

Ahora bien, esta primera señal demodulada (que posee la misma forma de onda de la envolvente moduladora anteriormente mencionada) constituye también una señal pulsada, y por ende, modulada en amplitud, donde la frecuencia de la nueva señal portadora varía en un intervalo comprendido entre los 100 y los 900 HZ, aproximadamente, constituyendo propiamente el sonido del latido cardíaco fetal que se escucha a través de la bocina del equipo. Su envolvente la constituye la señal cardíaco fetal, cuyo margen de frecuencia está comprendido desde 1 hasta 3.5 Hz (60-210 ppm). La portadora de frecuencia variable se compone de grupos de varios ciclos completos que van desde 5 hasta 30 ciclos, aproximadamente.

Consecuentemente, si se aplica nuevamente demodulación de amplitud a dicha señal portadora, se trasladan a banda base las componentes de menor frecuencia, contenidas en su envolvente, obteniéndose la información cardíaco fetal de baja frecuencia junto con todas las demás componentes espurias, anteriormente mencionadas. Si además se realiza un filtrado pasabajo adecuado, donde se atenúen las señales indeseables (ruidosas), se obtiene una señal donde predominan las componentes de más baja frecuencia (componentes provenientes de las paredes, músculos y válvulas del corazón fetal).

No obstante, después de haber obtenido estas componentes de baja frecuencia, se hace necesario aplicar el método de procesamiento de señal conocido como Auto-correlación en Tiempo Real, que permite obtener la información de la frecuencia cardíaco fetal de forma continua y confiable.

### CARACTERÍSTICAS DEL EQUIPO

Según la Figura 1, el oscilador maestro genera una señal sinusoidal con una frecuencia de 2 MHz ± 100 KHz, la cual es amplificada a 2 Volt pp y enviada a un sensor piezo-cerámico. Este último convierte la señal eléctrica en una señal acústica, que viaja por los tejidos del abdomen de la embarazada hasta llegar al corazón del feto, donde es reflejada, y parte de ella recibida por

dicho sensor. Este sensor es capaz de transformar ahora la señal acústica en una señal eléctrica S(t), la cual contiene información relacionada con la intensidad y la frecuencia del ultrasonido reflejado en dicho corazón fetal:

$$S(t) = A \cos(\omega_0 t + \phi_c) + B_f \cos(\omega_0 t + \omega_f t + \phi_f) + B_r \cos(\omega_0 t - \omega_r t + \phi_r) \quad (1)$$

donde:

- B<sub>f</sub> y B<sub>r</sub> son las amplitudes de las componentes de movimiento directas e inversas respectivamente.
- A amplitud de la señal del oscilador.
- ω<sub>0</sub> frecuencia del oscilador.
- ω<sub>f</sub> y ω<sub>r</sub> frecuencias de corrimiento Doppler (directa e inversa).
- φ<sub>c</sub>, φ<sub>f</sub> y φ<sub>r</sub> fases iniciales de las señales respectivas.

A continuación, dicha señal se amplifica en 39 dB y se demodula en amplitud en el detector del receptor, al cual le llega también una pequeña porción de la señal del oscilador maestro, como resultado de la inducción electromecánica en el sensor. Esta forma de demodulación elimina la señal de 2 MHz. del oscilador y provoca el solapamiento de las bandas laterales en banda base, quedando finalmente la señal que responde al término B cos(ω<sub>d</sub> + φ<sub>d</sub>). Seguidamente, esta señal demodulada se pre-amplifica en 25 dB y se filtra mediante un filtro pasabajo. A partir de este punto, la señal audible obtenida puede ser debidamente procesada y/o amplificada en potencia para su escucha.

### Etapa de Preprocesamiento

Los circuitos empleados en el preprocesamiento de la señal Doppler audible son los encargados de adecuarla convenientemente para entregársela

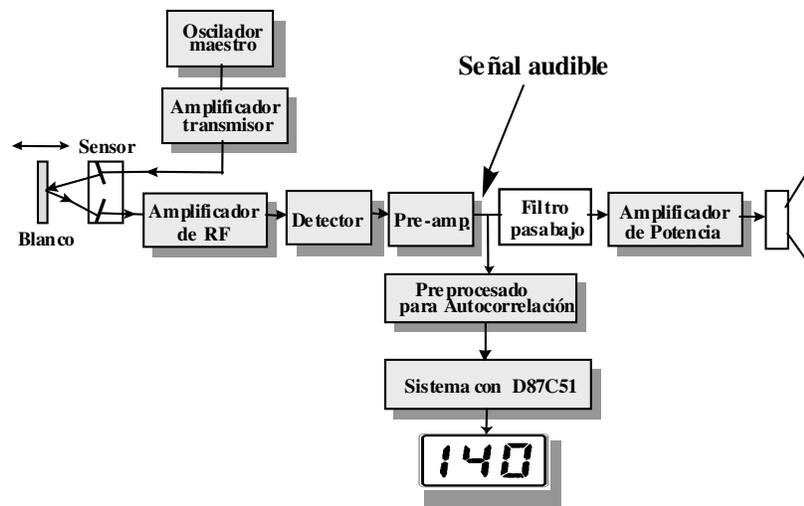


Figura 1. Esquema de bloques de un Detector de Latido Fetal.

posteriormente a la etapa de procesamiento. Existen varias formas de preprocesar la señal Doppler, según se ilustra en las referencias [7] y [11]. La implementación circuital mostrada en este trabajo es completamente original.

En la Figura 2 se muestran las señales en cada punto de referencia. La señal Doppler audible, proveniente del detector de amplitud del equipo, constituye la portadora de una señal modulada en amplitud cuya frecuencia varía en un intervalo comprendido entre los 100 y los 900 HZ y su envolvente es de tipo pulsatoria.

A continuación, la señal se amplifica debidamente y se comprime en amplitud, a manera de control automático de ganancia (CAG), para un margen del voltaje de entrada de más de dos décadas.

Entonces, esta señal comprimida se amplifica, se rectifica en onda completa y se introduce en un filtro pasabajo con una frecuencia de corte de 0.6 Hz a -3 dB y una pendiente de -6 dB/octava. Finalmente esta se amplifica nuevamente.

Como resultado, la envolvente pulsátil obtenida (forma de onda 3), posee un margen de frecuencias comprendido entre 1 y 3.5 Hz (60-210 LPM). A continuación se vuelve a rectificar la señal en onda completa para desacoplar la componente de corriente directa, se aplica filtrado pasabajo donde se atenúan estas componentes espurias (ruidosas) y finalmente, se amplifica y conforma la señal resultante donde predominan las componentes de más baja frecuencia que van a ser enviadas al procesador (forma de onda 5).

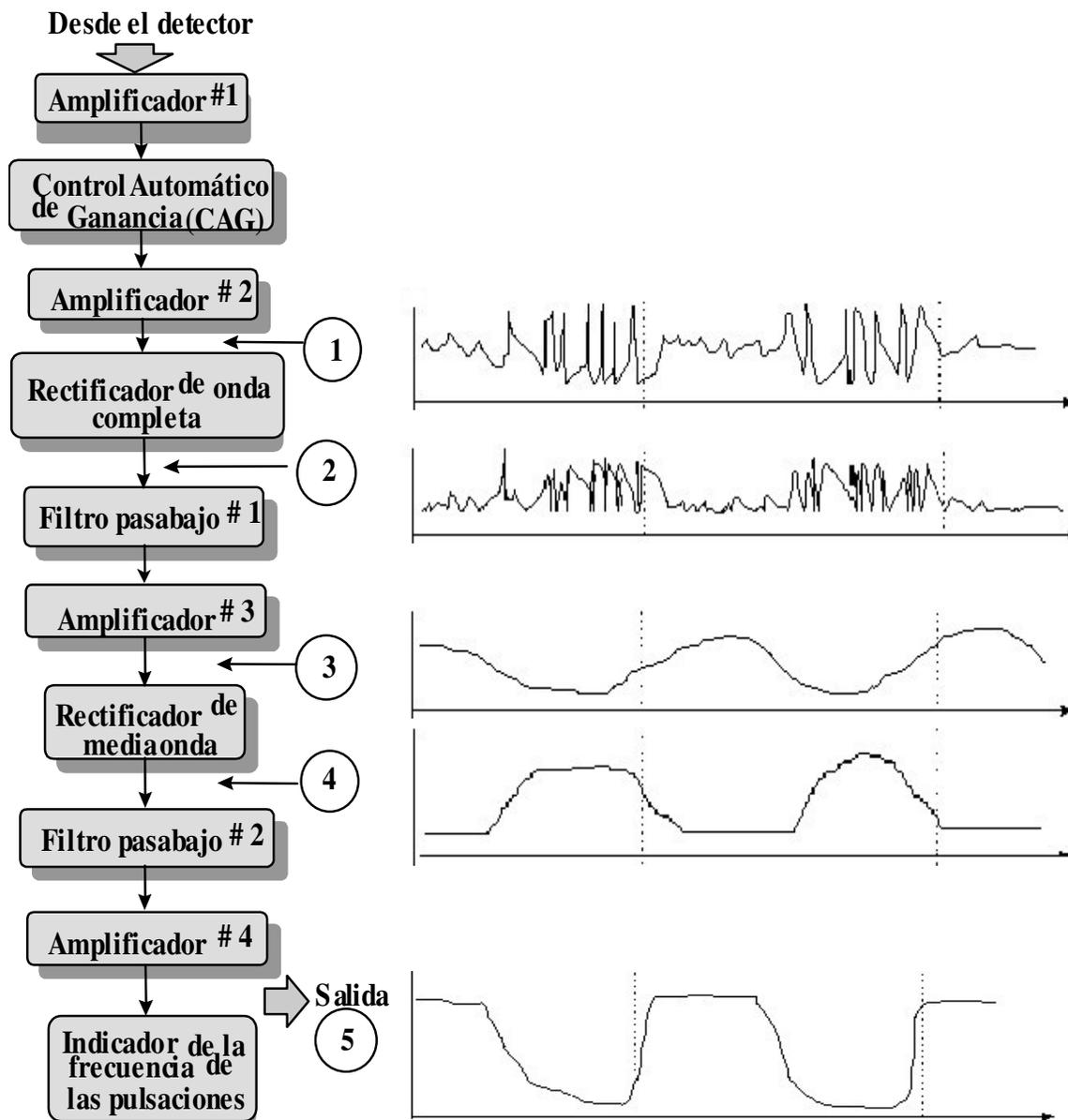


Figura 2. Esquema de bloques que representa la estructura del preprocesamiento empleado.

## Etapa de Procesamiento

En la Figura 3 se muestra una configuración circuital típica de adquisición y procesamiento de señal. El microcontrolador D87C51, una memoria RAM externa, lógica de interfase y un convertor A/D de 8 bit como periféricos fundamentales. El convertor A/D digitaliza la señal proveniente del preprocesador y se la entrega al microcontrolador.

Consecuentemente, se implementó el algoritmo de la Autocorrelación en Tiempo Real en lenguaje ensamblador del D87C51 [6]. Dicho algoritmo se compone de un programa principal y 10 subrutinas según se expone en la referencia [5]. Se empleó una frecuencia de muestreo de 200 Hz (una muestra cada 5 mseg.), como valor de compromiso entre velocidad de procesamiento en el microcontrolador y cantidad de puntos en la curva de autocorrelación, factor que determina la resolución del sistema.

Por ese motivo, cuando se calcula el valor de la frecuencia (o el periodo) fetal, a partir del valor máximo absoluto, obtenido de la evaluación de los valores discretos de autocorrelación, previamente calculados, existirá un error absoluto de lectura entre este valor calculado y el valor real, el cual está en dependencia del valor de frecuencia fetal a medir. En la Figura 5 se observa que el valor del error absoluto de lectura cambia no linealmente con la frecuencia de la señal. Por ejemplo: desde 60 hasta 117 ppm. el error introducido es nulo, desde 115 hasta 155 el error oscila entre 1 y 2 valores, desde 155 hasta 195 el error puede llegar hasta 3 valores y finalmente, por encima de 195 el error puede llegar hasta 4 valores.

Por otro lado, se escogió la máxima frecuencia de reloj permisible para el microcontrolador de 12 MHz y se seleccionó un factor de ponderación  $N = 8$  como valor óptimo, de acuerdo con los resultados obtenidos en las pruebas de simulación realizadas [6]. El tiempo de cálculo de una autocorrelación circular elemental, cuyo algoritmo se muestra en la figura 4, fue de solo 33  $\mu$ seg. Seguidamente, este resultado se deposita en uno de los 150 registros ubicados en memoria RAM. Los cálculos de la curva de autocorrelación se efectúan como si existieran realmente 200 registros, más se emplean solo 150 en aras de agilizar el proceso y tomando en consideración que los primeros 25 registros están asociados a valores mayores que 210 ppm los cuales no son de utilidad durante el diagnóstico

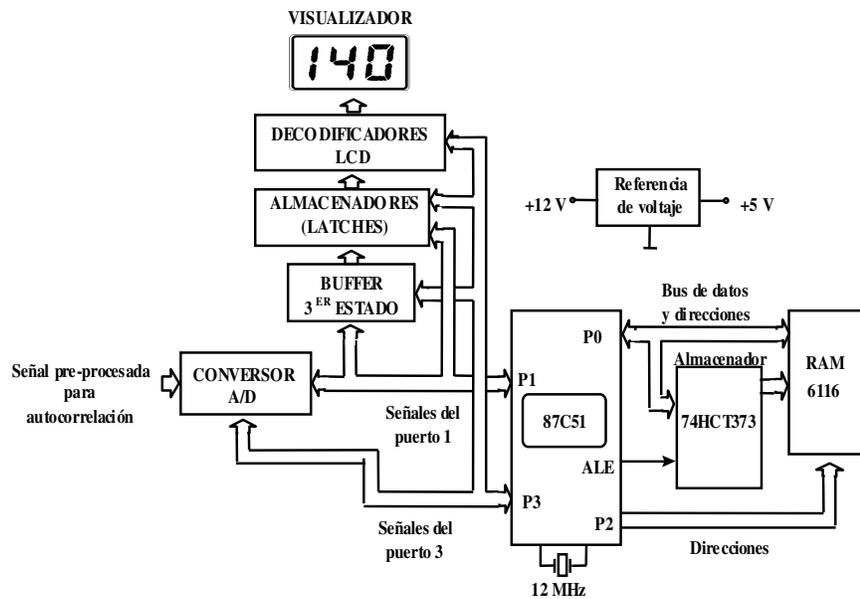


Figura 3. Esquema de bloques de la etapa de Procesamiento.

obstétrico. La actualización de estos 150 registros constituye una ronda de refrescamiento y actualización, efectuándose 240 rondas por cada lectura mostrada en el visualizador.

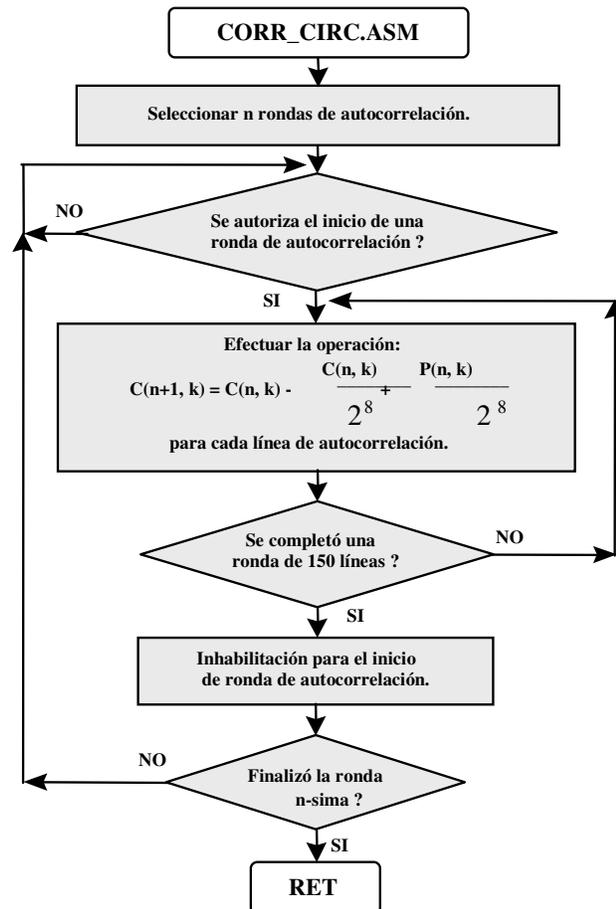
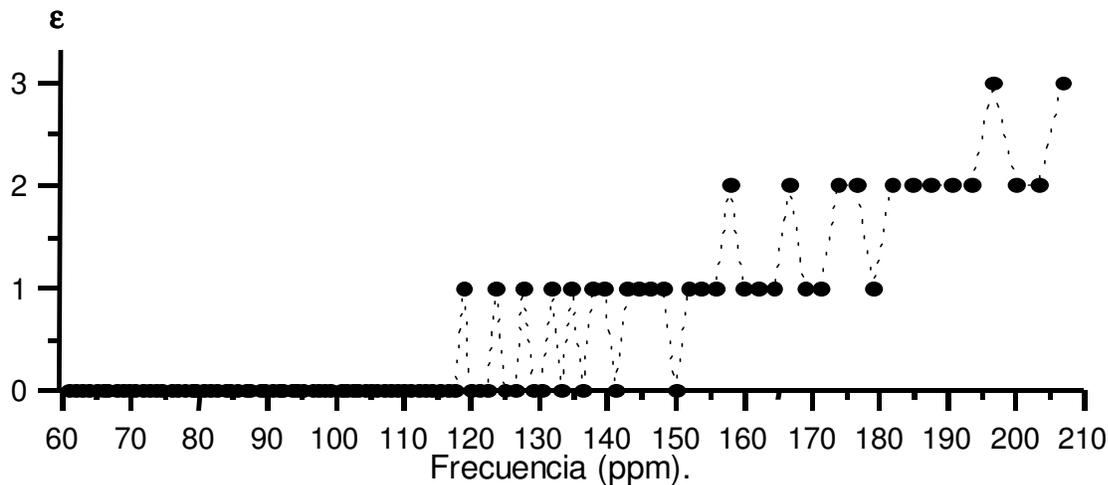


Figura 4. Algoritmo de la autocorrelación circular elemental.



**Figura 5.** Representación del valor absoluto de lectura en función del valor de frecuencia fetal calculada.

Como resultado, se obtuvo una lectura de frecuencia fetal cada 1.21 seg., mostrada en un visualizador LCD. El margen de medición se extiende desde 60 hasta 210 ppm., el cual está comprendido dentro del intervalo de medición de equipos comerciales análogos.

**Resultados obtenidos en la implementación con D87C51**

- Se obtienen lecturas de la frecuencia fetal cada 1.21 seg.
- Lecturas con un grado satisfactorio de coincidencia respecto a las obtenidas con equipos comerciales análogos.
- Implementación relativamente costosa por emplear varios circuitos integrados.
- Software de mediana complejidad.
- Sistema implementado con un valor constante del factor de ponderación  $N = 8$ , lo que garantiza un tiempo de cálculo constante para la auto-correlación circular elemental.
- Insuficiente velocidad de procesamiento del micro-controlador lo que impide elevar el valor de la frecuencia de muestreo.
- El empleo de dispositivos integrados periféricos aumenta el consumo eléctrico y el tamaño del circuito impreso.

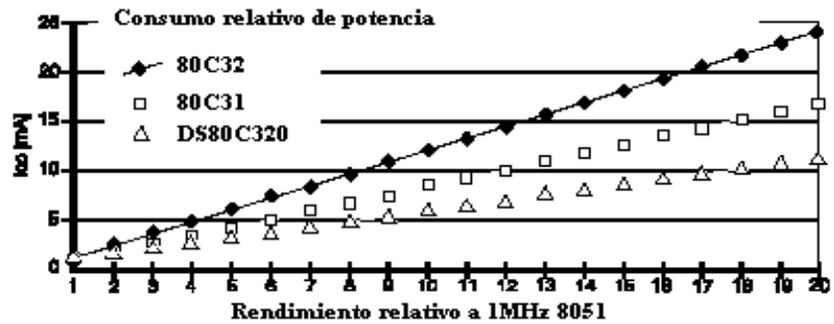
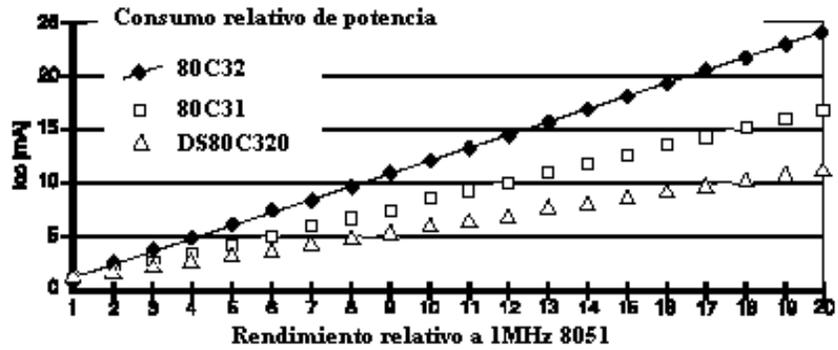
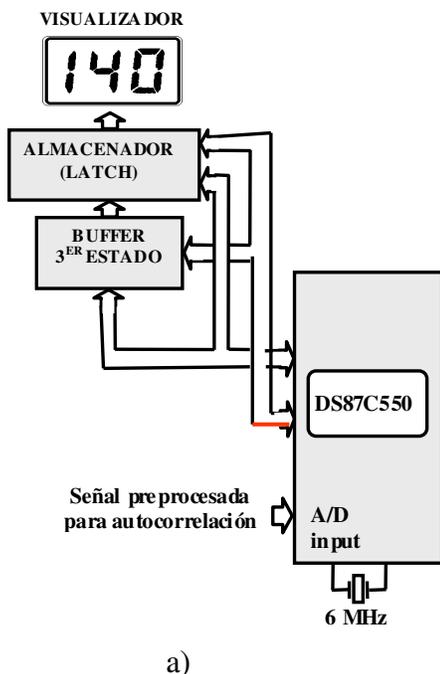
**EMPLEO DEL MICROCONTROLADOR DS87C550 DE MAXIM-DALLAS**

El microcontrolador DS87C550 de Maxim-Dallas es un dispositivo de nueva generación y de alta velocidad, el cual es totalmente 8751 compatible

incluyendo el juego de instrucciones y el compilador. Es 2.5 veces más veloz como promedio, respecto al tiempo de ejecución de todo el juego de instrucciones del 8751 típico, para iguales frecuencias de reloj, lo cual implica que operando a más baja frecuencia puede obtenerse el mismo desempeño en igual intervalo de tiempo, con menor consumo de corriente. Puede operar hasta 33 MHz, lo cual permite incrementar la velocidad de ejecución de una ronda elemental de autocorrelación en unas cuatro veces. Posee además memoria ROM reconfigurable, convertidor A/D con tiempo de conversión de 20 mseg. y varios modos de trabajo que facilitan el bajo consumo.

**Ventajas que introduce en el Detector Fetal Optimizado**

- Simplifica la arquitectura de la etapa de procesamiento con la presencia del conversor A/D y las memorias internas RAM, lo que ayuda a reducir el costo y el volumen del equipo, ver Figura 6.
- El factor de aumento en la velocidad de operación respecto al D87C51 es 2,12, lo que permite emplear la mitad de la frecuencia de reloj y por ende, reducir el consumo de corriente a la mitad.
- Mayor facilidad en el manejo de dispositivos periféricos y en la implementación de nuevas posibilidades de operación, ya que posee 6 puertos bidireccionales y seis niveles de interrupciones externas.
- El empleo del watchdog facilita el autoapagado.
- Los modos de trabajo de bajo consumo contribuyen a la portabilidad y al empleo de baterías.



**Figura 6.** a) Arquitectura simplificada empleando el DS87C550, b) y c) muestran los consumos, teniendo en cuenta al dispositivo como un todo y considerando memoria RAM interna, respectivamente.

Finalmente, las características técnicas del Detector Fetal son:

♥ Intensidad ultrasónica emitida	≤ 20 mW / cm <sup>2</sup>
♥ Tensión de transmisión	2 V. pp.
♥ Ganancia del receptor	+ 39 dB
♥ Relación señal ruido (BW = 150 KHz.)	73 dB
♥ Modo de operación	Continuo
♥ Frecuencia nominal del ultrasonido	2 MHz
♥ Potencia sonora máxima	1 W
♥ Funciones	Salida para auriculares Indicador de baja batería Apagado automático Congelamiento de lectura 64 niveles de volumen Recarga de baterías Encendido / Apagado Aumentar volumen Reducir volumen Congelar lectura
♥ Mandos	4 x AA (4,8 – 6 V)
♥ Alimentación por baterías	0,9 W
♥ Potencia de consumo máxima	60 – 210 ppm.
♥ Margen de indicación de frecuencia fetal	7,8 x 14,8 x 4,5 cm
♥ Dimensiones	Sin sensor 0,38 Kg Con sensor 0,57 Kg.
♥ Peso	
♥ Cumple con las normas IEC 60601-1	

La intención de optimizar la arquitectura y las posibilidades del equipo ha quedado demostrada en este trabajo. No obstante, las etapas de preprocesamiento pueden ser mejoradas empleando otros dispositivos de bajo consumo y estructuras circuitales. Además, el método de procesado empleado está vigente hoy día en equipos comerciales debido a la simplicidad de su implementación.

Finalmente, el diseño queda abierto al empleo de otros dispositivos y métodos, tanto para el preprocesado como para el procesado.

### CONSIDERACIONES FINALES

Es importante destacar que el Detector de Latido Fetal es un equipo ampliamente utilizado a nivel mundial. Sin embargo, actualmente el sistema de salud en nuestro país no cuenta con este tipo de equipo, habiéndose comprobado en la práctica que la tenencia de este, por parte de los obstetras, constituye una herramienta de gran utilidad en el trabajo

cotidiano. Puede emplearse, tanto en hospitales maternos como en otros centros de salud del país y en el extranjero. Su producción tiene ventajas potenciales en lo relacionado con precio de venta y

servicio posventa, tomando en cuenta el precio de equipos similares. Tiene gran incidencia social en cuanto a que contribuye a reducir la mortalidad infantil en recién nacidos.

### REFERENCIAS

- [1] ATKINSON, P. and J. WOODCOCK (1986): "Doppler Ultrasound and Its Use in Clinical Measurements", Academic Press Inc.
- [2] BOTELLA, J. y J. CLAVERO (1983): "Tratado de Ginecología", Tomo 1, Editorial Científico Técnica.
- [3] DALMAU, A.: Comunicación personal.
- [4] EVANS, D.H. and W. McDicken (1999): "Doppler Ultrasound: Physics, Instrumentation and Signal Processing", 2<sup>nd</sup> Edition, John Wiley & Sons, Ltd.
- [5] JIMÉNEZ, A.A. (1998): Tesis de Maestría en Sistemas Automatizados de Medición "Sistema Detector de Latidos Cardíacos Fetales", ISPJAE, La Habana, Cuba.
- [6] JIMÉNEZ, A. **et al.** (2000): "Resultados Obtenidos en la Simulación del Método de Autocorrelación en Tiempo Real con la Finalidad de Adecuar su Implementación en Microcontroladores", **Reporte de Investigación** No.122, ICIMAF, Noviembre.
- [7] KOHDEN, N.: "Service Manual of PARTOCORDER, model OMF-7101".
- [8] OPPENHEIM, A.V.; R.W. SCHAFER and J. BUCK (1999): "Discrete-Time Signal Processing", 2/e Prentice Hall Inc.
- [9] PROAKIS, J.G. and D.G. MANOLAKIS (1992): "Digital Signal Processing: Principles, Algorithms and Application", Mac Milland Publishing Co., New York.
- [10] TAKEUCHI, Y. (1976): "U.S. patent 3991365", 9 Nov.
- [11] TAKEUCHI, Y. and M. HOGAKI (1978): "An Adaptive Correlation Ratemer: A New Method for Doppler Fetal Rate Measurements", Ultrasonics, May.
- [12] EVANS, D.H. (1989): "Doppler Ultrasound", John Wiley & Sons, Ltd.

# EL APORTE DE LOS FÍSICOS AL DESARROLLO DE LA MÚSICA

Daniel Stolik, Facultad de Física, Instituto de Materiales y Reactivos para La Electrónica (IMRE), Universidad de La Habana

## RESUMEN

Este trabajo está dirigido a reconocer el aporte hecho por los físicos en el desarrollo de la música, en saludo al año Internacional de la Física. Es altamente conocido el papel que ha jugado la física en la comprensión del universo, los fenómenos naturales y el desarrollo de la tecnología. Sin embargo su influencia en el desarrollo de la cultura y el arte, incluyendo la música es menos conocida. En este trabajo se describe cuando comienza y como se desarrolla un tratamiento de los fundamentos científicos del sonido y la música, proceso en el que notables físicos, reconocidos por su contribución a la formación de la física como ciencia, hicieron aportes fundamentales. Se parte de la reflexión de que la música tiene una base objetiva, que obedece a leyes naturales físicas. Se destaca el hecho de que con Galileo, además de comenzar la física como ciencia, también comienza el análisis y la comprensión científica de la música. En el artículo se destacan los aportes más importantes de los físicos al respecto hasta la aparición de los sintetizadores y las nuevas tecnologías en el siglo XX.

## ABSTRACT

This work is devoted to recognize the contribution made by the physicists to the development of the music, in greeting to the: International Year of Physics. It is very well-known the role played by Physics in the understanding of the universe, the natural phenomena and the development of technology. Less well-known is their influence in the development of the culture and the art, including the music. Specifically this work describes the beginning of a treatment of the scientific basis of the sound and music. Many physicists, who are recognized by its contribution to the formation of the physics as a science, made fundamental contributions to music. I started from the point of view that the music has an objective base and obeys natural laws of the physics. It is very interesting the fact that with Galilean, not only began the physics as science, also began the scientific understanding of the music. In this respect we remark the contributions made by very important physicists, until the appearance of the synthesizers and the new technologies in the XX century.

## 1. BASE OBJETIVA DE LA MÚSICA

Antes de abordar el tema del título de este artículo es conveniente analizar algunos aspectos, por ejemplo: ¿Cuál es el límite del radio de acción entre la ciencia y el arte?, tema muy polémico en el que existen muchos criterios e interpretaciones distintas, por lo que trataremos de reforzar aquellos elementos que pueden revelarse como más evidentes; en este sentido creo que no habría dudas en reconocer la importancia de utilizar los métodos físicos cuantitativos para comprender el significado de los sonidos, las escalas, los instrumentos y cualquier otro aspecto constitutivo de la música, pero no sólo con el propósito de entenderla y explicarla científicamente, sino también de enriquecerla tecnológicamente e inclusive musicalmente. En realidad, aun hoy en día existen conductas extremas: músicos que no quieren saber de tecnologías, tecnólogos que desdeñan la música como arte, musicólogos que ignoran muchos aspectos científico-técnicos de la música, profesores de física de los distintos niveles de educación (incluyendo escuelas de arte) que ejemplifican poco con el comportamiento físico de la música, sonidistas de experiencia práctica pero con poca base científico-técnica de los fenómenos ligados al sonido, constructores de instrumentos musicales que se guían solamente por manuales y tablas pero

con poco conocimiento del porque físico de los propios instrumentos que construyen. Soy del criterio que hay que: a) poseer una mente abierta a la diversidad musical en términos de géneros, timbres, escalas, consonancias, etc.; b) reconocer la tremenda importancia de las manifestaciones sensoriales, ligadas a los sentimientos y espíritu humanos; c.) entender las bases objetivas científico-técnicas que la soportan. Sobre este último particular baste recordar lo siguiente:

- La mayoría de los instrumentos musicales son invenciones técnicas del hombre (sea violín, piano, etc.).
- La formación, propagación y detección del sonido son fenómenos físicos estudiados y conocidos que se cuantifican matemáticamente.
- La señal sonora se transforma y manipula eléctricamente a voluntad del hombre.
- Se estudia científicamente la relación existente entre las sensaciones y los estímulos por la Psicofísica.
- Las tecnologías aplicadas a la música, incluyendo la electrónica, se desarrollan impetuosamente y se aplican con un alto grado de digitalización y computarización.

Por último, la duración de las etapas con que han ido introduciéndose las tecnologías, que también han influido en el desarrollo de la música, se ha reducido espectacularmente, por ejemplo:

- Los instrumentos musicales surgieron hace miles de años.
- La ciencia física y las matemáticas modernas surgieron hace unos 400 años.
- Las tecnologías electrónicas hace alrededor de 100 años.
- La electrónica integrada hace unos 40 años.
- La tecnología MIDI (interfase digital de instrumentos musicales) hace 20 años.
- Las nuevas tecnologías continúan incorporándose con una velocidad impresionante.

## **2. ¿CUÁNDO Y CÓMO ENTRÓ LA MÚSICA A FORMAR PARTE DE LA VIDA DEL HOMBRE?**

Aunque no está muy claro, se supone que en la inflexión del habla están los antecedentes del canto, que en la necesidad de comunicarse a grandes distancias por medio de la percusión está el origen de la sensación del ritmo por el hombre primitivo y que posiblemente en factores tales como el golpe en la cuerda del arco de caza o el silbido provocado por el sople del viento en la caña hueca está el descubrimiento del placer por elementos de la música tonal (1). Existen claras evidencias de la existencia de instrumentos en la antigua Mesopotamia, liras de más de 5000 años en Ur, bandas de músicos sumerianos de hace 4600 años, también en Egipto, por ejemplo, orquestas con instrumentos de cuerda y viento de hace 4000 años (1), (2), (3). Hasta hace relativamente poco tiempo se entendía que en unos 10000 años podrían ubicarse las primeras flautas de un solo tono, sin embargo un hallazgo de 1995 parece indicar la existencia de una flauta hecha de un hueso de oso joven de 4 notas de hace más de 40000 años (4), de la época del hombre de Neandertal, del que se pensaba no poseía cultura musical alguna. Lo que en cualquier variante nuestro sentido común nos indica, es que la aparición de la música y de los primeros instrumentos musicales primitivos fue producto de un proceso extremadamente lento. Particularmente la música alcanza una gran diseminación de forma muy arraigada en la sociedad de la antigua Grecia (5).

## **3. LA MÚSICA EN LA ANTIGUA GRECIA Y PITÁGORAS**

Una figura real del siglo VIII a.n.e., anterior a Pitágoras fue Terpandro, quien se considera el creador de la música griega clásica y posiblemente también de las siete notas diatónicas (6). Ganó los concursos de música celebrados en los 26 juegos olímpicos en Esparta, Año 676 a.n.e., o sea hace

unos 2800 años (en esa época también se competía en música en las olimpiadas).

Pitágoras (580-500 a.n.e.) ha tenido una trascendencia en la cuantificación de los intervalos musicales, aunque existe confusión de adjudicarle el descubrimiento de los intervalos de quinta y cuarta, ya que en realidad Terpandro, muy anteriormente los manejaba con los nombres griegos correspondientes de la época. Realmente lo que los Pitagóricos le adjudican a Pitágoras fue la formación inicial de la famosa escala Pitagórica que subsiste hasta nuestros días y más específicamente la definición de que si se disminuye a la mitad la longitud de una cuerda obtenemos el intervalo de octava, si se disminuye a  $\frac{2}{3}$  obtenemos el de quinta, si se disminuye a  $\frac{3}{4}$  obtenemos el de cuarta y que el intervalo entre la 4a y la 5a es un tono pitagórico igual a un  $\frac{9}{8}$ . Es muy interesante el papel que jugó la música en la antigua Grecia, aspecto que no abordaremos por su extensión, solo mencionaremos a Aristóteles (384-322 a.n.e.), quien ya sugería que el sonido se propagaba a través del movimiento del aire, aunque constituía una hipótesis basada más en la filosofía que en la experimentación física. También mencionaremos a Ctesibio (270 a.n.e.), Griego catalogado como "físico" e inventor (7). Se le atribuye el descubrimiento de la elasticidad del aire y de distintos dispositivos de aire, también ha trascendido por la notable invención del "Hidraulic", órgano musical en el que el aire era dirigido a pasar por los tubos del instrumento debido a un mecanismo forzado por el peso del agua.

Hasta aquí la corta referencia a los Griegos. Lamentablemente este buen paso de desarrollo de las ciencias y las artes, incluyendo la música, no continuó en la triste edad media, periodo que constituyó unos 1500 años prácticamente perdidos, fundamentalmente debido a los prejuicios religiosos, tema que tampoco abordaremos.

## **4. COMIENZO DEL APORTE DE FÍSICOS AL CONOCIMIENTO DE LA MÚSICA**

GALILEO. Con el renacimiento comienza de nuevo la alegría de vivir, se disparan las artes y las ciencias. Es conocido que con Galileo Galilei (1564-1642) comienza realmente la física como ciencia (8), pero es menos conocido que también con Galileo comienzan los primeros tratamientos científicos del sonido y de la música, del estudio moderno de las ondas y la acústica. Galileo eleva a nivel de ciencia el estudio de las vibraciones y de la correlación entre la altura o tono y la frecuencia de la fuente del sonido (relación entre frecuencia, longitud, diámetro, densidad y tensión de las cuerdas). En realidad el interés de Galileo por el sonido fue inspirado por su padre Vincenzo Galilei, matemático, cantante, músico, compositor, teórico de la música de reconocida reputación histórica (7). Como veremos, después de

Galileo, el conocimiento científico acústico de la música tuvo un rápido desarrollo.

MERSENNE. Físico-matemático francés. Martín Mersenne (1588-1648), (9), que estudió las vibraciones en las cuerdas; sus resultados se plasman en las tres leyes de Mersenne, publicadas en 1636 y que comenzaron a formar la base de la acústica musical. Calculó los intervalos de la escala musical del temperamento igual, que realmente se empezó a generalizar mucho después con el Clave Bien Temperado de Bach.

También hizo una primitiva medición de la velocidad del sonido en el aire. Actualmente disponemos de un buen conocimiento de los instrumentos de la época gracias a su descripción en su publicación "La armonía universal".

Con Galileo y con Mersenne queda establecido la dependencia de la frecuencia de una cuerda vibrante de: la tensión, la longitud y de la densidad lineal de la cuerda tensada. Lo que en términos posteriores se definió como:

$$f_1 = \frac{1}{2L} \sqrt{\frac{T}{\mu}} \quad (1)$$

T – Tensión en Newton.

L – Longitud de la cuerda en metros.

$\mu$  – Masa por unidad de longitud en Kg/m .

DESCARTES. Rene Descartes (1596-1650). Filósofo-físico-matemático francés. Uno de los padres del pensamiento moderno. Escribió el "Compendium musicae latino", uno de los escritos más notables de su tiempo sobre el tema de la música. En su correspondencia se encuentran frecuentes referencias a la música (7).

TORRICELLI y PASCAL. El físico italiano Evangelista Torricelli (1608-1647) en 1643 y posteriormente el físico francés Blas Pascal (1623-1662) (10), midieron la presión de la atmósfera, pero aun se estaba muy lejos de que se determinara que el sonido se propagaba en el aire gracias a una onda de fluctuación de la presión atmosférica. Aún en el siglo XVII muchos científicos y filósofos creían que el sonido se propagaba a través de unas partículas invisibles que se originaban en la fuente del sonido y se movían en el espacio hasta el oído.

El concepto del sonido como una onda estaba destinado a cambiar completamente el cuadro anterior, pero había que demostrarlo experimentalmente.

GUERICKE. En 1650 el físico alemán Otto Von Guericke (1602-1686) inventó la bomba para crear vacío parcial (10). Realizó los primeros estudios que revelaron que la luz viajaba en el vacío pero el sonido no.

BOYLE. En 1660 el científico anglo-irlandés Robert Boyle (1627-1691) mejoró la tecnología de vacío y pudo observar como la intensidad del sonido originado por un timbre (tipo reloj) colocado en una campana neumática disminuía a medida que el aire era extraído. Boyle concluyó que un medio como el aire era necesario para la propagación de las ondas sonoras (10) (11) (12). Aunque la conclusión es correcta, hay una imprecisión en la interpretación del experimento. No obstante, quedó establecido que el sonido se propaga más como una onda que como un flujo de partículas. La imprecisión se debe a que aún en las actuales bombas mecánicas de vacío, el remanente de aire es suficiente para la propagación de la onda sonora. En realidad la disminución de la densidad del aire hace que la vibración del timbre se va tornando incapaz de transmitirse al aire. El problema es de acoplamiento de impedancia por diferencia de las densidades entre el aire y el material sólido del timbre que vibra, o sea, que primero desaparece la transmisión de las vibraciones en la interfase sólido-aire. Para vacíos más altos la interpretación de que desaparece la propagación es absolutamente correcta.

HUYGENS. Christiaan Huygens (1629-1695). Físico holandés. Antes de 1661 elaboró un elegante método para calcular las longitudes de las cuerdas para cualquier afinación de la escala que fue aplicado a la escala del tono medio. Aplicó los logaritmos en el cálculo de las longitudes de las cuerdas y de las dimensiones de los intervalos musicales. Planteó la posible naturaleza consonante de distintos intervalos musicales como aproximaciones de los intervalos de sexta y de cuarta aumentada respectivamente, en la escala del tono medio y en la del temperamento igual. Demostró la estrecha relación entre ambas escalas. Describió la escala de 31 tonos en su *Lettre touchant le cycle harmonique* (Rotterdam 1691) y *Novus cyclus harmónicos* (Leiden 1724).

Una vez reconocido que el sonido era una onda, la medición de su velocidad se convirtió en un importante objetivo.

BORELLI. Giovanni Alfonso Borelli (1608-1679), físico italiano que en 1650 obtuvo (junto con Viviani, otro físico italiano) un buen valor de la velocidad del sonido, 350 m/s. Por ser de amplio dominio, en este artículo no es necesario describir los grandes resultados obtenidos en la física por los notables físicos que se mencionan, no obstante, en ocasiones, haremos alusión a los aportes a la física de algunos de ellos, que a pesar de ser también notables se conocen menos, por ejemplo, Borelli realizó trabajos de física y astronomía, incluyendo la publicación de un tratado en 1666 sobre la influencia de la atracción de los satélites de Júpiter, También fue el primero en sugerir la trayectoria parabólica de los cometas.

HOOKE. El notable físico inglés Robert Hooke (1635-1703) fue empleado por Robert Boyle para construir las bombas de vacío, cinco años después formulaba su famosa ley sobre la elasticidad de los cuerpos. Hooke en 1681 fue el primero en demostrar experimentalmente que la altura o tono que percibimos como sensación auditiva, de sonidos graves a agudos, está directamente relacionado con la frecuencia de las vibraciones que origina la fuente de sonido. Se basó en que la frecuencia del sonido que produce un cartón golpeado por una rueda dentada girando a velocidad constante, produce la sensación de la altura de una nota determinada, al aumentar o disminuir la velocidad angular de la rueda dentada se producen sonidos más agudos o más graves correspondientemente. De esta forma quedó demostrado la relación directa y objetiva existente entre el estímulo (la frecuencia del sonido) y la sensación (la altura o nota del sonido).

SAUVEUR. El físico francés Joseph Sauveur (1653-1716) a finales del siglo XVII principios del XVIII realizó estudios detallados sobre la relación entre frecuencia y altura (tono), también sobre las ondas en cuerdas tensadas. Fue el primero que sugirió una explicación física de la disonancia en la música. Promovió la aceptación del término "acústica" y fue quien sugirió oficializar este nombre para el estudio del sonido. Propuso fijar como frecuencia de referencia en la escala musical, la nota DO con una  $f = 256$  Hz, que se conoce como "escala científica" en que las frecuencias son función exponencial de 2, dicha escala se utiliza hasta los días de hoy en función de investigaciones científicas de la música por su comodidad.

NEWTON. El ilustre físico inglés Isaac Newton (1643-1727) fue el primero en realizar un tratamiento matemático del sonido en 1687 (en su famoso trabajo "Principia"), posterior a la demostración de que la propagación del sonido a través de cualquier fluido solo dependía de propiedades físicas (elasticidad, densidad). También calculó, a partir de consideraciones teóricas, la velocidad del sonido en el aire.

LEIBNITZ. Gottfried Wilhelm Leibnitz (1646-1716). Lo traemos a colación ya que, independientemente de Newton, desarrolló la teoría del cálculo matemático, lo que posteriormente permitió a otros utilizar el cálculo diferencial en el tratamiento de las vibraciones y las ondas.

Un dato interesante es que en 1738 se obtuvo en la Academia de Ciencias en París un valor más preciso de la velocidad del sonido en el aire: 332 m/s, increíblemente bueno, teniendo en cuenta las herramientas rudimentarias de la época. Compárese con la última medición que en 1986 dio el valor de 331.29 m/s a  $0^{\circ}$  C.

Simultáneamente con todos los estudios antes mencionados en el mundo del sonido, los teóricos

fueron desarrollando la teoría matemática de las ondas, elemento indispensable requerido para el desarrollo de la física moderna, incluyendo la acústica y dentro de esta la música. Aportaron a la teoría mecánica de las ondas y de las vibraciones sonoras Johan Bernoulli (1667-1748), Brook Taylor (1685-1731), Daniel Bernoulli (1700-1782), Leonhard Euler (1707-1783), Jean Le Rond d'Alambert (1717-1783), Joseph-Louis Lagrange (1736-1813) Sophie Germain (1776-1831), Siméon-Denis Poisson (1781-1840). Al mismo tiempo hicieron cosas más específicas relacionadas directamente con la música:

TAYLOR. En 1708 Brook Taylor desarrolló la teoría matemática de las cuerdas vibrantes de acuerdo con las observaciones experimentales previas. Fue el primero en expresar matemáticamente el movimiento de una cuerda vibrante sobre la base de principios mecánicos.

EULER. En 1738, además de profundizar teóricamente en el fenómeno mecánico de las vibraciones, fue el primero en plantear una teoría científica de la armonía musical, aunque fallida, tiene el mérito de ser uno de los pocos intentos de su época para establecer los estudios racionales dentro del mundo musical. Llegó hasta proponer una medida definida cuantitativa de la disonancia de un acorde.

DALAMBERT, quien en 1740 realizó la derivación de la ecuación general de onda, realizó aportes relevantes a la música que son ampliamente reconocidos (1), (2), (7). Realizó la primera teoría física de la armonía musical que publicó en *Éléments de musique* en 1752. En 1754 publicó sus reflexiones sobre la música, titulado *Réflexions sur la musique en général et sur la musique française en particulier*. También publicó tratados en acústica, física del sonido y como enciclopedista francés de su época escribió varios artículos sobre música para la *Encyclopédie*.

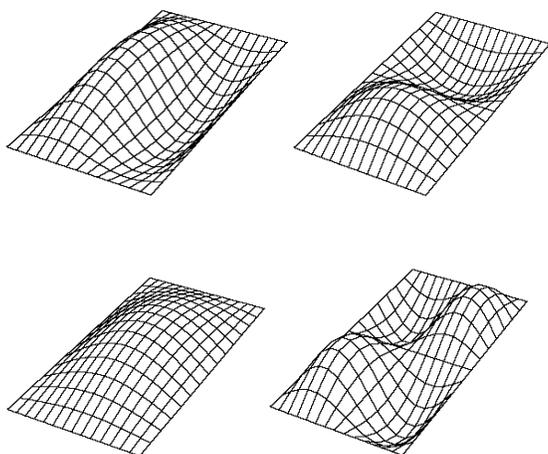
BERNOULLI JOHAN Y LAGRANGE. Contribuyeron al igual que Dalambert y Euler, al conocimiento del tono y del timbre del sonido producido por un instrumento musical y también a la naturaleza de la propagación del sonido en distintos medios. Lagrange y Euler aplicaron las nuevas ecuaciones diferenciales a las ondas en las cuerdas y en el aire.

POISSON. Extendió el desarrollo matemático de las vibraciones a las membranas tensadas.

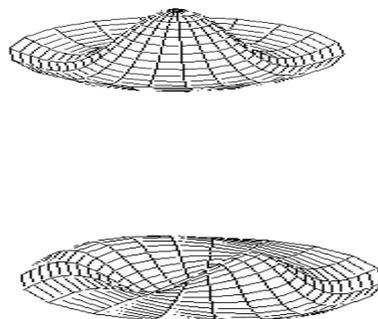
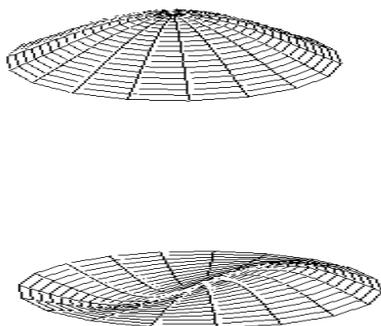
SOPHIE GERMAIN. Posiblemente una de las mejores explicaciones teóricas de las vibraciones fue dada en 1816 por la físico-matemática francesa Sophie Germain, de tal elegancia y complejidad que hasta los errores en el tratamiento al problema no fueron identificados hasta 35 años después, en este caso, por el gran físico alemán Gustav Robert Kirchhoff (1824-1887).

CHLADNI. El físico alemán Ernst Florenz Friedrich Chladni, (1756-1827) determinó la velocidad del sonido en distintos metales. Realizó los cálculos exactos, por un nuevo método, de la escala musical del temperamento igual. Pero sobre todo introdujo la técnica de observar experimentalmente los patrones de ondas estacionarias esparciendo arena en platos vibrantes, técnica que actualmente se utiliza en los instrumentos musicales, sobre todo de percusión. El mecanismo es sencillo, los modos de vibración en superficies originan patrones que tienen distintas líneas nodales (que no vibran) donde las partículas de arena se van concentrando debido a las vibraciones del resto de la superficie en cuestión, quedando así sobre la superficie, la figura del patrón altamente simétrica, delineada por las partículas de arena que quedan acumuladas en las líneas nodales.

Ejemplos de modos de vibración de una membrana rectangular:



Ejemplos de modos de vibración de una membrana circular.



## 5. EL SONIDO DESBORDA EL MARCO MECÁNICO-ACÚSTICO

El desarrollo que tuvo el electromagnetismo estaba destinado a provocar unas transformaciones tecnológicas sin precedentes en la historia de la humanidad. Las bases y condiciones para el mismo se sientan en un periodo que va desde finales del siglo XVIII hasta el XIX. El sonido y la música no van a escapar de estas revoluciones tecnológicas. Por solo citar un ejemplo de principio, las manipulaciones que se pueden hacer con las réplicas de las ondas de frecuencias sonoras electromagnéticas no se pueden hacer con las ondas sonoras mecánicas. Como recordaremos, entre los físicos más importantes que originaron las bases de tal desarrollo del electromagnetismo y con ello también propiciaron la revolución posterior del sonido y la música están los físicos: franceses Carlos Augusto de COULOMB (1736-1796) y Andrés Maria AMPERE (1775-1836), el danés Cristian OERSTED (1777-1851), el alemán Carlos Federico GAUSS (1777-1855), el inglés Miguel FARADAY (1791-1867), los alemanes Wilhelm WEBER (1804-1891) y Gustavo KIRCHHOFF (1824-1887), el escocés James Clark MAXWELL (1837-1879), el holandés Hendrick Antoon LORENTZ (1855-1921). El alemán Gustavo HERTZ (1857-1894). Pero además algunos de los gigantes del Electromagnetismo también incursionaron en forma directa en el mundo del sonido y la música. Veamos a continuación algo al respecto:

VOLT. El físico italiano Alejandro Conde de Volta (1745-1827) en 1800 trató de estimular eléctricamente la sensación de la audición conectando una pila a dos varillas metálicas que introdujo en sus oídos. Describió que sintió como “una sacudida en su cabeza”, y que poco después oyera “un ruido semejante al de una sopa espesa de hervir”. Fue tan desagradable que Volt no repitió el experimento ni siguió con esta investigación, por lo que la anécdota se refiere sólo a su gran interés por el tema.

BIOT. El físico francés Jean-Baptiste Biot (1775-1862) en 1808 dirigió los trabajos para la determinación de la velocidad del sonido en sólidos.

SAVART. El físico francés Félix Savart (1791-1841) continuó los trabajos de Hooke y perfeccionó la técnica de determinación de la frecuencia del sonido por la rotación de una rueda dentada, conocido hoy como discos de Savart. Realizó importantes estudios sobre los intervalos musicales e inclusive existe una unidad de intervalo musical llamado Savart (13).

FOURIER. Físico-matemático francés Jean-Baptiste-Joseph Fourier (1768-1830). Como se conoce, desarrolló la teoría sobre el análisis (descomposición) de una onda compleja periódica en sus componentes espectrales (armónicos). Lo contrario al análisis de Fourier constituye componer la señal a partir de la superposición de distintos armónicos, lo que se denomina síntesis de Fourier, de ahí el nombre de los sintetizadores de sonidos en la música. Fourier, sin proponérselo explícitamente como objetivo, hizo uno de los aportes más significativos en la comprensión física del sonido y del timbre musical.

OHM. El físico alemán Georg Simon Ohm (1787-1854) fue quien aplicó el análisis de Fourier al proceso de la audición, fue el primero en sugerir que el oído es sensible a las componentes espectrales. También planteó que el oído era sensible a las amplitudes y no a las fases de los armónicos de un tono complejo, lo que constituye una de las leyes más importantes del sonido y la música, conocida como Ley de Ohm de la audición, que no tiene que ver con su muy famosa ley de Ohm eléctrica.

WHEATSTONE Sir Charles Wheatstone (1802-1875), Sí, el mismo del puente eléctrico, de familia fabricante de instrumentos musicales. Inventó la concertina, un acordeón perfeccionado de gran extensión mundial. En 1836 intentó mantener la duración del sonido del piano con vibraciones forzadas por medio del paso de aire a través de una abertura en el extremo de la cuerda luego de retirado el martillo.

## 6. ESCALA Y AFINACIÓN MUSICALES

Vimos la escala científica propuesta por Saveur. En 1834 el físico alemán J. H. SCHEIBLER propuso fijar la altura (tono) de referencia de la escala musical a partir de la nota LA con una frecuencia de 440 Hz. En 1859 una comisión gubernamental francesa asesorada por el francés Jules Antón LISSAJOUS definió el standard legal para la nota La 435 Hz. En 1936 la American Standards Association adoptó la Escala a partir de LA 440 Hz, y que se generalizó en 1939 en Conferencia Internacional. (13).

## 7. LA PROBLEMÁTICA ESTÍMULO-SENSACIÓN

También en el siglo XIX comienzan los estudios científicos sobre el controvertido problema estímulo-sensación (3).

FECHNER. Gustavo Teodoro Fechner (1801-1887), fue un físico alemán sumamente interesante. Como no es muy conocido por los físicos veamos algunos de sus aportes a la física (10). Fue el primero en plantear en 1845 una teoría electrónica de la inducción electromagnética, desarrollada ulteriormente por Weber y Faraday. También el primero en plantear que la corriente eléctrica era el producto del movimiento de cargas eléctricas discretas. Ya en 1855 dijo: "La teoría ondulatoria es un momento necesario de la física, la atomística es un momento necesario de la teoría ondulatoria". Sin embargo, es ampliamente conocido como el padre de la psicofísica, que es la ciencia que trata sobre las relaciones cuantitativas entre el estímulo físico objetivo y la sensación subjetiva del hombre al respecto. Por ejemplo, desarrolló las expresiones sobre la diferencia mínima perceptible (JND-just noticeable difference) de dos frecuencias al acercarse hasta confundirse la sensación en una sola frecuencia, algo similar ocurre con la sonoridad (amplitud) de la onda sonora que puede ser discriminada por el oído, conocida como Ley de Fechner. Mas recientemente, en el marco de esta ciencia, se han precisado mejor estas relaciones cuantitativas entre estímulo-sensación.

HELMHOLTZ Ludwig Ferdinando Hermann Helmholtz (1821-1894), el famoso físico alemán, posiblemente el más conocido por los musicólogos, dio por primera vez las explicaciones científicas de la consonancia, el timbre, la naturaleza de las vocales y otros muchos aspectos físicos de la música. Contribuyó notablemente a la comprensión del mecanismo de la audición y la psicofísica del sonido musical. Su libro "Sensaciones del tono" (1863) es un súper clásico de la teoría de la música (14), uno de los más leídos que continúa inclusive editándose. Sus trabajos marcan una profundización en el estudio científico moderno de los fenómenos musicales.

## 8. ACELERACIÓN DEL DESARROLLO TEÓRICO Y TECNOLÓGICO

KUNDT. El alemán Augusto Kundt (1839-1894), desarrolló un número importante de técnicas para la investigación de las ondas sonoras, incluyendo el conocido tubo que lleva su nombre, donde en forma sencilla se ve la formación de la onda estacionaria que se establece por las huellas de arena fina suelta dentro del tubo de cristal.

KOENIG. Karl Rudolph Koenig Físico prusiano, creativo experimentador y quien diseñó en el siglo XIX muchos de los instrumentos utilizados para las investigaciones en audición y en música.

RAYLEIGH. John William Strutt. Físico inglés (1842-1919), Premio Nóbel de Física en 1904, por sus investigaciones de las densidades de los gases

más importantes y por su descubrimiento del argón. Más conocido como Conde Rayleigh. Realizó una enorme variedad de investigaciones acústicas, muchas de ellas incluidas en sus dos tomos del tratado "Teoría del Sonido" que publicó en 1877-78, monumento mayor de la literatura sobre acústica. Realizó trabajos básicos sobre la propagación de las ondas sonoras en gases, fluidos y sólidos, y los fenómenos de vibración y resonancia. Su importante reflexión sobre la audición binaural fue la base del desarrollo ulterior del sonido estereofónico.

EDISON. Hasta el tercer cuarto del siglo XIX, no existía constancia diferida auditiva (grabación) de los sonidos, como el habla y la música. En 1877 el norteamericano Tomas Alba Edison (1847-1931) comienza la grabación en cilindros, aparece el fonógrafo. Estas grabaciones y reproducciones del sonido eran completamente mecánicas, se movía el cilindro a mano a velocidad uniforme, grabándose y acercándose mucho a la bocina ya que la vibración de las voces era la fuente de energía que movía la aguja del fonógrafo. El fonógrafo mejorado de Edison en 1909 de cilindros de cera, podía tocar dos tipos de cilindros: los antiguos de 2 min. y los nuevos de larga duración (4 min.), con el doble de surcos. Ambos giraban a 160 rpm. También el alemán Emile BERLINER en 1877 inventó la grabación pero en platos, no en cilindros como Edison.

En menos de 100 años, las grabaciones de sonido evolucionarían en forma espectacular, de mecánicas a: eléctricas, ópticas, magnéticas y memorias sólidas con increíble calidad y fidelidad.

BELL. En la misma prolija época en que apareció la primera grabación, lo hace también la transmisión alámbrica del sonido, el físico norteamericano Alejandro Graham Bell (1847-1922) comienza la telefonía en 1876, se perfecciona por BERLINER en 1877 y el micrófono por el inglés David HUGHES (1831-1900) en 1878.

La invención del micrófono fue sumamente importante (15), sencillamente porque por primera vez las ondas sonoras se podían convertir, aún sin amplificación, en sus réplicas eléctricas y viceversa (el parlante). Una vez transformada la señal sonora en su similar eléctrica, ésta podría someterse a múltiples manipulaciones antes de convertirla nuevamente en onda sonora mecánica.

Estaba por comenzar la transmisión inalámbrica de las ondas electromagnéticas. Aspecto sumamente importante ya que existía la posibilidad de transformar la señal sonora mecánica en electromagnética y así transportarla a grandes distancias.

MARCONI. Guglielmo Marconi (1874-1937). Físico italiano inventor de un exitoso sistema de radio telegrafía en 1896, por lo que en 1909 recibió el

premio Nóbel de Física. Posteriormente trabajó en el desarrollo de las comunicaciones de longitudes de onda más cortas, base de las modernas comunicaciones a largas distancias.

POPOV. El físico ruso Alexandr Stepanovich Popov (1859-1906) independientemente de Marconi desarrolló un sistema de radio telegrafía en 1895, por lo que estimo merece la coautoría por la invención del equipo de radio (sin hilos).

Haremos una pausa del tema electromagnético, para preguntarnos ¿cómo era la calidad del sonido en las salas de teatro y otros locales? Ya el físico JOSEPH HENRY (1797-1878) en 1856 había incursionado teóricamente en esta problemática, pero fue a finales del propio siglo XIX, que se dio el paso más importante al respecto.

SABINE. Wallace Sabine (1868-1919). Físico norteamericano (16), considerado el padre de la arquitectura acústica moderna. En 1895 descubrió que el producto del tiempo de reverberación multiplicado por la absorción total del local es proporcional al volumen del local, denominada Ley de Sabine y la unidad de absorción del sonido es el sabine.

La primera aplicación Sabine la realizó en el Boston Symphony Hall que se inauguró en el 1900 y constituyó un gran éxito acústico.

Estaba por llegar el siglo XX, y con él estaba naciendo la electrónica. En 1895 LORENTZ postuló la existencia del electrón. El inglés John Ambrose FLEMING en 1904 inventó el diodo de tubo al vacío.

DE FOREST, Lee de Forest (1873-1911) físico norteamericano, en 1906 puso una rejilla al diodo, inventa el tríodo y nace el amplificador electrónico. Alrededor del 1911, después de distintos perfeccionamientos tecnológicos, nace la electrónica práctica.

Para los nuevos instrumentos musicales el tríodo daba perspectivas sin precedentes, debido, por un lado, a la posibilidad de amplificar la señal y por el otro, a que los circuitos osciladores, sobre todo a partir del heterodino en 1917, eran capaces de replicar cada armónico del espectro musical, posteriormente sintetizarlos y simular timbres.

THEREMIN. Lev Sergueyevich Termen (Theremin) (1896- 1993), físico ruso, que inventó uno de los primeros instrumentos electrónicos, con circuitos osciladores de válvulas y filtros, considerado el primer sintetizador de sonidos musicales en 1920. Se ejecuta sin contacto directo de las manos, sino por medio de la inducción electromagnética, mediante el cambio de la capacitancia entre las manos y las antenas del instrumento. Mostrado a Lenin en 1922, a Einstein en 1927. Theremin fue

honrado en el festival de música electrónica en Francia en 1989 y en la Universidad de Stanford (aún vivo) en 1991. Usado en la música de muchos filmes de ciencia ficción, en grabaciones de grupos como el de rock de Estados Unidos: Beach Boys.

**NERNST.** El físico alemán Hermann Walther Nernst (1864-1941), padre de la físico-química moderna. Fue uno de los principales iniciadores de los pianos eléctricos y sus principios fueron aplicados a toda una serie de instrumentos de este tipo (7). El piano neo-Bechstein o Siemens-Bechstein, desarrollado bajo su dirección se presentó en 1931. Similar al piano, sin tabla resonante y con martillos de menor fuerza. Las vibraciones se recogen por electroimanes, amplificadas electrónicamente y transformadas de nuevo en ondas sonoras por altoparlantes. De timbre ajustable a voluntad vía condensadores variables, entre otras facilidades. El perfeccionamiento ulterior de este piano siguió bajo su dirección. Nernst recibió el premio Nóbel de Química en 1920.

**RAMAN.** Sir Chandrasekhara Venata Raman (1888-1970), físico indio. Desarrolló una teoría descriptiva, publicado en 1918, sobre el movimiento de la cuerda frotada por arco de los instrumentos musicales de cuerda (violín, viola, celo,..), cuyo comportamiento vibracional es no lineal (17). Publicó estudios sobre tambores de la India en 1934. Recibió el premio Nóbel de física en 1930 por su trabajo sobre la dispersión de la luz y por el descubrimiento del efecto que lleva su nombre.

El número de físicos trabajando en temas relacionados con la música fue creciendo, aspecto que no podemos detallar en este artículo por razones de tiempo y espacio, solo mencionaremos algunos muy específicos.

**BEKESY.** Georg Von Bekesy (1899-1972), físico húngaro-norteamericano. Demostró como se propaga, resuena y analizan las componentes espectrales de la onda sonora en el espacio de la conocida membrana basilar del oído interno. Recibió el Premio Nóbel en 1961 de fisiología y medicina por su descubrimiento sobre el significado físico de como el sonido es analizado y convertido en señales eléctricas dentro del oído interno, que se comporta como un analizador espectral. Su libro "Experiments in Hearing" publicado en 1960 (18), es la obra magna sobre la teoría del oído.

## 9. LA ERA DE LAS NUEVAS TECNOLOGÍAS

En el siglo XX se desarrollan a plenitud: las Comunicaciones, el Control, las Componentes y la Computación, en lo que muchos llamaron el desarrollo de las cuatro "C".

**SCHOKLEY, BARDEEN, BRITAIN.** A la altura del 1949 se descubre el efecto transistor por tres

físicos norteamericanos: William Schokley (1910-1989), John Bardeen (1908-1991), Walter Houser Brittain (1902-1987), William Schokley (1910-1989), por lo que recibieron el premio Nóbel en 1956. A partir de los circuitos transistorizados continuaron desarrollándose rápidamente los instrumentos musicales electrónicos y los equipos de grabación, reproducción y transmisión. Se perfeccionó el sintetizador que era aún de grandes dimensiones.

**MOOG.** Robert Moog (1934-2005), físico norteamericano recientemente fallecido. En 1964 fue el primero en diseñar varios tipos de sintetizadores compactos de precio moderado y que suministraba un extenso rango de manipulación del sonido. Además de producir ondas sinusoidales, triangulares, cuadradas y dientes de sierra, el sintetizador MOOG tenía generadores de ruido blanco, amplificadores con control de voltaje, filtros pase banda y secuenciador. Esta tecnología analógica se convirtió en la base de los sintetizadores modulares y portátiles producidos en masa en las décadas 60s y 70s. El Polymoog, podía producir acordes y apareció en 1976. Época en que comenzaron a producirse sintetizadores compactos comerciales por varias empresas importantes.

En el campo de los componentes se dio un sostenido desarrollo de los circuitos integrados. Por solo mencionar una de sus características (19), digamos que el nivel de integración en 1960 era de varias decenas de transistores equivalentes por  $\text{cm}^2$ , en 1966 era de cientos, en 1969 miles, en 1975 decenas de miles, en 1985 cientos de miles, en 1995 varios millones, en el 2000 decenas de millones y actualmente más de 100 millones de transistores equivalentes por chip.

La digitalización y la computación evolucionaron a una velocidad vertiginosa. Se dispararon las nuevas tecnologías de la información y las comunicaciones, el hardware, el software en todas las esferas, el procesamiento digital de señales, incluyendo la científica también en la música.

En el siglo XX, no obstante constituir una pequeña parte de los físicos, cientos de físicos se han dedicado a estudios multifacéticos de la música, también plasmados en cientos de artículos científicos y decenas de libros y monografías, comportamiento que continúa actualmente en el siglo XXI.

## 10. LOS FÍSICOS TAMBIÉN DISFRUTAN LA MÚSICA

Además del aporte al desarrollo científico y tecnológico de la música, a los físicos, al igual que a todos, les gusta: oír, apreciar, cantar, bailar, ejecutar y hasta posiblemente componer música. Podemos citar a notables físicos aficionados a tocar algún

instrumento musical: Piano, el alemán premio Nóbel de Física de 1918 Max Plank (1858-1947). Tumbadora y bongoes, el norteamericano Nóbel de Física de 1965 Richard Feynman (1918-1988). Violín, el alemán Albert Einstein (1879-1955) Nóbel en 1921.

## 11. CONCLUSIÓN

Esperamos que este trabajo, que consideramos trata un tema poco conocido, pueda contribuir a que el lector tenga una justa valoración sobre el papel jugado por los físicos en el desarrollo de la música.

## REFERENCIAS

1. JEANS, Sir JAMES (1953): **Science and Music**, Cambridge University Press, London, England.
2. WOOD, ALEXANDER (1964): **The Physics of music**, Methuen & Co., 6<sup>th</sup> edition, London, England.
3. STEVENS, S. y F. WARSHOFKY (1971): **Sonido y Audición**, Colección científica de Time-Life, Time Inc., Offset Multicolor S:A: México, DF.
4. La Flauta Neandertal, Reportes en: [www.apnet.com/insight/04031997/grapha.htm](http://www.apnet.com/insight/04031997/grapha.htm); "Science" Nov. 1996, "Muy Interesante" Feb1997.
5. ASKILL, JOHN and D. VAN NOSTRAND (1979): **Physics of Musical Sound**, Co., New York.
6. VELÁZQUEZ, G.O. (1966): **100 Biografías en la Historia de la Música**, IPN, México.
7. SCHOLE, P.A. (1981): **Diccionario Oxford de la Música**, Editorial Arte y Literatura, Ciudad Habana.
8. Diccionario enciclopédico de Física, Redactor principal A.M. Projorov, Ediciones "Enciclopedia Soviética" Moscú, 1983.
9. MAGIE, W.F. (1969): **A Source Book in Physics**, Harvard University Press, Cambridge, Massachussets.
10. SPASSKI, B.I. (1963): **Historia de la Física** (en ruso), Universidad de Moscú.
11. Enciclopedia Encarta 2004.
12. Enciclopedia Británica 1999.
13. Handbook of Chemistry and Physics, 67<sup>th</sup> Edition, 1986-1987, CRC Press. Florida.
14. HELMHOLTZ, H. von (1954): **On the Sensation of Tone**, Dover Publications.
15. STOLIK, D. (2004): Textos del Curso "Fundamentos físicos de la música", Facultad de Fisica, Universidad de La Habana.
16. SABINE, W.C. (1964): **Collected Papers on Acoustic**, Dover Pub.
17. FLETCHER, N.H. (1999): "The nonlinear physics of musical instrument", **Rep. Prog. Phys.** 62, 723/764, UK, IOP Publishing Ltd.
18. BEKESY, G. (1960): **Experiments in Hearing**, McGraw-Hill.
19. MILLMAN, J. (1979): **Microelectronics**, Mc.Graw-Hill Book Co., New York.

# QUANTIFYING THE IMPACT OF PHYSICS THROUGH SCIENTIFIC PUBLICATIONS

Ernesto Altshuler, “Henri Poincaré” Group of Complex Systems and Superconductivity Laboratory, Physics Faculty-IMRE, University of Havana, Cuba

## ABSTRACT

I present some of the parameters internationally used to measure the impact of scientific articles, and discuss their advantages and disadvantages. I introduce data related to the impact of scientific work in Physics generated in the area of Latin America, and, specifically, in Cuba. Although this paper cannot be taken as a definitive bibliometric study, I hope it offers some clues to conduct a systematic investigation of the impact of Cuban Physics in the international context, using scientific publications as a measuring tool.

## RESUMEN

En este trabajo presento algunos parámetros utilizados internacionalmente para medir el impacto de los artículos científicos, y discuto sus ventajas y desventajas. Introduzco datos relacionados con el impacto del trabajo científico en el campo de la Física generada en el área de Latinoamérica y, específicamente, en Cuba. Aunque este artículo no puede ser tomado como un estudio bibliométrico definitivo, espero que ofrezca algunas ideas sobre cómo conducir una investigación sistemática del impacto de la Física cubana en el contexto internacional, utilizando las publicaciones científicas como herramienta de medición.

## 1. INTRODUCTION

During a recent visit of Leon Lederman<sup>1</sup> to the Physics Faculty, University of Havana, he presented to our students the hypothetical case of a scientist who is offered all the necessary tools to reach the understanding of the ultimate truths of the Universe. The only condition is that the scientist should be confined to a desert island, and should keep his findings to himself. His firm answer is: NO!. The story has to do with the need of transcendence, an old concern of humans so elegantly posed by Diderot in 1875: “Posterity is, for the philosopher, what the Other World is for the religious man”.

In the academic arena, *publication* is one of the highest expressions of the need of transcendence. If in past ages books constituted the typical means to spread original scientific work (remember Galileo’s “Dialogues on two new sciences” and Newton’s “Principia”), publication of articles in peer reviewed scientific journals constitutes a standard in the contemporary road to scientific transcendence –including in the term, beyond personal recognition, the need of dissemination of frontline knowledge to ultimately improve mankind.

An important aspect of contemporary publications is that they open one way to quantify the impact of the scientific work of a person, or even of a country. I should underline, however, that there are other parameters that can be used for that purpose, such as the % of the National Gross Product devoted to

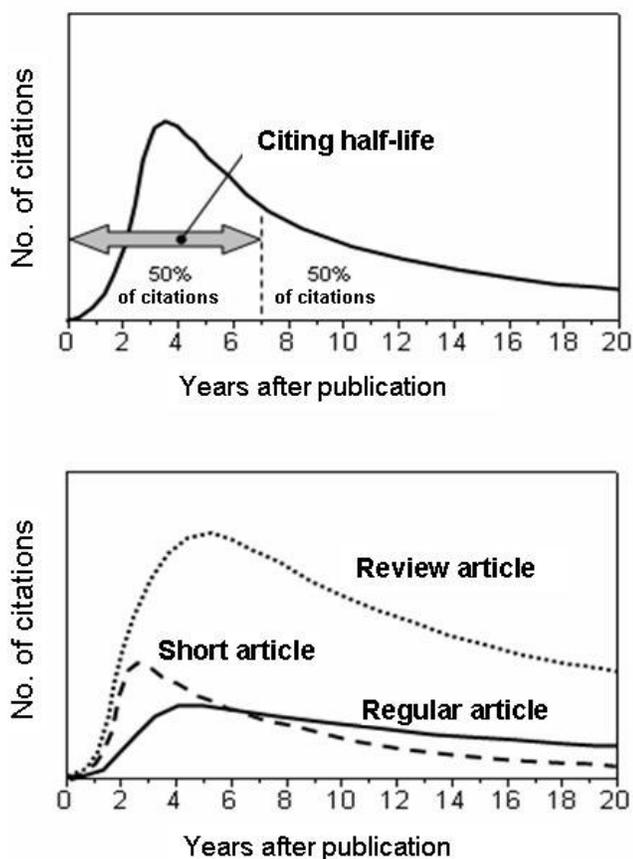
research, the number of scientists per capita, or the number of patents.

In the next sections, I shall describe how the impact of scientific publications is measured, with emphasis in the *impact factor* and *half-life* parameters, and present their values for several journals in the Physical and Chemical sciences. I discuss some limitations of these parameters related to social, cultural and even political issues that unavoidably influence the scientific arena. Then, I shall explain some criteria used to quantify the scientific impact of individuals, based on their publications in peer-reviewed journals. Along the road, I offer data relative to scientific publications by Latin American authors and, specially, Cuban scientists. The latter –still very far from a true bibliometric study– are published by the first time, to the author’s knowledge.

## 2. A COUPLE OF BIBLIOMETRIC INDICES

In a few words, the impact of a scientific paper can be measured by counting how many times and for how long it is cited in later scientific papers. Figure 1 shows two hypothetical graphs illustrating the number of times an average paper published in a scientific journal is cited by other papers published in the same or other journals [1]. The graphs indicate that the number of citations first increase and then slowly decreases.

<sup>1</sup>1988 Physics Nobel Prize winner for his contributions to the study of neutrino and leptons.



**Figure 1.** Time evolution of the impact for hypothetical articles.

It must be stressed that Figure 1 shows just an average behavior: there are, for example, “gray” articles that are never cited –unfortunately a case more common than we would like to accept. In fact, the road to a large number of citations does not necessarily follow the hypothetical curves shown in Figure 1. For example, record-citation paper “A model of Leptons”, published by Nobel prize winner Steven Weinberg in 1967 has followed a very tortuous path to fame. In a recent article series by James Riordon, the authors analyzes the 10 most cited papers published in the journal *Physical Review Letters* [2]: Weinberg’s was paper No. 1, with 4602 citations. The curious thing about this paper is that it received most of its citations only after 1971 –4 years after its publication. The reason was the following. The paper proposed a theory establishing a relation between electromagnetic and weak interactions. However, Weinberg was not able to show at that time that the theory was renormalizable, which limited its practical usefulness. Only in 1971 Gerard’t Hooft was able to prove it, opening the doors of Weinberg’s 1967 paper to the book of records... and giving a final push to Weinberg’s nomination to the Nobel Prize in 1979.

Nobody should be surprised by the fact that scientific journals have their own “pedigree”: some are regarded more prestigious than others. In this world, where everything is ranked –from MTV music hits to the most relevant personalities of the Millennium–, scientific journals do not escape the mainstream. Amongst the most popular rankings of scientific journals are the ones reported by the U.S.-based *Institute for Scientific Information* (ISI), resting on the analysis of tens of thousands of publications from all over the world.

Let us assume that the graph shown in Figure 1 represents the average time behavior of the citations to papers published in journal X. The time elapsed from the time of publication of the article and the moment when it has reached a 50% of the its total number of citations, is called *citing half-life* of journal X. In the hypothetic example shown in the upper section of Figure 1, its value is 7 years. This parameter gives an idea of the permanence in the scientific scenario of the papers published in journal X. So, it might be said that a large half-life journal publishes papers that tend to “make History”.

However, a different parameter is the one used to establish the ranking of scientific journals: the *impact factor*, introduced by ISI in the 1960’s [3]. The impact factor of journal X is the average number of citations per year that an average paper of journal X receives, within a given interval of time after the publication of the article. For example, if journal X has published 500 papers during 1997 and 1998, and those articles have received a total of 1000 citations during 1999, the impact factor of journal X during 1999 is  $1000/500 = 2.00$ . Of course, the articles published in a journal with high impact factor strongly attract the attention of the scientific community.

From now on, I will call “IF04” the impact factors reported by ISI in 2004, and “<IF>”, the impact factors averaged within the period 1974-2000 [4]<sup>2</sup>.

In the field of Physics, the review article journal *Reviews of Modern Physics* (IF04 = 32.77; <IF> = 16,61) boasts the highest absolute impact factor, while *Physical Review Letters* (IF04 = 7.22; <IF> = 6,57) has the highest impact factor amongst the journal publishing original papers in the field of Physics. In both cases, numbers are supported by an established prestige: while future Nobel winners tend to publish in *Physical Review Letters* some of their key original results, their Nobel lectures tend to be published in *Reviews of Modern Physics*. *Applied Physics Letters* (IF04 = 4.31; <IF> = 3,42) is the most important journal in the field of the applications of Physics. Finally, it is worth noting that *Nanoletters*

<sup>2</sup>We have taken this particular period since the full data can be accessed freely in the webpage of reference [4], and because it comprises the lapse where many of the Cuban papers in those journals were published.

(IF04 = 8.45) –devoted to the “hot” area known as nanoscience– has emerged as a key journal in the fields of Physics and Chemistry in the last years.

### 3. LIMITATIONS OF IMPACT FACTORS

#### 3.1. The issue of diversity

While the prestige of the publications mentioned above is very well established, it is dangerous to evaluate the importance of a given journal only by the size of its impact factor. One of the elements that make any comparison difficult is that there are journals specialized in the publication of just one type of article. While most scientific journals publish *original papers* of medium or short sizes, some specialize just in short contributions –like *Physical Review Letters* and *Applied Physics Letters*– and others only in review articles, such as *Reviews of Modern Physics*. Figure 1 suggest that articles appearing in the former type of journals tend to receive many citations in a very short time, while those published in the latter ones receive a large number of citations, but distributed within a longer period of time. The editors of review journals often invite recognized scientists to write papers that examine an entire subfield of the discipline, typically pointing out strengths and weaknesses, and proposing future lines of research. Those papers are popular amongst researchers entering a given field of research, and eventually tend to shape the scientific work of a scientific community.

The existence of different disciplines associated to diverse scientific communities constitutes another obstacle to compare scientific impacts based exclusively on the impact factor. For example, there are areas that generate a huge amount of citations due, amongst other things, to the great number of scientists involved: Medicine and Pharmaceutical Science are two examples. It is not strange that a scientific paper reporting a new drug or medical procedure is cited many times due to the fact that the new drug or procedure is statistically studied in hundreds of hospitals, producing new publications. It is my opinion, however, that some comparisons are reasonable among the exact sciences, which is supported by the compilation of average impact factors of Physics and Chemistry journals illustrated in Table 1. It shows that the impacts within the different ranges selected are quite similar, perhaps with some “advantage” by Chemistry journals. Appendix 1 contains a list of the impact factors in the period 2000-2004 of most of the journals shown in Table 1. Some of them are not listed either because they disappeared, or because their impact factors have decreased below 5.00. Some new journals have been added to the list, in the light of their high impact factor during the last years.

#### 3.2. Scientific prejudices

Another element that makes difficult the comparison of scientific results exclusively by means of impact factors is the “scientific prejudices” of the community involved –composed, after all, by humans.

<b>Table 1.</b> Comparison between the impact factors of some Chemistry and Physics journals (averages taken within the period 1974-2000).	
<b>CHEMISTRY</b>	<b>PHYSICS</b>
⟨IF⟩ > 10	
12.50 CHEM REV 11.59 ELECTROANAL CHEM 11.43 SURF SCI REP 10.91 ORG REACTIONS	16.47 REV MOD PHYS 12.61 SOLID STATE PHYS 11.43 SURF SCI REP
10 > ⟨IF⟩ > 9	
9.33 ACCOUNTS CHEM RES 9.38 PROG INORG CHEM	9.71 ADV PHYS 9.96 ANNU REV ASTRON ASTR
9 > ⟨IF⟩ > 8	
8.55 PROG PHYS ORG CHEM 8.26 ADV ORGANOMET CHEM 8.25 ADV CHEM PHYS	8.25 ADV CHEM PHYS 8.02 ADV NUCL PHYS
8 > ⟨IF⟩ > 7	
7.99 ADV INORG CHEM 7.93 ADV CYCLIC NUCL RES 7.18 ADV INORG CHEM RAD 7.02 ANNU REV PHYS CHEM	7.10 PHYS REP 7.02 ANNU REV PHYS CHEM
7 > ⟨IF⟩ > 6	
6.78 ADV PHOTOCHEM 6.47 ADV PHYS ORG CHEM 6.24 ADV CATAL 6.02 STRUCT BOND 6.04 TOP STEREOCHEM	6.57 PHYS REV LETT 6.43 J HIGH ENERGY PHYS
6 > ⟨IF⟩ > 5	
5.63 Z ANGEW CHEM INT EDIT 5.59 CHEM SOC REV 5.45 MASS SPECTROM REV 5.30 ORGANOMET CHEM REV A 5.29 MAT SCI ENG R 5.12 J PHYS CHEM REF DATA 5.00 PROG MACROCYCL CH	5.82 EUR PHYS J C 5.81 REP PROG PHYS 5.45 MASS SPECTROM REV 5.29 MAT SCI ENG R 5.12 J PHYS CHEM REF DATA

Scientific prejudices influence, of course, where and when a discovery is published, and also the attention it receives by the scientific community. A good example is the publication of the discovery of high temperature superconductors (HTc's). During the

1950's, much respected BCS theory established an upper limit for the critical temperature of a superconductor. No experimental physicists in his (or her) right mind would dare to attempt to find a superconductor with a critical temperature above approximately 30 degrees Kelvin (30 K). One exception was Alexander Müller, from IBM at Zürich, who had been quietly working on perovskite-type materials, where he suspected superconductivity could be found. After years of work with his colleague Johannes Georg Bednorz, they managed to synthesize a ceramic material with a critical temperature near 35 K. Perhaps due to the possibility of a negative reaction of the physics community, the results were published in 1986 under the low-profile title "Possible High- $T_c$  superconductivity in the Ba-La-Ca-O system", in *Zeitschrift für Physik* – a journal with a relatively modest impact factor ( $\langle IF \rangle = 2,32$ ) if compared, for example, with *Physical Review Letters*. In fact, the paper was unnoticed for most of the scientific community.

Fortunately C. W. Chu, from the University of Houston, understood the importance of the result, and started to investigate frantically related compounds with the collaboration of colleagues from the University of Alabama. It resulted in the discovery of a ceramic superconductor with a critical temperature higher than 90 K (i.e., above the boiling point of nitrogen, a very important achievement for applications). The result was rapidly published in *Physical Review Letters* in 1987, under the strong title "Superconductivity at 93 K in a new mixed-phase Y-B-C-O compound system at ambient pressure", and produced an instantaneous earthquake in the field of Physics. The climax was reached shortly after the publication of the paper, in a round-the-clock meeting in New York known as the "Woodstock of Physics", in analogy with the legendary rock festival of 1969. The importance of Bednorz and Müller's paper in *Zeitschrift für Physik* finally became widely acknowledged, and the authors received the Nobel prize in Physics in 1987 [5].

### 3.3. Idiomatic limitations

As in many other aspects of social life, historical periods have been marked by the domination of certain cultures also in the scientific arena. Consequently, certain languages have dominated science in different eras. In the XIX century the languages of the most developed European countries dominated science: English, German and French. That situation changed dramatically after Second World War with the exodus from Europe to the United States of scientists of Jewish origin and from other ethnical, political and religious backgrounds persecuted by the Nazis [6]. Physicists Albert Einstein and Enrico Fermi are perhaps the most relevant amongst them. The result was the

transformation of the U.S. into a super-power in Physics from the second half of the XX century. This situation undoubtedly contributed to the domination of scientific publications *in English*, edited, in the case of Physics, by the main U.S. in the field: *The American Physical Society* (publishers of *Reviews of Modern Physics*, *Physical Review Letters*, etc.), and *The American Institute of Physics* (publishers of *Applied Physics Letters*, etc.).

Although most referees and editors of prestigious scientific journals work under high professional and ethical standards, some of them recognize that a manuscript containing idiomatic defects can influence an editorial decision. The rationale is the following: "if the authors have been careless with grammar and spelling, why should one discard that they have been also careless in their experiments or calculations?" [7]. That reasoning is, of course, valid to some extent. But, at any rate, non-English speaking scientists have nowadays *two* barriers to overcome when writing a scientific paper: the difficulties of science itself, *and* those of a foreign language.

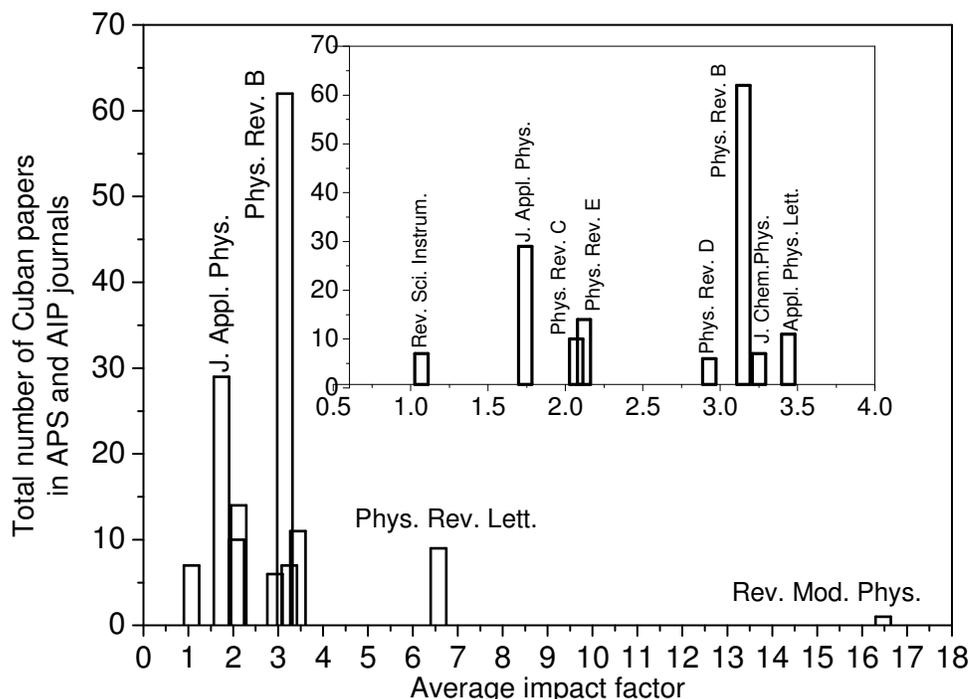
### 3.4. Economical, social and political issues: The Geography of scientific impact

Some scientific journals publish papers in all the branches of science. Amongst them, *Nature* ( $IF_{04} = 32,18$ ;  $\langle IF \rangle = 16,07$ ) and *Science* ( $IF_{04} = 31,85$ ;  $\langle IF \rangle = 14,68$ ), are the most prominent ones. They reject between 90 and 95% of the manuscripts they receive, and a sizable amount of their decisions are purely editorial (i.e., the papers are not examined by referees). Beyond the numerical values of their impact factors, most scientists agree that they are the most visible journals in the field of natural sciences. Based on statistics of 1994, approximately 85% of the papers published by *Science* were signed by authors from an "elite" of countries: United States (~20%), United Kingdom (~17%), France (~15%), Germany (~14%), Canada (~12%) and Japan (~7%). Scientific powers like Russia or China modestly contributed to the remaining percent [7].

Regarding the "global" geography of high impact papers authored by Third World and, particularly, by Latin American scientists, the global situation can be easily predicted. The share of the latter in the scientific production in all fields of science, measured through the number of papers in indexed journals is less than 5% of the world production [8]. Although I do not have data for Cuba embracing publications in all indexed journals at the time of writing this paper, it is safe to say that, in spite of the declining economy of the early 1990's, our scientific community have

started to increase its share of publications in internationally recognized journals ever since, achieving more and more space and impact.

number of small contributions, specially from Las Villas Central University. Authors from the University of Havana are also involved in 70% of the



**Figure 2.** Number of articles appeared in APS and AIP journals, signed by physicists working in Cuban institutions (period 1994-2004).

Figure 2 shows some data extracted from the American Physical Society (APS) and American Institute of Physics (AIP) websites, related with articles published in the period 1994-2004 on journals edited by those organizations, where several of the most prestigious journals in Physics are included. I have only included in the statistics the hits returned when the word “Cuba” is typed in the space “affiliation” of the searching engine. Since the search is free, I invite the reader to perform his (or her) own investigation. Notice that my selection criterion has several clear limitations, such as neglecting important journals (those edited by *Elsevier Science*, to put one possible example), but also some less evident, such as neglecting papers where Cuban researchers signed as members of the foreign institution where they were doing scientific work at the time of the publication. As a complement to Figure 2, Table 2 shows the data of the first Cuban publications appeared in APS and AIP journals, selected by using the search criterion described above.

It is worth noting that, in a first approximation, 80% of the publications comprised in Figure 2 are signed by at least one author from the University of Havana (basically from the Physics Faculty or from the Institute of Materials and Reagents), while the remaining 20% is roughly shared by the ICIMAF and the InSTEC (also Havana-based institutions), with a

“pioneering papers” of Table 2. The inspection of the list of papers comprised in Figure 2 and Table 2 are two-fold alarming. First, they show an overwhelming concentration of Physics authors in Havana. Second, there is a relatively small number of authors contributing to those papers within each institution.

Journal	<IF>	Reference
J. Chem Phys.	3.25	31, 467 (1959)
Appl. Phys. Lett.	3.44	46, 853 (1985)
J. Appl. Phys.	1.74	59, 2114 (1986)
Phys. Rev. B	3.15	37, 4583 (1988)
Phys. Rev. D	2.93	40, 1255 (1989)
Phys. Rev. E	2.12	49, 4027 (1994)
Phys. Rev. Lett.	6.57	76, 42 (1996)*
Phys. Rev. C	2.07	55, 2471 (1997)
Rev. Sci. Instrum.	1.07	69, 3634 (1998)
Rev. Mod. Phys.	16.47	76, 471 (2004)

\*It is important noting reference 67, 2335 (1991), where a

Cuban author has not signed “Cuba” as “affiliation”, but as “permanent institution”.

But let us go back to the effect of economic limitations on the impact of publications generated from Third World countries. One evident effect is the one resulting from lack of material support and ageing of available equipment. A glance at the contents of the papers contained in Figure 2 reveals also “second order” effects. For example, 70% from the total of papers are “theoretical” and only a 30% involve experimental work. While 60% of the theoretical papers are exclusively signed by Cuban institutions, the same number for experimental papers is as small as a 15%. These approximate figures suggest two tendencies in Cuban contemporary Physics –at least the one published in some of the highest impact journals. First, the prevalence of “theoretical” investigation (defined as that where no direct experimental work is involved) Second, that the experimental work depends heavier on foreign collaboration than the theoretical one. An interesting analysis of these and other elements has been presented in an article recently appeared in the *Physics Today*, September 2006, p. 42.

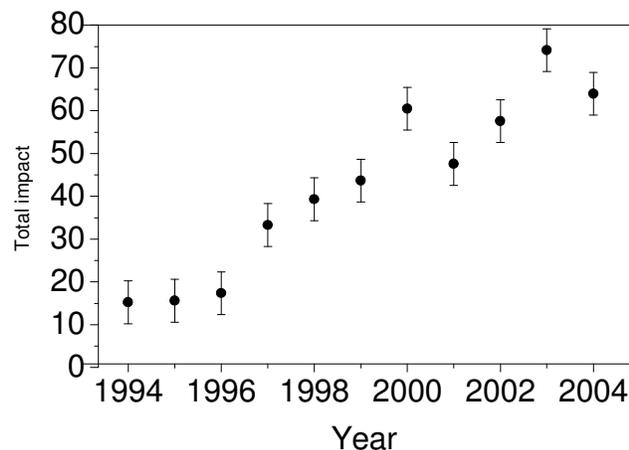
Although these tendencies are no surprising for the majority of my readers, here we have been able to quantify them, at least partially, with the help of bibliographic databases. Finally, it is worth noting that the great proportion of papers published by Cuban authors in *Physical Review B* certainly reflects the high volume of scientific work in Solid State Physics, specially semiconductor physics.

Figure 3 attempts to quantify the evolution of the *impact* of the publications accounted in Figure 2. The vertical axis corresponds to the “total average impact” for each year within the period 1994-2004, defined as:

$$FI_T = \sum_{i=1}^N FI_i$$

where  $N$  is the total number of articles published during a certain year, and  $FI_i$  is the average impact factor [4] of the journal where paper  $i$  was published. Notice that a systematic increase of the total impact of Cuban publications in the journals under analysis starts quite sharply in 1996. If we fit it to a straight line in the period 1996 - 2004, the slope is approximately 6 impact units per year. Some questions immediately arise: is the 1996’s “awakening” connected to the “recovery” of the Cuban economy from the lowest point of its crisis, known as “the special period in peace time”? Is it connected with the widening of the relations with “the West” that took place in the same period and holds until today? Or it is perhaps related to the increasing “expertise” of Cuban researchers in the “art” of publication in international scientific journals? Should we expect this tendency to continue within the next years?

Other elements related to the economy that conspire against scientific productivity is the lack of continuity of our own scientific publications, which makes difficult to fulfill the requirements to be indexed in international databases. To the author’s knowledge, the only Cuban journal indexed by ISI is the *Cuban Journal of Agriculture (Revista Cubana de Agricultura)*. The Cuban Journal of Physics (*Revista Cubana de Física*) was founded in 1981. After an period of irregularities during the 1990’s, it has been revitalized, including the preparation of a database with the totality of the numbers since its foundation.



**Figure 3.** Total impact of articles appeared in APS and AIP journals, signed by physicists working in Cuban institutions (period 1994-2004).

Another element indirectly related with economy that influences the impact of Physics can be posed as another question: to what extent the brain drain to the most developed countries influenceS the impact of publications by Third World countries in Physics?. As far as I know, there is no trustable statistics to construct an answer. But one can safely say that, within the last 20 years, it is easy to identify an increase in Chinese names, and, since the early 1990’s, in Russian names, on scientific papers signed by first-world institutions. In the case of Cuban Physics I would say that a glance at the list of authors of the papers included in Figure 2 allows identifying a number of names that are nowadays signing their papers as members of foreign institutions –mainly European, Brazilian or Mexican–, a situation that has increased during the last 10 years. However, there is a “hard core” of Cuba-based physicists that, in spite of many difficulties, has managed to keep an increasing presence in the world of high-impact scientific publications.

In the area of “non-economic” factors conspiring against the scientific productivity measured through publications in indexed journals, there are also “endemic prejudices”. One example is the relatively small importance given to publications in indexed

journals as a way to measure scientific productivity of individuals and institutions, mainly before the 1990's. That tendency is somehow associated to the strong will to drive scientific work towards *direct* economical applications, sometimes resulting in useful technological research of marginal scientific quality. It might be also linked to the difficulties to use the impact of scientific articles as a single standard for natural sciences and for social and economical sciences, where publication in indexed journals can be handicapped by ideological and political considerations. I do not subscribe, however, the idea that the production and impact of publications should be taken as the *only* measuring standard for scientific proficiency.

An extreme that illustrates how political issues can influence scientific publications from third world countries, is a recent "reinterpretation" of the U.S. embargo laws against a number of countries (Cuba amongst them) saying that manuscripts submitted to U.S. journals from those countries could not be edited by them. An editor would not, in principle, correct the grammar or spelling of a paper submitted, say, by an Iranian or Cuban author. In practice, it would mean freezing any paper from the embargoed countries submitted to the journal. Some Cuban manuscripts were reportedly frozen by indexed journals in the process. The decision was subject of protest by the international scientific community, as it was regarded an offense to scientific ethics (9), resulting in its dismissal, at least *de facto*. Differently from other scientific organizations, it should be said that the *American Physical Society*, the *American Institute of Physics*, and the *American Society for the Advancement of Science* (which publishes *Science* magazine) refused from the very beginning to accept the prohibition. My own experience is that APS never hesitated even during the editing process of its top-visibility journals, even at the worst of the "crisis".

Beyond the abyss between developed and developing countries, the differences in economical and political power amongst highly developed countries shows clear fingerprints in the impact of scientific journals, especially after II World War. In the field of Physics, even when the European Physics Society has tried to challenge the domination of *Physical Review Letters* by pushing forward its equivalent *Europhysics Letters*, the attempt has failed –at least for the time being– since the latter has been unable to surpass an impact factor of approximately half of that of *Physical Review Letters*. APS journal *Reviews of Modern Physics*, on the other hand, is well beyond any competitor in the field of Physics (and in most scientific fields), with an impact factor of 32.7 in 2004.

#### 4. QUANTIFYING THE INDIVIDUAL IMPACT

In "highly competitive" scientific markets, the individual impact measured through scientific publications is often used to hire professionals, to decide promotions and, eventually, to get research grants and projects –particularly international ones. In the case of Cuba, it is used as an element to grant prizes and small financial support through projects –sizable support for science is basically decided on other grounds, though. In any case, I believe that the scientific impact measured through scientific publications is a useful tool to self-evaluate the reach of one's scientific work. Table 3 –inspired on some elements reported in reference 10– includes a number of useful parameters that quantify the individual scientific impact, as well as some of their advantages and disadvantages:

Criteria in Table 3 would deserve an extensive analysis, but I will restrict myself to mention three examples indirectly related to the discussion. Few persons would hesitate to classify as "geniuses" writer Miguel de Cervantes and musician Joaquín Rodrigo. However, they earned such status *basically on a single work*: the novel popularly known as "El Quijote" and the musical piece "Concierto de Aranjuez". Something similar can be stated about physicist Luis de Broglie and his wave-particle duality.

I would add further criteria that might be useful in the Cuban context. There, where so many international collaborations take place, sometimes the role of the foreign counterpart does not reduce to the pure scientific collaboration or to provide part of the material support, but includes the control of the very publication process: elaboration of the manuscript, exchange with the editors, and response to the referees. Such tendency delays the development of publishing skills by Cuban scientists. So, a relevant index to quantify the "autonomy" of our authors (we might call it "scientific tropicality factor, STF"), would be the number of the articles where Cuban scientists are "corresponding authors", or the number of citations those papers have received.

With the aim to eliminate the disadvantages of many of the criteria included in Table 3, J. E. Hirsch has recently proposed a new parameter to measure individual scientific work: *the h-index* (10,11). *A scientist possesses an index h, if h of his (or her) papers have been cited, at least, h times* (the remaining papers have been cited less times). Figure 4 illustrates a way to visualize the h-index. The horizontal axis contains the number of papers by the author in question, starting by the one that has received a maximum number of citations. The vertical axis contains the number of citations corresponding to each of the papers in the horizontal

axis. Then, the area under the curve corresponds to the total impact of the articles published by the researcher. If now we draw a 45 straight line crossing the origin of coordinates, both the “x” and “y” coordinates of the intercept between the straight line and the curve correspond to the *h* index.

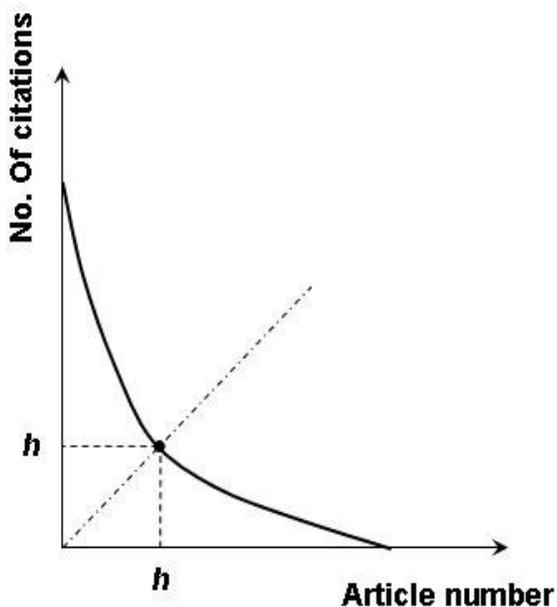
Assuming a “linear” model (i.e., the researcher produces articles of a similar impact, at a fixed rate along his (or her) career), it can be shown that the index *h* is given by (10):

$$h \sim mn$$

where *n* is the year, and *m* is a number that depends on the researcher. It means, of course, that it is *m*, the parameter that allows comparison between researchers of different ages. Within the linear model, *m* is given by (10):

$$m = \frac{c}{1 + \frac{c}{p}}$$

where *p* is the number of papers per year published by the researcher, and *c* is the number of new citations that each of those papers receives each year. As most linear models applied to “social” phenomena, this one pictures reality in an oversimplified fashion: a glance at curves in Figure 4 demonstrates how far from reality the model can be.



**Figure 4.** Visualizing the *h*-index.

Beyond models, straight calculations of *m* and *h* based on hard data give encouraging results about their capacity to quantify the quality and quantity of individual scientific work. In fact, until August 2005, the top *m* and *h* indices in Physics corresponded to Ed Witten (*h* = 110, *m* = 3.89). Following Hirsch, *m* ~ 1 (i.e., *h* ~ 20 after 20 years of scientific work) is a

fingerprint of a successful scientist, while *m* ~ 3 (i.e., *h* ~60 after 20 years of scientific work) corresponds to truly unique individuals.

A further element that supports the choice of *m* and *h* indices is that 84% of Nobel winners in Physics with papers published in journals indexed by ISI have a *h* factor of, at least, 30. Following Hirsch (10) it indicates that the Nobel Prize is not just the result of a struck of luck, but of sustained scientific work.

## CONCLUSION

In my opinion, the advantages to quantify the scientific impact based on the publications in indexed journals are beyond any doubt. Cuban scientists and, specially, Cuban physicists, are encouraged to increase the quantity and quality of their contributions to indexed journals, and then increase the relative position of our country in that field, at least within the Latin American context. Considering the material limitations in which our scientific work takes place, the only way is to put on it an extra dose of originality, and to increase our professionalism in the art of publication of new results: sensible paper writing, wise selection of the target journal, and intelligent management of the editorial process.

Although the impact factor and other bibliographic indices are unavoidable parameters to measure the real impact of science, they cannot be taken as the only way to measure the scientific results of an individual, and institution, or a country. In the latter case, other numbers as the % of the National Gross Product devoted to research and the number of scientists per capita should be also taken into account<sup>1</sup>. Other non-quantifiable factors such as the social or cultural impact of a scientific discovery are extremely relevant<sup>2</sup>.

Finally, I would like to stress that this article does not pretend to constitute an exhaustive bibliometric study of the impact of Cuban Physics: for one thing, the sample of publications under analysis has been very limited. It would be excellent if a real expert performs a definitive study on the subject, which would surely contribute to evaluate the international impact of Cuban Physics, and its perspectives in the short and medium terms.

## ACKNOWLEDGMENTS

The author thanks the “Abdus Salam” ICTP for supporting the writing of the first version of this manuscript. M. Sánchez-Colina is acknowledged for pushing me to write this paper and for the revision of the manuscript. The detailed revision of the manuscript by A. R. Ruiz-Salvador, and his search for extra data included in the Appendix, greatly improved this article.

<b>CRITERIA</b>	<b>ADVANTAGES</b>	<b>DISADVANTAGES</b>
Total number of papers published in indexed journals	Measures lifetime productivity	It does not measure the true impact of papers, it does not allow to compare scientists from different ages
Total number of citations of published papers	Measures total impact in a lifetime	Can be “inflated” with a few highly cited papers, where the scientist may have had small contribution. It does not allow to compare scientist from different ages
Number of citations per paper	Allows to compare scientists from different ages	Penalizes high productivity (assuming that producing a small number of papers with very high impact is a bad idea!)
Number of “most relevant” papers	It concentrates attention on what is important	The quality of the papers taken as “relevant” is arbitrary. In general, the selection criteria must be adapted somehow to the age of the scientist. It does not give a true image of the whole scientific career of the individual
Number of citations of each of the “most relevant” papers	It concentrates attention on the impact of relevant scientific work	Besides the disadvantages of the former criteria, this one is not described by a single number

#### APPENDIX

#### Impact of some Physics and Chemistry journals within the period 2000-2004<sup>3</sup>

<b>Journal</b>	<b>IF2000</b>	<b>IF2001</b>	<b>IF2002</b>	<b>IF2003</b>	<b>IF2004</b>	<b>&lt;IF00-04&gt;</b>
REV MOD PHYS	12.774	12.762	23.672	28.172	32.771	<b>22.030</b>
ANNU REV ASTRON ASTR				16.000	18.839	
SURF SCI REP	14.952	14.091	13.238	12.650	21.35	<b>15.256</b>
ADV PHYS	13.611	16.200	13.952	13.087	15.333	<b>14.437</b>
PROG MATER SCI*	4.667	14.000	11.600	12.00	10.467	<b>10.547</b>
ANNU REV PHYS CHEM	9.237	7.609	10.255	10.500	11.944	<b>9.909</b>
SOLID STATE PHYS	9.250	9.667	6.600	7.000	16.000	<b>9.703</b>
MAT SCI ENG R	6.083	6.143	11.893	10.032	14.233	<b>9.677</b>
PHYS REP	7.110	8.341	12.645		14.742	
REP PROG PHYS	9.000	8.879	7.618	8.409	7.842	<b>8.349</b>
MASS SPECTROM REV	7.600	8.391	6.750	7.364	8.743	<b>7.769</b>
ANNU REV NUCL PART S*	5.040	6.690	7.179	8.667	7.739	<b>7.063</b>
ADV NUCL PHYS	4.667	6.667	10.571	8.750	4.500	<b>7.031</b>
PHYS REV LETT	6.462	6.668	7.323	7.035	7.218	<b>6.941</b>
ADV MATER*	5.522	5.579	6.801	7.305	8.079	<b>6.657</b>
J HIGH ENERGY PHYS	4.196	8.664	6.854	6.057	6.503	<b>6.455</b>
ANNU REV FLUID MECH*	6.486	5.486	6.450	5.108	6.694	<b>6.045</b>
PROG NUCL MAG RES SP*	5.062	7.192	4.808	5.971	6.885	<b>5.984</b>
NUCL PHYS B*	4.225	6.226	5.409	5.297	5.819	<b>5.395</b>
J PHYS CHEM REF DATA	8.756	4.488	3.333	4.000	4.788	<b>5.073</b>
ADV ATOM MOL OPT PHY*	4.941	4.576	4.524	4.107	7.214	<b>5.072</b>
<b>PHYS TODAY*</b>	<b>5.298</b>	<b>4.790</b>	<b>5.000</b>	<b>5.020</b>	<b>5.211</b>	<b>5.064</b>

<sup>1</sup>A recent study by RAND based on these and other parameters classified Cuba as a “scientifically proficient” country (12).

<sup>2</sup>A beautiful example that shows how top-quality science combines social usefulness with bibliometric impact, is the recent development of a fully-artificial vaccine by a Cuban team of researchers (*Science* 305 (2004) 522).

<sup>3</sup>The symbol <IF00-04> corresponds to the impact factor averaged over the period 2000-2004. Only journals with <IF00-04> higher than 5.00 have been included in the tables.

## Chemistry

Journal	IF2000	IF2001	IF2002	IF2003	IF2004	<IF00-04>
CHEM REV	20.036	21.044	20.993	21.036	20.233	<b>20.668</b>
SURF SCI REP	14.952	14.091	13.238	12.650	21.35	<b>15.256</b>
ACCOUNTS CHEM RES	13.262	12.781	15.901	15.000	13.154	<b>14.019</b>
PROG MATER SCI*	4.667	14.000	11.600	12.000	10.467	<b>10.547</b>
ANNU REV PHYS CHEM	9.237	7.609	10.255	10.500	11.944	<b>9.909</b>
CHEM SOC REV	10.747	9.137	8.718	9.569	10.836	<b>9.801</b>
MAT SCI ENG R	6.083	6.143	11.893	10.032	14.233	<b>9.677</b>
PROG INORG CHEM	10.714	16.500	4.286	8.500	7.200	<b>9.44</b>
ADV CATAL	11.000	6.846	10.923	7.889	9.75	<b>9.282</b>
ANGEW CHEM INT EDIT	8.547	8.255	7.671	8.427	9.161	<b>8.412</b>
MASS SPECTROM REV	7.600	8.391	6.750	7.364	8.743	<b>7.769</b>
CATAL REV*	6.562	8.471	6.455	5.708	9.750	<b>7.389</b>
ALDRICHIM ACTA*	5.900	7.846	6.333	7.077	8.833	<b>7.198</b>
ADV ORGANOMET CHEM	9.588	7.417	5.467	7.200	5.500	<b>7.034</b>
ADV MATER*	5.522	5.579	6.801	7.305	8.079	<b>6.657</b>
ADV INORG CHEM	11.545	9.567	3.933	4.095	3.769	<b>6.582</b>
NAT PROD REP*	5.295	5.772	5.900	7.529	7.89	<b>6.477</b>
J AM CHEM SOC*	6.025	6.079	6.201	6.516	6.903	<b>6.345</b>
PROG POLYM SCI*		3.738	7.279	7.759	8.482	
ADV POLYM SCI*		6.053	5.389	6.955	7.320	
PROG NUCL MAG RES SP*	5.062	7.192	4.808	5.971	6.885	<b>5.984</b>
TOP CURR CHEM*	5.960	5.800	5.000	5.784	5.283	<b>5.565</b>
COORDIN CHEM REV*	3.763	5.224	5.853	5.951	6.446	<b>5.447</b>
<b>J PHYS CHEM REF DATA</b>	<b>8.756</b>	<b>4.488</b>	<b>3.333</b>	<b>4.000</b>	<b>4.788</b>	<b>5.073</b>

## REFERENCES

- 1) AMIN, M. and M. MABE: (2000): "Impact factors: use and abuse", in **Perspectives in Science Publishing Newsletter**, Elsevier Science.  
([http://www.elsevier.com/framework\\_editors/pdfs/Perspectives1.pdf](http://www.elsevier.com/framework_editors/pdfs/Perspectives1.pdf))
- 2) American Physical Society Newsletter: Physical Review Letter's Top Ten (2003)  
(<http://www.aps.org/apsnews/topten.cfm>) (2003)
- 3) Institute of Scientific Information (ISI) essays: The Impact Factor  
(<http://www.isinet.com/essays/journalcitationreports/7.html/>)
- 4) POPESCU, I.I. (2002): Science Journal Ranking by Average Impact Factors  
([http://alpha2.infim.ro/~ltpd/Jo\\_rankingb.htm](http://alpha2.infim.ro/~ltpd/Jo_rankingb.htm))
- 5) HAZEN, R. M. (1989): The breakthrough: the race for the Superconductor. Ballantine Books,
- 6) ALTSHULER, J. (2003): Ciencia Aria y Ciencia judía, en "A propósito de Galileo", Fondo de Cultura Económica.
- 7) GIBBS, W. W. (1995): Trends: Lost science in the Third World. **Scientific American** 273 92
- 8) Latin America: A growing presence. **Science Watch**, Sept.-Oct, 2001.  
([http://www.sciencewatch.com/sept-oct2001/sw\\_sept oct2001\\_page1.htm](http://www.sciencewatch.com/sept-oct2001/sw_sept oct2001_page1.htm))
- 9) BRUMFIEL, G. (2004): Publishers split over response to U.S. trade embargo ruling. **Nature** 427, 663.  
([http://www.nature.com/nature/journal/v427/n6976/full/427663a\\_fs.html](http://www.nature.com/nature/journal/v427/n6976/full/427663a_fs.html))
- 10) HIRSCH, J.E.: An index to quantify an individual's scientific research output.  
(<http://arxiv.org/abs/physics/0508025>)
- 11) BALL, P. (2005): Index aims for fair ranking of scientists. **Nature** 438, 900.
- 12) WAGNER, C.S.; I. BRAHMAKULAM; B. JACKSON; A. WONG and T. YODA. (2001): Science and Tech. Collaboration: Building Capacity in Developing Countries?, RAND Science and Tech. Report.

# INFLUENCIA DEL RETARDO EN UN TUMOR SOMETIDO A UN TRATAMIENTO PERIÓDICO

Ramón Espinosa Riquelme<sup>1</sup>, Daniel Rodríguez Pérez<sup>1</sup>, Oscar Sotolongo-Costa<sup>2\*</sup>,  
J. Antonio Santos Miranda<sup>3</sup> y José Carlos Antoranz<sup>1,2</sup>

<sup>1</sup>Departamento de Física Matemática y Fluidos, UNED, Madrid, España

<sup>2</sup>Cátedra "Henri Poincaré" de Sistemas Complejos, Universidad de La Habana, Cuba

<sup>3</sup>Servicio de Oncología, Radioterapia. H.G.U. "Gregorio Marañón", Madrid, España

## RESUMEN

Se presenta un modelo de interacción entre el sistema inmunitario y una población de células tumorales. Se estudia la estabilidad del sistema cuando se añade un tratamiento periódico con citoquinas. En dicho modelo se ha tenido en cuenta el efecto producido por el tiempo transcurrido desde que el sistema inmune advierte la presencia del tumor, hasta que actúa contra él (retardo). Se comparan los resultados obtenidos con y sin retardo. En general, el retardo introduce inestabilidades en algunas soluciones del sistema respecto de las correspondientes al sistema de respuesta instantánea.

## ABSTRACT

We model the interaction between the immune system and tumor cells under a periodic therapy, when there is a delay between the triggering of the immune system and the detection of the tumor cells. The results are compared with those of a previous paper, concluding that the delay introduces instabilities in some solutions of the system relative to those of the undelayed system.

## 1. INTRODUCCIÓN

La lucha contra los tumores malignos es, en el momento actual, una de las tareas más importantes de la ciencia, dada su alta mortalidad y su cada vez mayor incidencia en la población, en lo que incide el alargamiento de la esperanza de vida y factores de riesgo como el consumo creciente de diversas drogas y factores medioambientales, genéticos y otros. Esta tarea es una de las que más preocupa a la sociedad actual y a los profesionales de la salud debido a que el surgimiento de tumores tiene un carácter en muchos casos impredecible y también a que su tratamiento resulta muy complicado, al trabajar sobre células que han escapado de la vigilancia del sistema inmunitario del individuo.

La inmunoterapia es una de las vías que se ha escogido para el tratamiento de los tumores malignos, y se define como la utilización del sistema inmunitario y de sus productos para controlar o destruir las células tumorales. La inmunoterapia actúa estimulando a las células inmunitarias que deben encargarse de eliminar o, al menos, mantener bajo control al tumor. Los llamados linfocitos-T son los encargados de efectuar el trabajo, pero para ello otras células, en general conocidas como células presentadoras de antígenos, deben activarlos; su proliferación requiere a menudo una coestimulación mediada por unas hormonas llamadas citoquinas [1, 2].

Es obvio que en este proceso debemos considerar, como un parámetro fundamental, el tiempo transcurrido desde que el sistema inmune se percata de que hay células extrañas hasta que los linfocitos-T actúan sobre ellas. Consideramos este retardo como el parámetro de control en el sistema a estudiar [3, 4]. En [5] se presentó un modelo de interacción entre el tumor y el sistema inmune que incluía conjuntamente la acción depresora del tumor sobre el sistema inmunológico y la acción de una terapia periódica externa con citoquinas. Dicho modelo arrojó como resultado diferentes comportamientos dinámicos que concuerdan cualitativamente con observaciones clínicas.

En nuestra opinión, el tiempo que media entre la presencia del antígeno y la reacción defensiva del organismo no ha sido considerado suficientemente en la literatura, a pesar de haber resultados de mucho interés [6, 7] donde al modelo propuesto en [8] se le incluye dicho efecto y se demuestra su importancia en la modelación del sistema inmune. Así, el presente trabajo está basado en el modelo inicial, ya publicado en [5], al cual se le ha añadido una interacción con memoria al término que determina la actuación del sistema inmune contra el tumor. Como veremos posteriormente, este estudio se realiza en el sistema sin forzar (sin tratamiento) y en el sistema forzado (con tratamiento periódico). Por último, se han estimado los rangos de variación de los parámetros biológicos que determinan las constantes del sistema

y se realizaron simulaciones numéricas sólo dentro de dicho rango. Sin duda, la variedad de soluciones del sistema es más rica de lo que se muestra en este trabajo pero no corresponderían a soluciones biológicas para este modelo.

## 2. EL MODELO

### 2.1 Construcción del modelo

El modelo propuesto inicialmente en [5] está basado en un modelo de competición de dos especies, de Lotka-Volterra, al que se le añaden algunos términos adicionales:

$$\begin{aligned} \frac{dx}{dt} &= aX - bXY \\ \frac{dY}{dt} &= dXY - fY - kX + u + F \cos^2 wt \end{aligned} \quad (1)$$

donde  $X(t)$ , y  $Y(t)$  son las poblaciones de células tumorales y linfocitos-T, respectivamente. La tasa de crecimiento de las células tumorales es proporcional a  $X$  y la tasa de mortalidad es proporcional a la frecuencia de interacción con los linfocitos. La tasa de crecimiento de los linfocitos es proporcional a la interacción con las células tumorales y al flujo por unidad de tiempo de linfocitos a la zona de localización del tumor ( $u$ ). Su tasa de mortalidad está dada por dos términos: muerte natural de linfocitos ( $-fY$ ) y aumento de la masa tumoral ( $-kX$ ). Una característica de este sistema es que la inmunodepresión, expresada a través de este término, se ha tomado en cuenta en igual forma que en [5], es decir, de la forma más sencilla (y probablemente muy cruda). La inmunodepresión puede llevar a consecuencias fatales ya que puede aniquilar la totalidad del sistema inmune, lo que se expresa como  $Y(t) = 0$ . En este modelo la población de linfocitos puede llegar a tomar valores negativos una vez anulada, de modo que en este caso no tiene sentido continuar la integración.

Para la introducción de los efectos producidos por el tratamiento periódico con citoquinas, elegimos la función  $F \cos^2 wt$ , como primer modelo de función periódica positiva, donde  $w$  es la frecuencia de aplicación del tratamiento y  $F$  el valor neto de la dosis.

El término  $dXY$  es especialmente importante para este estudio. Describe la interacción entre las dos poblaciones,  $X$  y  $Y$ , y  $d$  es la frecuencia con que son reconocidas las células malignas por el sistema inmunológico. Incluiremos el efecto del retardo al igual que en [6], que es la generalmente aceptada en la literatura. En [6] puede hallarse referencias a otros trabajos. De acuerdo con lo dicho en la sección anterior, introducimos el tiempo de retardo en el argumento de las funciones  $X$  y  $Y$  de este término quedando el nuevo sistema:

$$\frac{dX}{dt} = ax - bXY \quad (2)$$

$$\frac{dY}{dt} = dX(t-T)Y(t-T) - fY - kX + u + G(t)$$

donde  $G(t) = F \cos^2 wt$ , y  $T$  es el tiempo que tarda el sistema inmunológico en actuar sobre las células tumorales una vez reconocidas. El tiempo de retardo no aparece en el otro término de interacción,  $bXY$ , ya que este término representa la acción directa de los linfocitos-T, ya formados, contra el tumor y, por tanto, se produce en tiempo  $t$ .

Si  $F = k = u = 0$ , recuperamos el modelo clásico de Lotka-Volterra, cuyo tiempo característico es  $t_c = 1/\sqrt{af}$ , el cual utilizamos para adimensionalizar el sistema (2), en la forma:

$$\frac{dx}{d\tilde{t}} = \alpha x - xy \quad (3)$$

$$\frac{dy}{d\tilde{t}} = x(\tilde{t} - \tau)y(\tilde{t} - \tau) - \frac{1}{\alpha}y - \kappa x + \sigma + H(\tilde{t})$$

donde

$$\tilde{t} = t \cdot t_c \quad x = \frac{Xd}{\sqrt{af}} \quad y = \frac{Yb}{\sqrt{af}}$$

y

$$\kappa = \frac{kb}{d\sqrt{af}} \quad \alpha = \sqrt{\frac{a}{f}} \quad \sigma = \frac{ub}{af} \quad V = \frac{Fb}{af} \quad \beta = \frac{w}{\sqrt{af}}$$

con  $H(\tilde{t}) = V \cos^2 \beta \tilde{t}$ . A partir de ahora  $\tilde{t} = t$ .

### 2.2. Estimación de algunos parámetros

Para saber qué rangos emplear en las simulaciones numéricas, es de gran utilidad estimar o al menos conjeturar los órdenes de magnitud de los parámetros que hay en el sistema. Tomamos como unidad: células, para  $X$  y  $Y$ , y días para  $t$ . Se sabe que el tiempo medio de duplicación de un tumor sólido está en torno a los 70 días en media [9]. Obviamente esta media no es muy significativa. Podemos encontrar tumores con un tiempo de doblaje de 20 días y otros de 100 días y aún mucho mayores., por lo que este valor será útil sólo para tener una idea de los órdenes de magnitud e intentar trabajar numéricamente con valores que no estén muy alejados de la realidad.

Como en ausencia de linfocitos hemos supuesto un crecimiento exponencial, tenemos que el parámetro  $a = \text{Ln}2/T_{\text{doblaje}}$  estará medido en días<sup>-1</sup>. De modo que  $a$  queda acotado en el intervalo conservador  $a \in [10^{-4}, 10^{-3}]$  días<sup>-1</sup>. Experimentalmente, pueden

encontrarse valores para el tiempo de vida de los linfocitos, siendo éste de unos 30 días [8]; por lo tanto podemos suponer  $f \in [10^{-2}, 10^{-1}]$  días<sup>-1</sup>. El tiempo que los linfocitos toman para proliferar es entre 2 y 12 días. [7]. Podemos tomar este tiempo como el retardo del sistema inmune

Suponiendo que hay un estado estacionario establecido, en ausencia de tumor, la segunda ecuación del modelo muestra en el estado inicial:  $u - fY = 0$ . Se sabe que en el bazo de un animal de laboratorio hay aproximadamente  $10^8$  linfocitos, pero sólo una parte de ellos estarán en condiciones de ser linfocitos-T citotóxicos. Asumimos que este porcentaje es un 3% [3, 6, 10], quedando  $Y_0 \approx 3 \cdot 10^5$  células.

Con estos datos, estimamos que  $u \approx 1.2 \cdot 10^4$  células-día<sup>-1</sup>. Otros parámetros del sistema se estimarán como en [8, 10], de forma que  $b \in [10^{-9}, 10^{-7}]$  (células.día)<sup>-1</sup>, y  $d$ , del mismo orden que  $b$  y medido en las mismas unidades. Por último, en el sistema (2) podemos suponer que hay un estado estacionario establecido en ausencia de células del sistema inmunitario, y por tanto  $k \approx 1.2 \cdot 10^{-3}$  días<sup>-1</sup>.

Todos estos estimados concuerdan con valores particulares reportados en la literatura ([11] y referencias contenidas).

Esto lleva a los siguientes rangos de valores para los parámetros del sistema (3):

$$\alpha \in [10^{-2}, 1], k \in [10^{-2}, 10], \sigma \in [10^{-2}, 10^3], \tau \in [10^{-3}, [10^2], x_0 \in [10^{-3}, 10^2].$$

y en los cálculos tomaremos para estos parámetros valores típicos que se enmarquen en estos rangos.

### 3. RESULTADOS

#### 3.1. Resultados conocidos para el sistema sin retardo y sin tratamiento ( $G(t) = 0$ )

Resumimos ahora los resultados conocidos para el sistema (1), que en forma adimensional y con  $G(t) = 0$  se escribe:

$$\frac{dx}{dt} = \alpha x - xy \tag{4}$$

$$\frac{dy}{dt} = xy - \frac{1}{\alpha} y - \kappa x + \sigma$$

(para detalles, véase [5]). La estabilidad de los estados estacionarios de este sistema puede resumirse así:

- Para el estado  $L_0 = (0, \alpha \sigma)$

Si  $\sigma > 1 \Rightarrow$  Nodo ESTABLE

Si  $\sigma < 1 \Rightarrow$  Punto silla INESTABLE (E1)

- Para el estado  $L_1 = \left( \frac{1-\sigma}{\alpha-\kappa}, \alpha \right)$

INESTABLE (SILLA) para  $\frac{\kappa}{\alpha} > \sigma > 1 \quad \sigma > \frac{\kappa}{\alpha} > 1$  (E2)

ESTABLE (NODO o FOCO) para  $\frac{\kappa}{\alpha} < \sigma < 1$  (E3)

INESTABLE (NODO o FOCO) para  $\sigma < \frac{\kappa}{\alpha} < 1$  (E4)

#### 3.2. Estabilidad del sistema con retardo y sin tratamiento ( $G(t) = 0$ )

El sistema (2), en ausencia de tratamiento y adimensionalizado toma la forma:

$$\frac{dx}{dt} = \alpha x - xy \tag{5}$$

$$\frac{dy}{dt} = x(t-\tau)y(t-\tau) - \frac{1}{\alpha} y - \kappa x + \sigma$$

Los siguientes resultados de tipo analítico sobre la estabilidad de los estados estacionarios, han sido probados mediante las técnicas habituales utilizadas en sistemas de ecuaciones diferenciales con retardo [12-14].

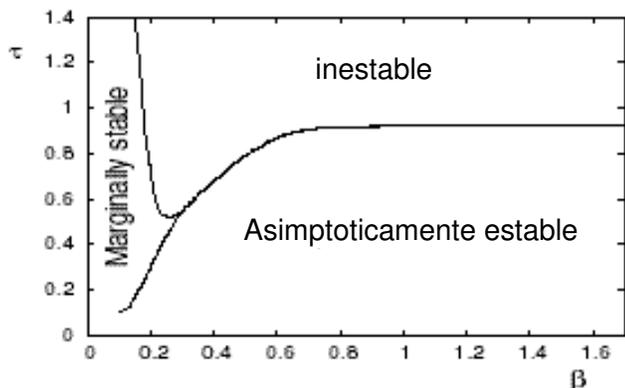
- El estado estacionario  $L_0$  no cambia su estabilidad debido al retardo. Es decir, se cumple (E1)  $\forall \tau \geq 0$ .
- El estado estacionario  $L_1$  se mantiene inestable para todo  $\tau \geq 0$ , en los casos dados por (E2).
- Para el rango de estabilidad dado por (E3), existe un valor de  $\tau$  (designémoslo por  $\tau_c$ ), tal que el estado estacionario  $L_1$  se vuelve inestable. Para el rango de valores de los parámetros dado por (E4), el estado estacionario  $L_1$  mantiene su inestabilidad.

Además, podemos calcular el valor concreto del retardo para el cual encontramos el cambio de estabilidad. Usando las expresiones expuestas en [12], se puede hallar  $\tau_c$  en función, por ejemplo, del parámetro de crecimiento del tumor. Igual puede hacerse para los otros parámetros  $k$  y  $\sigma$ . Un análisis más detallado de esto y la dinámica del sistema será publicado posteriormente.

#### 3.3. Resultados para el sistema con retardo, cuando se añade el tratamiento

Integrando numéricamente el sistema (3), se obtuvieron los valores de la frecuencia y la amplitud, para los cuales el sistema es inestable, y por tanto el tumor crece de forma incontrolada, y también los valores de dichos parámetros para los cuales el sistema evoluciona favorablemente, es decir, el estado estacionario  $L_1$  es estable, y la población de células tumorales está controlada.

Mediante la integración numérica del sistema con retardo (3), se obtienen los valores de la frecuencia del tratamiento y del retardo para los que se obtiene un sistema estable, o inestable, fijando la amplitud del tratamiento en  $V = 0.5$ . La inestabilidad del sistema se corresponde con un crecimiento de la masa tumoral incontrolable, perdiéndose el equilibrio entre ambas poblaciones. La estabilidad del sistema corresponde a una masa tumoral controlada por el tratamiento periódico con citoquinas (véase Figura 1).



**Figura 1.** Representación en el espacio de parámetros  $\beta$ ,  $\tau$ , de los puntos de inestabilidad, y estabilidad. Los valores de los otros parámetros son:  $\kappa = 0.95$ ,  $\sigma = 0.5$ ,  $\alpha = 1.5$ ,  $V = 0.5$ , todos dentro del rango fisiológico estimado. Las condiciones iniciales fueron  $x_0 = 0.1$ ,  $y_0 = 0.75$ . Las líneas delimitan las regiones de estabilidad asintótica donde el tamaño del tumor permanece controlado después de  $t = 100$ . Observar la región "marginamente estable" donde el crecimiento del tumor ocurre después de un  $t_{max} = 10$ .

Las curvas de la Figura 1 dividen al plano  $(\beta, \tau)$  en tres regiones. Por debajo de una de las curvas obtenemos la región donde el sistema permanece estable en un tiempo finito y es biológicamente aceptable. Por encima de la otra curva el sistema es inestable en tiempo finito (menor que un  $t_{max}$  escogido), lo que significa la anulación total del sistema inmunológico y el crecimiento incontrolable del tumor. Existe una curiosa zona intermedia, que llamaremos de "estabilidad marginal", entre ambas curvas que correspondería a puntos de inestabilidad, pero estas inestabilidades se dan en tiempos muy largos que biológicamente carecen de significado. Ésta es una región difícilmente clasificable, aunque pudiese ser considerada como región de estabilidad ya que el tiempo para el que se desestabiliza la solución es del orden de las decenas de años,

lo que no tiene valor práctico a los efectos clínicos. El límite entre la región marginalmente estable e inestable depende fuertemente de las condiciones iniciales y de  $V$ .

#### 4. CONCLUSIONES

De los resultados obtenidos en la sección anterior, podemos concluir que la introducción de un término de retardo, enriquece la dinámica del sistema. Debido a la introducción de este parámetro, el sistema puede volverse inestable. Fijando el retardo que induce a dicha inestabilidad, y variando la amplitud y la frecuencia del tratamiento, obtenemos un sistema estable, es decir, un control absoluto del tratamiento sobre el crecimiento del tumor. Así, el conocimiento de los tiempos de retardo y los parámetros del sistema permite incrementar la efectividad de los tratamientos con citoquinas.

La eficacia del tratamiento no lleva a una anulación completa de la masa tumoral, sino que mantiene un estado controlado del tumor, permitiendo la introducción de otros métodos (cirugía, radiación, etc).

Las simulaciones numéricas realizadas para el sistema con tratamiento y con retardo, muestran en el plano  $(\beta, \tau)$  una curva que separa la zona de inestabilidad de la zona de estabilidad. Aunque, para valores bajos del retardo, también hay una zona inestable que será menor cuanto mayor sea  $V$ .

Es muy importante notar que para algunos valores de la frecuencia de aplicación de las citoquinas hay valores del retardo que hacen inútil el empleo de este tipo de inmunoterapia. Este modelo indica, como ya se ha visto, que la aplicación de la terapia con citoquinas requiere una valoración de los tiempos de retardo en el sistema inmune del organismo en su comparación con los tiempos característicos del sistema, que en última instancia vienen aquí expresados en unidades del tiempo del sistema Lotka-Volterra.

La dinámica del sistema en el caso de la presencia del retardo es muy variada, y abre un gran campo de trabajo, en este ya dilatado dominio de la dinámica de poblaciones. Puede además escogerse unidades diferentes para adimensionalizar el sistema o considerar el retardo como perturbación, lo que simplifica algunos análisis, o permite elaborar símiles mecánicos como daremos a conocer en trabajos posteriores.

#### REFERENCIAS

[1] TORTORA, G.J. and S. R. GRABOWSKI (1999): Principles of Anatomy and Physiology, Tenth Edition, Oxford.

[2] FLAVIO SALAZAR, O. (2000): "El sistema inmune, herramienta estratégica en la batalla contra el cáncer", **Revista Chilena de Pediatría** 71.

- [3] FORYS, U. (2002): "Marczhuck's model of immune system dynamics with application to tumor growth", **Journal of Theoretical Medicine** 4, 85.
- [4] BODNAR, M. and U. FORYS (2003): "Time delays in proliferation process for solid avascular tumor", **Mathematical And Computation Modelling** 37, 1201.
- [5] SOTOLONGO-COSTA, O. et al. (2003): "Behavior of tumors under nonstationary therapy", **Physica D** 178, 242.
- [6] GALACH, M. (2003): "Dynamics of the tumor -immune system competition- the effect of time delay", **Int. J. Appl. Math. Comp. Sci.** 13, 395.
- [7] BYRNE, H.M. (1997): "The effect of time delays on the dynamics of avascular tumor growth", **Math. Bios.** 144, 83.
- [8] KUZNETSOV, V.A.; I. MAKALKIN; M.A. TAYLOR and A.S. PERELSON (1994): "Nonlinear dynamics of immunogenic tumors: parameter estimation and global bifurcation analysis", **Bulletin of Mathematical Biology** 56, 295.
- [9] BEGG, A. and G.G. STEEL (1977): "Cell proliferation and growth rate ", Oxford.
- [10] KIRSCHNER, D. and J.C. PANETTA (1998): "Modelling immunotherapy of the tumor - immune system interaction", **Journal of Mathematical Biology** 38, 235.
- [11] PILLIS, L.G. de; A.E. RADUNSKAYA and C. L. WISEMAN (2005): "Validated mathematical model of cell-mediated immune response to tumor growth", **Cancer Research** 65, 7950.
- [12] BERETTA, E. and Y. KUANG (2002): "Geometric stability switch criteria in delay differential systems with delay dependent parameters", **SIAM Journal on Mathematical Analysis** 33, 1144.
- [13] HALE, J. (1977): **Introduction to functional differential equations**, vol. 99 of Applied Mathematical Sciences, Springer-Verlag, New York, Incorporated.
- [14] GOPALSAMY, K. (1992): **Stability and oscillations in delay differential equations of populations dynamics**, Kluwer Academic Publishers, Dordrecht.

# BIOMAGNETISMO vs. MAGNETOBIOLOGÍA

A. González Arias, Dpto. Física Aplicada, Universidad de La Habana<sup>1</sup>

## RESUMEN

Bajo el título *Biomagnetismo* es posible encontrar actualmente referencias muy disímiles en la literatura científica; por ejemplo, un estudio serio y profundo sobre la posible interacción de las palomas mensajeras con el campo magnético terrestre, y también un comercial sobre una supuesta terapia magnética para aliviar el dolor, de muy dudosa efectividad. También el biomagnetismo puede aparecer ligado a la "bioenergía", término muy abusado y distorsionado en la literatura pseudocientífica. Una revisión de la literatura científica sobre el tema conduce a que el biomagnetismo aparece asociado, cada vez más de manera exclusiva, al estudio de los campos electromagnéticos de baja frecuencia generados por el cerebro y el corazón humanos. En lo que sigue se analizan las diferencias entre biomagnetismo y magnetobiología. Este último término ha sido acuñado recientemente para designar el estudio de la interacción de diversos animales con el campo magnético terrestre.

## ABSTRACT

Nowadays, under the title *Biomagnetism*, it is possible to find very dissimilar references in the scientific literature; for instance, a serious and insightful study about the possible interaction of pigeons with the earth magnetic field, and also some commercial advertising about a very doubtful magnetic therapy for removing pain. Biomagnetism may also appear related to "bioenergy", a very abused and distorted term in pseudoscientific literature. A review on the subject leads to the conclusion that in scientific literature the biomagnetism appears mostly related to the study of the low-frequency electromagnetic fields generated by the human heart and brain. In the following, the differences between biomagnetism and magnetobiology are analyzed. The last term has been coined recently to design the study of the interaction of different animals with the earth magnetic field.

## 1. INTRODUCCIÓN

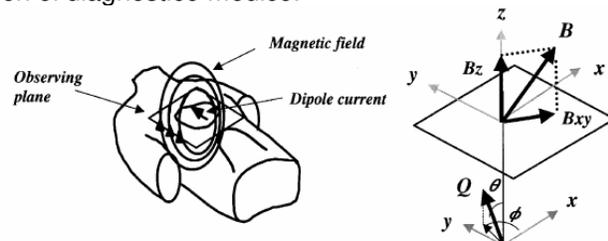
Bajo el título *biomagnetismo* es posible encontrar actualmente referencias sobre temas muy disímiles, cuyo denominador común es la interacción del campo magnético con los organismos vivos. Así, bajo este título puede aparecer tanto un comercial sobre una supuesta terapia magnética para aliviar el dolor, de dudosa efectividad, como también un estudio mucho más serio sobre la posible interacción de las palomas mensajeras con el campo magnético terrestre. Cuando la interacción biomagnética se refiere a aplicaciones médicas, es preferible referirse a ella como una *terapia magnética* (o, mejor aún, *electromagnética*, por ser esta última la única reconocida universalmente por la comunidad científica)<sup>1</sup>. Y para designar la investigación de la sensibilidad de las palomas y otros organismos vivos al campo magnético terrestre se ha reservado el término *magnetobiología*.

El término *biomagnetismo* se tiende a asociar cada vez más de manera exclusiva al estudio de los campos electromagnéticos de baja frecuencia generados por el cerebro y el corazón humanos.

## 2. BIOMAGNETISMO

El biomagnetismo es un método de diagnóstico y no una terapia. No sirve para curar, sino para diagnosticar.

Las funciones cerebrales y cardiacas están relacionadas con la presencia de corrientes eléctricas muy débiles, que van acompañadas de campos eléctricos y magnéticos. Cuando las corrientes son variables o pulsantes, los campos se encuentran en forma de *radiación electromagnética*: ondas invisibles formadas por la interacción de los campos eléctrico y magnético que se transforman uno en otro continuamente y se propagan en todas direcciones. La componente eléctrica se puede medir con electrodos colocados en la piel; se llega así al *electrocardiograma* y al *electroencefalograma*, de uso corriente en el diagnóstico médico.



**Figura 1.** Diagrama esquemático de la dirección relativa de los campos y dipolos magnéticos asociados a las corrientes en el corazón humano<sup>2</sup>.

El biomagnetismo trata de las mediciones de la componente *magnética* asociada a estas corrientes. En este caso no se utilizan electrodos, sino un *magnetómetro SQUID* (del inglés *Superconducting Quantum Interference Device*), instrumento capaz de

<sup>1</sup> [arnaldo@fisica.uh.cu](mailto:arnaldo@fisica.uh.cu)

medir campos magnéticos increíblemente pequeños, de intensidad 100 millones de veces menor que el campo magnético terrestre.

En la tabla siguiente se muestran, de mayor a menor, algunos valores de la intensidad de la inducción magnética pulsante y el lugar de origen<sup>3</sup>. El tesla es la unidad de la inducción magnética (B) en el Sistema Internacional de Unidades. El hertz (Hz) es una medida de frecuencia, igual a una vibración por segundo. Las frecuencias de las radiaciones emitidas por el cuerpo humano nunca son superiores a 1000 Hz.

El SQUID ha sido utilizado para diagnosticar la epilepsia, pues permite registrar actividades magnéticas sobre la cabeza incluso cuando el electroencefalograma no detecta efectos apreciables. En los casos donde el paciente presentaba una intensa actividad eléctrica, pero muy difusa y poco localizada, las mediciones magnéticas lograron localizar el foco epiléptico en el interior de la masa cerebral.

Este instrumento también permite detectar la presencia de impurezas ferromagnéticas en los tejidos y órganos. Si se aplica un campo magnético intenso a una región del cuerpo que contenga partículas susceptibles de magnetización, en los tejidos aparece un campo magnético propio, muy débil, pero que no desaparece después que el campo externo se retira, a causa de la magnetización remanente característica de los materiales ferromagnéticos. La técnica ha sido utilizada para buscar partículas contaminantes en los pulmones de trabajadores expuestos a atmósferas muy específicas (por ejemplo, mineros y soldados).

Otra aplicación del magnetómetro SQUID es la detección de sustancias paramagnéticas, también magnetizables, pero con mucha menor intensidad que las ferromagnéticas y que no mantienen la magnetización una vez retirado el campo externo. Sin embargo, una sustancia paramagnética en un tejido ocasiona una variación en la distribución del campo externo aplicado que puede ser detectada por el magnetómetro. Esta técnica se ha utilizado para medir la cantidad de hierro en el hígado en pacientes afectados de anemia y también para medir la cantidad de sangre que fluye por el corazón, ya que los glóbulos rojos o hematíes contienen hierro en forma de hemoglobina, que puede ser detectada por el SQUID.

La magnetometría SQUID es una técnica cara. El principal problema de esta técnica es que, al medir señales que son muchos órdenes menores que las de los "ruidos magnéticos" ambientales, provenientes de lámparas, equipos electrónicos y líneas de transmisión, es obligatorio aplicarla en un recinto magnéticamente aislado. Para eso es necesario construir una cámara de algún metal adecuado, y no de cualquier metal, en cuyo interior quepan paciente e instrumento, para que la cámara sirva de blindaje al "ruido magnético" ambiental y al campo magnético de la tierra. Dentro de la cámara no puede haber equipos electrónicos, solamente el sensor SQUID,

pues el ruido electrónico generado por cualquier otro equipo sería detectado por el sensor, oscureciendo la determinación. Por lo tanto, otro problema a resolver es el de instalar los conductores eléctricos necesarios para interconectar los equipos dentro y fuera de la cámara sin que penetre el ruido magnético externo, lo que tampoco resulta fácil. Estas características hacen que algunos opinen que la magnetometría SQUID difícilmente se pueda generalizar en un futuro como técnica rutinaria de análisis.

**Tabla 1**

B (Tesla)	Origen del campo	frecuencia de la radiación (Hz)
$7 \times 10^{-5}$	campo magnético de la tierra <sup>2</sup>	0
$10^{-7}$	fluctuaciones del campo magnético de la tierra y "ruido magnético" urbano	todo el intervalo
$10^{-9}$	partículas pulmonares	0.1 - 10
$7 \times 10^{-11}$	magnetocardiograma	0.1 - 100
$10^{-12}$	magnetocardiograma fetal	0.1 - 100
$10^{-12}$	magnetoencefalograma (ondas $\alpha$ )	8 - 12
$10^{-13}$	magnetoretinograma	0.1 - 800
$10^{-15}$	límite de sensibilidad del SQUID	todo el intervalo

En notación científica,  $7 \times 10^{-5}$  significa trasladar el punto decimal 5 lugares a la izquierda: 0.00007

### 3. MAGNETOBIOLOGÍA

En los últimos 20 años han aparecido múltiples informes describiendo el hallazgo de pequeñas cantidades de óxido ferroso-férrico  $Fe_3O_4$ , conocido vulgarmente como *magnetita* o imán natural, en el organismo de diversos animales; bacterias, anguilas, palomas y delfines, entre otros. Y se especula si estos animales utilizan el campo magnético de la tierra como una especie de brújula interna, que les sirve de orientación para desplazarse en su medio ambiente.

El menor organismo donde se ha comprobado la existencia de una capacidad de orientación en la dirección del campo magnético es el *Aquaspirillum magnetotacticum* o bacteria magnetotáctica<sup>4,5,6</sup>. Estas bacterias fueron descubiertas en 1975 por Richard P. Blakemore, quien notó que algunas de las bacterias que él observaba al microscopio siempre se movían hacia el mismo lado de su campo visual.

Cuando colocaba un imán cerca del portamuestras, las bacterias se movían hacia el polo norte del imán; las bacterias muertas también se orientaban en la misma dirección pero, desde luego, no se trasladaban al igual que las vivas. Blakemore encontró que estas bacterias son capaces de orientarse y viajar a lo largo de la dirección del campo porque tienen en su seno pequeñas partículas o *magnetosomas* compuestas esencialmente de magnetita; cada partícula es un pequeño imán permanente con su polo norte y su polo sur.

Las bacterias logran ordenar en su interior estos pequeños imanes uno detrás del otro para hacer un imán mucho más largo, y utilizan este imán como una pequeña brújula para alinearse a lo largo de la dirección del campo magnético.

Los magnetosomas presentan un intervalo de tamaños muy determinado, de 35 a 120 nm, y este estrecho rango conlleva consecuencias físicas importantes. Sólo las partículas magnéticas comprendidas en este intervalo son capaces de establecer monodominios magnéticos estables. Partículas menores de 35 nm son *superparamagnéticas* a temperatura ambiente y no se magnetizan de forma permanente en una dirección determinada. Por el contrario, las partículas mayores de 120 nm tienden a formar dominios múltiples, cada uno de ellos magnetizado en una dirección diferente, lo que reduce el valor total de la magnetización y proporciona una efectividad menor para generar campos magnéticos.

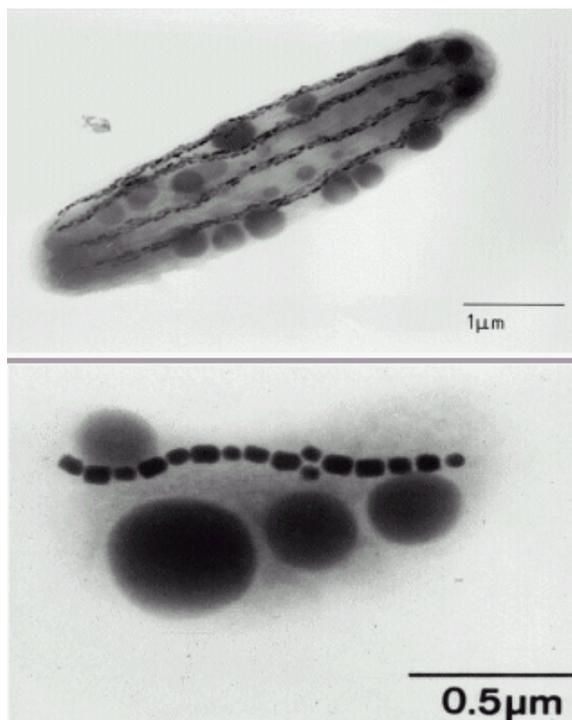
¿Para qué necesitan una brújula estas bacterias? Pues resulta ser que este tipo de bacterias prefiere vivir en medios donde hay poco o ningún oxígeno; son *anaerobias*. En un medio acuoso, el nivel de oxígeno decrece a medida que se avanza en profundidad, y las bacterias magnetotácticas utilizan sus brújulas internas para conocer donde se encuentra lo más profundo.

La dirección del campo magnético es paralela a la superficie de la tierra sólo en el ecuador; cuando nos alejamos del ecuador la dirección del campo se va inclinando hacia la superficie de la tierra, hasta llegar a ser perpendicular en los polos. Las bacterias que viven sumergidas en agua en el hemisferio norte, al orientarse hacia el norte, van efectivamente hacia lo más profundo. Y, como era de esperar, se comprobó que las bacterias del hemisferio sur, en vez de seguir el norte, se orientan al sur. En el ecuador se encuentra una mezcla de bacterias tipo "norte" y tipo "sur".

En experimentos realizados en el laboratorio, la inversión artificial de los campos norte y sur llevó a la inversión de la polaridad de las bacterias en un término de 8 semanas. Las nuevas generaciones de bacterias iban sustituyendo a las anteriores de manera tal que la orientación de sus magnetosomas era la que mejor favorecía la anaerobicidad.

Se sabe que las abejas poseen material magnético en la parte anterior del abdomen; el magnetismo parece desarrollarse en el estado de crisálida y persiste luego en los adultos. También hay mineral

magnético en la anguila, concretamente en los huesos del cráneo, de la columna vertebral y la faja pectoral<sup>7</sup>. En la cabeza y el cuello de las palomas mensajeras el material ferromagnético está presente en estructuras visibles a simple vista (0.5 – 2 mm), entre la duramadre y el cráneo; al ser observado al microscopio electrónico aparece un material opaco de aproximadamente 0.1 micrómetros de espesor, en forma de cristales negros de magnetita.



**Figura 2.** Arriba: bacteria magnetotáctica mostrando hileras de magnetosomas en su interior. Abajo: sección ampliada mostrando una hilera de magnetosomas. Las manchas difusas inferiores, de mucho mayor tamaño, son gránulos de azufre.



**Figura 3.** Diagrama esquemático de delfines siguiendo las líneas del campo magnético terrestre.

También se ha encontrado magnetita en mamíferos, especialmente en el Delfín Común del Pacífico (*Delphinus delphis*); aunque la sustancia magnética está presente en muchas partes de la cabeza, la

región más magnética se encuentra alrededor de 2 cm detrás de la cresta formada por la sutura de las partes occipitales, parietales y frontales, en la parte izquierda de la hoz cerebral, entre el cráneo y la duramadre<sup>8</sup>.

Se han llevado a cabo gran cantidad de experimentos para verificar el sentido de orientación magnética de diversos animales, sometiéndolos a la acción de campos magnéticos y estudiando su comportamiento en estas condiciones. Los experimentos preferidos son aquellos en que intervienen palomas mensajeras, aunque también se ha experimentado con insectos, peces, ... y hasta con personas.

Cada año cientos de ballenas y delfines quedan varados en las playas de todo el mundo. Algunos de estos animales son viejos y enfermos, pero también hay muchos jóvenes y fuertes. A pesar de que los encallamientos ocurren frecuentemente, no se ha podido encontrar una explicación satisfactoria en todos los casos. Si el animal está muerto, se puede suponer que murió en el mar y las corrientes lo arrastraron a la playa. Pero cuando los animales que quedan varados están vivos, resulta mucho más difícil encontrar una explicación. Una de las muchas teorías no comprobadas que existen para explicar

este comportamiento es que el encallamiento ocurre donde las líneas de inducción del campo magnético terrestre cruzan la línea de la costa, razón por la cual los animales pierden su orientación (figura 3).

Algunos experimentos proporcionan evidencia a favor de un sistema de orientación basado en el campo magnético de la tierra; otros, no tanto. Sin embargo, a pesar de los muchos experimentos y opiniones favorables en algunos casos, hasta el momento no se ha podido encontrar ningún mecanismo que explique estos posibles efectos.

En las bacterias magnetotácticas el mecanismo de orientación es muy simple; funciona incluso con la bacteria muerta. En un organismo más complejo debería existir alguna conexión de la sustancia magnética con el sistema nervioso para poder explicar la conducta de orientación observada, pero tal conexión aún no se ha encontrado. Como consecuencia, aunque no es difícil encontrar escritos que dan por sentado que muchas aves se guían en sus migraciones por una combinación visual y magnética, muchos biofísicos no creen que la sensibilidad al campo magnético de los animales superiores esté en realidad comprobada.

El tema aún se encuentra abierto a la investigación.

---

## REFERENCIAS

---

<sup>1</sup>A. González Arias, *Magnetismo y Pseudociencia en la Medicina*, Rev. Cubana de Física, Vol. 20, No. 1, (2003)

<sup>2</sup> Keiji Tsukada, Tsuyoshi Miyashita, Akihiko Kandori, Toshio Mitsui, Yasushi Terada, Masato Sato, Junko Shiono, Hitoshi Horigome, Satsuki Yamada & Iwao Yamaguchi, *An iso-integral mapping technique using magnetocardiogram, and its possible use for diagnosis of ischemic heart disease*, International Journal of Cardiac Imaging 16: 55-66, (2000)

<sup>3</sup> A. González Arias, *¿Qué es el Magnetismo?*, Acta Salmanticensis, Biblioteca de las Ciencias, Ediciones Universidad de Salamanca, España, (2001)

<sup>4</sup> R.B. Frankel, R.P. Blakemore, and R.S. Wolfe: *Magnetite in Freshwater Magnetic Bacteria*, Science 203, 1355-1357 (1979)

<sup>5</sup> R.B. Frankel: *Magnetotactic Bacteria*, Comments Mol. Cell. Biophys. 1, 293-310 (1982)

<sup>6</sup> D. Schueler and R.B. Frankel, *Bacterial magnetosomes: microbiology, biomineralization and biotechnological applications*, Applied Microbiology and Biotechnology 52, 464-473 (1999)

<sup>7</sup> Hanson M. et. al., *Magnetic Material in European Eel (Anguilla Anguilla)*, Comp. Biochem. Physiol. a. Comp. Physiol., 77, 2, 221—224, (1984)

<sup>8</sup> Zoeger J. et. al., *Magnetic Material in the Head of the Common Pacific Dolphin*, Science, 213, 892-894, (1981)