

MEMORIAS DE LA XX ESCUELA DE VERANO DE FÍSICA

junio 18-29, 2012

Instituto de Física, UNAM, 18-23 de junio
Instituto de Ciencias Físicas, UNAM, 24-29 de junio

Editores

JOSÉ RÉCAMIER
ROCÍO JÁUREGUI



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
2013

Esta edición fue preparada por el Instituto de Física
y el Instituto de Ciencias Físicas de la UNAM.

Primera edición electrónica: 2013

© D.R. UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Ciudad Universitaria, 04510, México D. F.
Instituto de Física
Instituto de Ciencias Físicas

Prohibida la reproducción parcial o total por cualquier medio
sin autorización escrita de su legítimo titular de derechos.

ISBN: 978-607-02-4377-6

Hecho en México

CONTENIDO

Introducción	5
Agradecimientos.	7
Profesores participantes	9
Instituciones participantes	11
Alumnos participantes	13
▪ Contribuciones	
Denis Boyer	
<i>Las caminatas aleatorias y sus aplicaciones</i>	15
F. Castillo, H. Martínez, O. Flores, B. Campillo	
<i>Propiedades de los plasmas</i>	37
Sergio Cuevas García	
<i>¿Cómo se mezclan los fluidos?</i>	51
Rubén Fossión	
<i>Estimación y cuantificación de la adaptabilidad homeostática en sistemas fisiológicos</i>	77
Gabriel Germán Velarde	
<i>Sobre la expansión actual del universo</i>	91
Agustín E. González	
<i>Simulaciones por computadora de la cristalización coloidal</i>	99
Eugenio Ley Koo	
<i>Operadores de escalera en 1, 2 y 3 dimensiones</i>	105
Frédéric Masset	
<i>Migración planetaria</i>	121
Nina Pastor	
<i>Módulos estructurales y percursoros de fibras: anticuerpos y amiloidosis</i>	133
Eduardo Ramos	
<i>Atmospheric turbulence and wind energy utilization</i>	139
José Récamier	
<i>Estados coherentes</i>	153
Humberto Saint Martin	
<i>Las simulaciones numéricas de sistemas moleculares</i>	165

INTRODUCCIÓN

La XX Escuela de Verano en Física fue organizada por el Posgrado en Ciencias Físicas, el Instituto de Física y el Instituto de Ciencias Físicas de la Universidad Nacional Autónoma de México. Se llevó a cabo en las instalaciones del Instituto de Física, en Ciudad Universitaria, del 18 al 23 de junio y en las instalaciones del Instituto de Ciencias Físicas, en Cuernavaca, Morelos, del 24 al 29 de junio de 2012.

En esta escuela se impartieron ocho cursos de cinco horas de duración cada uno y 25 conferencias. Los cursos y conferencias cubrieron un amplio espectro con temas como cosmología, plasmas, biofísica, óptica cuántica, nanobiotecnología, caminatas aleatorias, mecánica cuántica, cuerdas, sistemas complejos, astronomía y biología teórica entre otros.

Rocío Jáuregui, Instituto de Física
José Récamier, Instituto de Ciencias Físicas
Rafael Pérez Pascual, Instituto de Física
Universidad Nacional Autónoma de México
Ciudad Universitaria, abril, 2013

AGRADECIMIENTOS

Agradecemos a Guadalupe Ramírez por el apoyo secretarial y a Ulises Amaya por el apoyo con la página de la escuela.

Agradecemos los apoyos recibidos para la realización de esta escuela a la Universidad Nacional Autónoma de México a través de la Coordinación de la Investigación Científica, del Instituto de Física (IF) y del Instituto de Ciencias Físicas (ICF). Agradecemos también al programa de Becas de Movilidad Santander-Universia y a las Becas ECOES.

PROFESORES PARTICIPANTES

- Jorge Ascencio Gutiérrez, ICF, *Nanobiotecnología y física de materiales aplicados al desarrollo tecnológico de alto impacto social.*
- Pablo Barberis, IMAS, *Huellas de estados enredados en correlaciones de luz en cavidades ópticas.*
- María Ester Brandan, IF, *Malignidad a simple vista: resta de imágenes para visualizar cáncer en una mamografía.*
- Remigio Cabrera Trujillo, ICF, *Propiedades de átomos bajo presión.*
- Bernardo Campillo Illanes, FQ, *Procesos físicos en plasmas y sus aplicaciones.*
- Octavio Castaños Garza, ICN, *Transiciones de fase cuánticas en sistemas finitos.*
- Carmen Cisneros Gudiño, ICF, *Radiación sincrotrón.*
- Sergio Cuevas García, CIE, *¿Cómo se mezclan los fluidos?*
- Jaime de Urquijo Carmona, ICF, *Los plasmas de baja temperatura: ¿qué son?*
- Raul Esquivel, IF, *Autoensamblaje la construcción de nanoestructuras.*
- Rubén Fossion, IG, *Estudio de sistemas complejos dinámicos con series de tiempo y señales de alerta temprana: adaptabilidad vs colapso.*
- Gabriel Germán Velarde, ICF, *Sobre la expansión actual del universo.*
- Hortensia González, FC, *Algunos métodos de estudio de la variabilidad de la frecuencia cardiaca.*
- Agustín González Flores, ICF, *Estructura fractal de los núcleos de cristalización en cristales coloidales en 2D.*
- Shahen Hacyan, IF, *Física de viajes interestelares.*
- Guillermo Hinojosa Aguirre, ICF, *Experimentos con haces de iones: colisiones, fotones y tecnología.*
- Rocío Jáuregui, IF, *Elementos básicos de teoría de la información.*
- Antonio Juárez Reyes, ICF, *Perspectiva actual de la física molecular: Estudios fundamentales y usos en áreas estratégicas.*
- Hernán Larralde Ridaura, ICF, *Caminatas aleatorias.*
- Eugenio Ley, IF, *Operadores de subida y de bajada en sistemas cuánticos de una, dos y tres dimensiones.*
- Axel de la Macorra, IF, *Nuestro universo: ¿cómo sabemos lo que sabemos?*
- Arnulfo Martínez, IF, *Desarrollo de sistemas híbridos de microtomografía*
- Lorenzo Martínez Gómez, ICF, *El reto de llevar el conocimiento científico a la actividad económica y social. Un ejemplo del caso.*

- Horacio Martínez Valencia, ICF, *Procesos físicos en plasmas y sus aplicaciones.*
- Rafael Méndez Sánchez, ICF, *Caos cuántico y elástico: teoría y experimento.*
- Luis Mochán, ICF, *Propiedades macroscópicas y bandas fotónicas en medios inhomogéneos*
- Myriam Mondragón, IF, *Astropartículas.*
- Matías Moreno, IF, *Física de aceleradores.*
- Iván Ortega Blake, ICF, *Biofísica molecular.*
- Nina Pastor Colón, FC-UAEM, *Módulos estructurales y precursores en fibras: anticuerpos y amiloidosis.*
- Rosario Paredes, IF, *Fluidos cuánticos: fundamentos, experimentos y aplicaciones.*
- Carlos Pineda, IF, *Matrices aleatorias e información cuántica.*
- Saul Ramos, IF, *Supersimetría y gran unificación.*
- Eduardo Ramos Mora, CIE, *Energía eólica.*
- José Récamier Angelini, ICF, *Estados coherentes.*
- Humberto Saint Martin, ICF, *Biofísica molecular.*
- Genaro Toledo, IF, *Las partículas que quisieron ser estrellas.*
- Mayo Villagrán, CCADET, *Plasmas inducidos por láser.*
- Carlos Villarreal, IF, *Superconductividad.*

INTITUCIONES PARTICIPANTES

- ICF - Instituto de Ciencias Físicas, U.N.A.M.
- ICN - Instituto de Ciencias Nucleares, U.N.A.M.
- IF - Instituto de Física, U.N.A.M.
- IIMAS - Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, U.N.A.M.
- CCADET - Centro de Ciencias Aplicadas y Tecnología Avanzada, U.N.A.M.
- CIE - Centro de Investigación en Energía, U.N.A.M.
- FC - Facultad de Ciencias, U.N.A.M.
- FC UAEM - Facultad de Ciencias, Universidad Autónoma del Estado de Morelos
- IG - Instituto de Geriátría, Secretaría de Salud.

ALUMNOS PARTICIPANTES

- Rafael Alberto Díaz Hernández Rojas, FC-UNAM
- David Landa Marbán, FC UNAM
- José Luis Maya Bernal, FC UNAM
- Bruno Carballo Zama, FC UNAM
- José Cruz Rojas, FC UNAM
- Juan Carlos del Valle Rosales, FC UNAM
- Gustavo Amilcar Saldaña Moncada, FC UNAM
- José Ernesto Alba Arroyo, FC UNAM
- Jaime David Wong Campos, ITESM MTY
- Irving Enrique Reyna Nolasco, FC UNAM
- Marco A. Lechuga, U de G
- David Dávalos, U de G
- José Luis López López, FC UNAM
- Lorena Marisol García Iglesias, U de G
- Ramiro Contreras Martínez, U de G
- Jesús Carlos Toscano Figueroa, U de G
- Armando Enríquez Pérez Gavilán, ITESM
- Andrés Sánchez Pérez, U de G
- María Guadalupe Gómez Farfán, FC UNAM
- Katya Patricia Vázquez Rivera, FC UNAM
- Oriana Trejo Alvarez, FC UNAM
- Jaziel Soto Muñoz, UACH
- Marian Lemus García, UMSNH
- Ezequiel Manzo Martínez, UNSNH
- José Agustín Lozano Torres, U. Veracruzana
- Víctor Castillo Gallardo, UAM-A
- Olivia Nadia Martínez Vega, Tec. Zacatepec
- Fernando Guerrero Vélez, UANL
- Paola Vanessa Olguín Rodríguez, UAEM
- César Iván León Pimentel, UMSNH
- Efrén Anguiano Guzmán, U de G
- Miguel Ángel Mendoza Mejía, U. Coah.

- Aracely Martínez Longoria, U. Guanajuato
- Carolina Bohórquez Martínez, UABJO
- Ricardo Martínez Martínez, U de G
- Carlos Jessé Colón Villalón, UASLP
- Leonardo Báez Castillo, U. Sonora
- Francisco Perea Alvarado, UAM-A
- Ana Luisa Sandoval Jasso, UACH
- Elizabeth García Rosas, I. Tec. Qro.
- Guillermo Ramírez Arias, UPEMOR
- Ángel Eduardo Rodríguez Fernández, FC UNAM
- Pamela Patricia Ornelas Silva, U de G
- Ana Luisa Aguayo Alvarado, U de G

Las caminatas aleatorias y sus aplicaciones¹

Denis Boyer

Instituto de Física, UNAM

Los devenires son geografía, son orientaciones, direcciones, entradas y salidas. Son actos que sólo pueden estar contenidos en una vida y expresados en un estilo. A medida que alguien deviene, aquello en lo que deviene cambia tanto como él.

Gilles Deleuze

1 Introducción

La difusión es tal vez el mecanismo más básico de transporte en la materia y se encuentra al origen de muchos procesos, tanto de equilibrio como fuera de equilibrio, en termodinámica y física estadística. Las caminatas aleatorias ofrecen un marco teórico útil para entender la difusión a un nivel microscópico. Estas caminatas se han utilizado para describir el movimiento browniano de una partícula sujeta a colisiones con las moléculas de un fluido que la rodea, así como para entender la cinética de reacciones químicas, las propiedades estadísticas de avalanchas en medios granulares o las conformaciones de polímeros. En áreas distintas a la Física, las caminatas aleatorias aparecen de manera ubicua en la descripción de problemas donde el ruido y la incertidumbre juegan un papel fundamental, por ejemplo: cómo una proteína encuentra un sitio funcional en la célula, cómo los animales usan el espacio y los recursos de un ecosistema, cómo los precios evolucionan en mercados bursátiles o en problemas de adopción de opiniones por agentes sociales. A pesar de la simplicidad de su formulación probabilística original, los procesos difusivos siguen siendo un tema de investigación muy activo en la actualidad. En este capítulo presentaremos algunas técnicas y resultados básicos sobre la difusión, así como sus aplicaciones recientes en temas interdisciplinarios. Nos enfocaremos en el estudio de los patrones de movilidad individual en animales o humanos y discutiremos perspectivas en esa área. Muchos de nuestros problemas de interés involucran una partícula pero mencionaremos fenómenos colectivos que pueden emerger en sistemas compuestos de muchos elementos.

¹Otra versión de este texto ha sido publicada por el autor con el título “Procesos difusivos: de moléculas a animales” como un capítulo del libro electrónico de acceso libre: *Fronteras de la física en el siglo XXI*, O. Miramontes y K. Volke (Editores), CopIt ArXives, México, 2013.

2 Difusión simple

2.1 Caminatas aleatorias y Teorema Límite Central

Iniciamos considerando una partícula que puede ocupar posiciones discretas $x(t)$ equidistantes a lo largo de una línea, moviéndose a cada paso en tiempo discreto ($t \rightarrow t + 1$) al sitio más cercano ubicado a su derecha o a su izquierda con igual probabilidad $1/2$, ver Figura 1a. La probabilidad $P_t(n)$ de que el caminante esté en el sitio n ($n = 0, \pm 1, \pm 2, \dots$) al instante t obedece a una ecuación recursiva de dos variables:

$$P_t(n) = \frac{1}{2}P_{t-1}(n-1) + \frac{1}{2}P_{t-1}(n+1). \quad (1)$$

Esta expresión suma la probabilidad de que en el instante $t - 1$, la partícula esté en $n - 1$ y dé un salto hacia adelante y la probabilidad de que esté en $n + 1$ y dé un salto hacia atrás. Si el caminante se encuentra en la posición $n = 0$ al tiempo inicial $t = 0$, después de t pasos la probabilidad de que haya dado n_d pasos a la derecha y $t - n_d$ pasos a la izquierda en un orden cualquiera está dada por la fórmula del binomio:

$$P_t(n) = \binom{t}{n_d} \frac{1}{2^t} \quad (2)$$

En esta fórmula se tomó en cuenta el hecho de que cualquier secuencia dada de t pasos tiene probabilidad $(1/2)^t$ de ocurrir y que la posición n vale $2n_d - t$ (si los enteros n y t tienen la misma paridad). Se puede comprobar que (2) es solución de (1). A tiempos grandes ($t \gg 1$) usamos la aproximación de Stirling y obtenemos la bien conocida distribución gaussiana [1]:

$$P_t(n) \simeq \frac{1}{\sqrt{4\pi Dt}} e^{-n^2/4Dt}, \quad (3)$$

donde D ($= 1/2$ en este caso) es la constante de difusión. Se puede re-obtener este resultado a partir de un modelo un poco diferente, en el cual el tiempo es continuo y la probabilidad de saltar a la derecha en el intervalo de tiempo $[t, t+dt]$ es αdt (lo mismo sucede a la izquierda) [2]. Eligiendo la tasa $\alpha = 1$, podemos escribir una ecuación maestra para la evolución de $P_t(n)$:

$$\frac{\partial P_t(n)}{\partial t} = P_t(n+1) - 2P_t(n) + P_t(n-1). \quad (4)$$

El término $-2P_t(n)$ proviene del hecho de que el caminante deje su posición actual con probabilidad $2dt$ en $[t, t+dt]$. Para resolver (4) se introduce la transformada de Fourier discreta:

$$P(k, t) = \sum_{n=-\infty}^{\infty} P_t(n) e^{ikn}. \quad (5)$$

De (4), obtenemos

$$\frac{\partial P(k, t)}{\partial t} = [e^{ik} + e^{-ik} - 2]P(k, t). \quad (6)$$

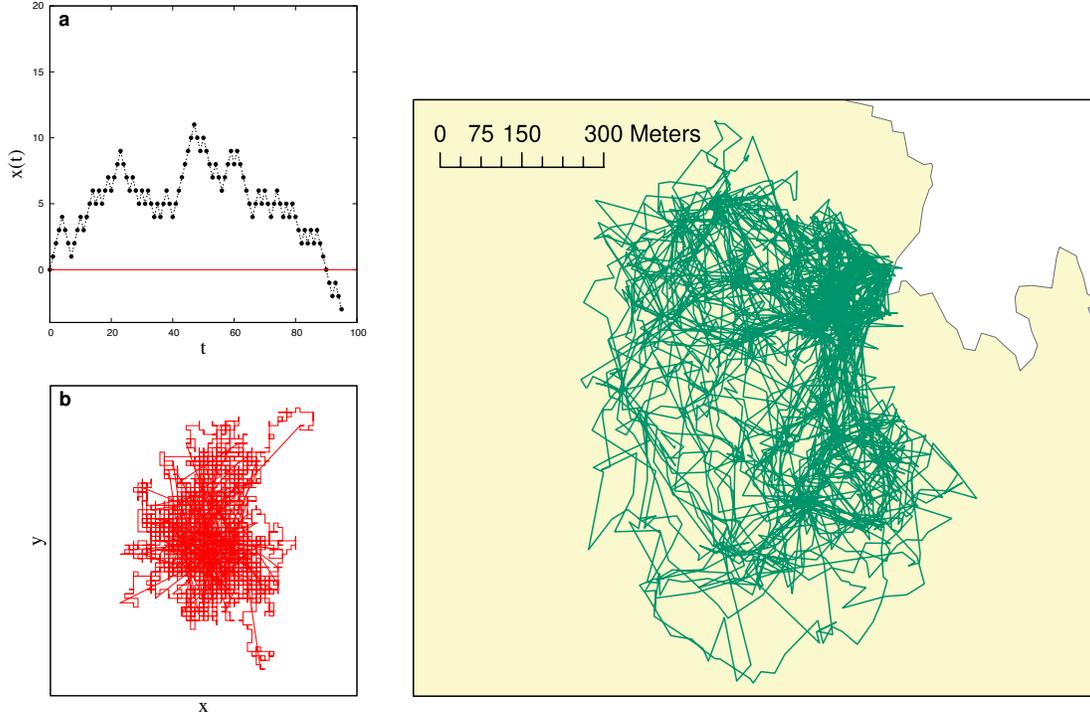


Figure 1: **Izquierda:** a) Posición en función del tiempo de una caminata aleatoria discreta en una dimensión espacial, que da pasos con igual probabilidad a la derecha o a la izquierda. En este ejemplo, el caminante regresa por primera vez a su punto de partida a $t = 90$. b) Caminata aleatoria en dos dimensiones espaciales con saltos intermitentes a lugares visitados anteriormente, ver sección 3.1. **Derecha:** Trayectoria recorrida durante un mes por un primate (mono capuchino) obtenida con un GPS en la isla de Barro Colorado, Panamá [cortesía de Meg Crofoot].

Dado que $P_{t=0}(n) = \delta_{n,0}$ entonces $P(k, t = 0) = 1$ y (6) se puede integrar con esta condición inicial:

$$P(k, t) = e^{2(\cos k - 1)t}. \quad (7)$$

La solución $P_t(n)$ se obtiene usando la identidad $e^{z \cos k} = \sum_{n=-\infty}^{\infty} e^{ikn} I_n(z)$ donde I_n son las funciones de Bessel modificadas. Comparando con (6):

$$P_t(n) = I_n(2t) e^{-2t}, \quad (8)$$

la cual converge a la forma asintótica (3) cuando $t \gg 1$, con $D = 1$. En un espacio continuo se sustituye n por una posición real x , reconociendo que el lado derecho de (4) tiene forma de una segunda derivada espacial. Esta ecuación se convierte una ecuación diferencial parcial:

$$\frac{\partial P(x, t)}{\partial t} = D \frac{\partial^2 P(x, t)}{\partial x^2}, \quad (9)$$

que es la bien conocida ecuación de la difusión. $P(x, t)$ es una densidad de probabilidad: $P(x, t)dx$ es la probabilidad de que la partícula se encuentre entre x y $x + dx$ al instante t . Por normalización, $\int_{-\infty}^{\infty} P(x, t)dx = 1$ para todo t . Si la partícula se encuentra al origen $x = 0$ en $t = 0$, entonces $P(x, t = 0) = \delta(x)$ ² y la forma gaussiana (3) es solución exacta de (9) para $t > 0$, sustituyendo $n \rightarrow x$. Una consecuencia importante de este resultado es que por paridad, la posición media de la partícula $\langle x \rangle = \int xP(x, t)dx$ es cero: no hay *en promedio* transporte en ninguna dirección preferente. Sin embargo, el segundo momento de la distribución $P(x, t)$, también llamado desplazamiento cuadrático medio $\langle x^2 \rangle$ se relaciona con el ancho de la distribución gaussiana y no es cero. En una trayectoria dada, los números de pasos a la derecha y a la izquierda no son iguales en general y la diferencia tiende a crecer con el tiempo debido a fluctuaciones estadísticas. En otras palabras, $\langle x^2 \rangle$ crece con el tiempo:

$$\langle x^2 \rangle(t) = \int_{-\infty}^{\infty} x^2 P(x, t)dx = 2Dt. \quad (10)$$

Esta es la ley de Smoluchowski-Einstein. La cantidad $l = \sqrt{\langle x^2 \rangle}$ representa la distancia típica entre el caminante y su punto de origen después de t pasos. Esta distancia crece como $t^{1/2}$, indefinidamente, aunque mucho más lentamente que una partícula en movimiento uniforme ($l \sim t$).

Además de describir la densidad de probabilidad del desplazamiento x para los modelos de caminatas aleatorias descritos arriba, la distribución gaussiana se aplica más generalmente a la suma de variables aleatorias reales independientes (es decir, sin correlaciones entre sí) cuando la suma contiene un gran número de términos. Esta universalidad se explica por el *Teorema Límite Central* (ver por ejemplo [3]). Consideremos una suma (posición) x que contiene N términos (pasos),

$$x = \sum_{n=1}^N u_n, \quad (11)$$

donde los u_n no necesariamente toman los valores ± 1 sino que están generados aleatoriamente a partir de una distribución de probabilidad $p(u)$ arbitraria³. La relación (1) toma la forma general:

$$P(x, N) = \int_{-\infty}^{\infty} P(x', N - 1)p(x - x')dx'. \quad (12)$$

Para llegar al valor x en N pasos, hay que llegar a algún x' en $N - 1$ pasos y luego añadir $x - x'$, lo cual ocurre con probabilidad $p(x - x')dx'$. La integral (12) es una convolución entre P y p . Definiendo la transformada de Fourier continua,

$$f(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx, \quad (13)$$

² $\delta(x)$ es la función delta de Dirac y δ_{ij} la función de Kronecker.

³La probabilidad que u tome un valor en $[U, U + dU]$ es $p(U)dU$.

obtenemos de (12) la relación $P(k, N) = P(k, N - 1)p(k)$. Esta relación se resuelve fácilmente por iteración⁴ y se obtiene $P(k, N) = [p(k)]^N$. Notamos que $p(k = 0) = \int p(u)du = 1$ por normalización y que $|p(k)| < 1$ si $k \neq 0$: entonces $P(k, N)$ es exponencialmente chico cuando $N \gg 1$, excepto para k cerca de 0. Esto explica la universalidad de las funciones gaussianas. Desarrollando en serie de Taylor con k pequeño obtenemos $e^{iku} \simeq 1 - iku - \frac{1}{2}k^2u^2$ y deducimos que $p(k) \simeq 1 - ik\langle u \rangle - \frac{1}{2}k^2\langle u^2 \rangle$. Para simplificar consideramos $\langle u \rangle = 0$ y suponiendo $\langle u^2 \rangle \equiv \int_{-\infty}^{\infty} u^2 p(u)du$ finito ($< \infty$), la distribución $P(x, N)$ que buscamos se obtiene por transformada inversa de Fourier de la función $[p(k)]^N$:

$$\begin{aligned} P(x, N) &\simeq \int_{-\infty}^{\infty} \left[1 - \frac{1}{2N} N k^2 \langle u^2 \rangle \right]^N e^{ikx} dk \\ &\simeq \int_{-\infty}^{\infty} e^{-\frac{N\langle u^2 \rangle}{2} k^2 + ikx} dk \\ &= \frac{1}{\sqrt{2\pi N \langle u^2 \rangle}} e^{-\frac{x^2}{2N \langle u^2 \rangle}}. \end{aligned} \quad (14)$$

En resumen, la suma de N variables aleatorias de varianza finita tiende a una distribución gaussiana, sin importar la forma analítica de la distribución $p(u)$ de estas variables. Los resultados anteriores se generalizan fácilmente en espacios d -dimensionales, en donde se sustituye x por el vector posición \mathbf{r} y se obtiene:

$$P(\mathbf{r}, t) = \frac{1}{(4\pi Dt)^{d/2}} e^{-\frac{\mathbf{r}^2}{4Dt}}, \quad \langle \mathbf{r}^2 \rangle = 2dDt. \quad (15)$$

La Figura 1b muestra un ejemplo de caminata en dos dimensiones.

2.2 Propiedades de primer paso

Consideramos un caminante con posición \mathbf{r} en una red discreta infinita dentro de un espacio de dimensión d (\mathbf{r} es por lo tanto un vector con d componentes enteros relativos). La función $P(\mathbf{r}, t)$ describe la probabilidad de *estar* en \mathbf{r} al instante t . Sin embargo, podemos plantear otra clase de preguntas. Si $\mathbf{r} = \mathbf{0}$ a $t = 0$, ¿Cuál es la probabilidad de que el caminante llegue *por primera vez* a un sitio \mathbf{r}_0 dado al instante t ? ¿Cuál es la probabilidad de que el caminante vuelva a su punto de origen? ¿Cuántos sitios *distintos* son visitados en promedio en t pasos?

Las propiedades de primer paso son importantes en problemas de búsqueda aleatorias, en donde el caminante puede estar representando un animal o una molécula y \mathbf{r}_0 un sitio de alimento o un sitio de reacción [4]. Las respuestas a las preguntas anteriores no son triviales y dependen de la dimensión espacial d . En un espacio no acotado, si $d \leq 2$, las caminatas aleatorias son *recurrentes*, es decir visitan muchas veces los mismos sitios, mientras que si $d > 3$ las caminatas son

⁴Con $N = 0$, la suma x vale 0 con probabilidad 1: $P(x, 0) = \delta(x)$ o $P(k, 0) = 1$.

transientes: los sitios visitados son pocas veces revisitados y existe una probabilidad > 0 de nunca regresar a un sitio que ya ha sido visitado. Heurísticamente, a partir de la ley de la difusión (10), la trayectoria de un caminante hasta un tiempo t se encuentra en un dominio de radio $l(t) \sim t^{1/2}$. Hay t visitas en t pasos (no necesariamente a sitios distintos) por lo tanto el número de visitas a un sitio dado dentro del dominio es $n_v(t) \sim t/l(t)^d \sim t^{1-d/2}$. Si $d < 2$, $n_v(t)$ crece con t y casi seguramente todos los sitios del dominio son visitados muchas veces; si $d > 2$, $n_v(t) \rightarrow 0$: muchos sitios en el dominio nunca serán visitados.

Estas propiedades de recurrencia se pueden derivar de manera rigurosa. Considerando un caminante que sigue el proceso de difusión en tiempo continuo descrito en (4), definimos la probabilidad $P_1(\mathbf{r}_0, t)dt$ de alcanzar por primera vez una posición \mathbf{r}_0 dada entre t y $t+dt$. La variable t suele llamarse en este contexto un tiempo de primer paso (en \mathbf{r}_0). En un espacio homogéneo:

$$P(\mathbf{r}_0, t) = \int_0^t P_1(\mathbf{r}_0, t')P(\mathbf{0}, t-t') dt' + \delta_{\mathbf{r}_0, \mathbf{0}} e^{-2dt}. \quad (16)$$

Para estar en \mathbf{r}_0 al tiempo t , es necesario llegar ahí por primera vez en algún tiempo $t' \leq t$ y haber regresado a ese origen en un tiempo $t-t'$. El segundo término representa el caso $\mathbf{r}_0 = \mathbf{0}$: la probabilidad de haberse quedado en el origen hasta t es $[\exp(-t)]^{2d}$, dado que el caminante puede saltar con tasa unidad a uno de $2d$ sitios más cercanos en la red. Como conocemos P , podemos obtener P_1 de (16). Con este fin, es conveniente usar la transformada de Laplace definida para cualquier función $f(t)$ como:

$$f(s) = \int_0^\infty f(t) e^{-st} dt, \quad (17)$$

donde s es la variable conjugada al tiempo. La convolución temporal en (16) se traduce como:

$$P_1(\mathbf{r}_0, s) = \frac{P(\mathbf{r}_0, s) - \delta_{\mathbf{r}_0, \mathbf{0}}/(s+2d)}{P(\mathbf{0}, s)}. \quad (18)$$

Eligimos ahora $\mathbf{r}_0 = \mathbf{0}$ y queremos determinar $P_1(\mathbf{0}, t) \rightarrow P_1(t)$, la distribución de los tiempos de primer retorno al origen (ver un ejemplo de primer retorno en la Fig. 1a). La cantidad $\mathcal{R} = P_1(s=0) = \int_0^\infty P_1(t)dt$ representa entonces la probabilidad de que el caminante regrese en algún momento a su punto de partida. Usando (18) y (15) se obtiene:

$$\mathcal{R} = 1 - \frac{1}{P(\mathbf{0}, s=0)}. \quad (19)$$

$P(\mathbf{0}, s=0) = \int_0^\infty P(\mathbf{0}, t)dt$ es la densidad de presencia en el origen integrada sobre todos los tiempos. De (15), $P(\mathbf{0}, t) \propto t^{-d/2}$ a t grande. Para $d \leq 2$ esa función decae más lentamente que $1/t$, por lo tanto $P(\mathbf{0}, s=0) = \infty$. Usando (19), concluimos que $\mathcal{R} = 1$: las caminatas siempre vuelven a pasar por su punto

de origen. En dimensiones mayores que 2, $P(\mathbf{0}, s = 0) < \infty$ y su valor preciso para redes hiper-cúbicas puede ser calculado a partir de (8). En $d = 3$, la caminata regresa al origen solamente con probabilidad $\mathcal{R} = 0.3405\dots$

También podemos derivar el comportamiento de la distribución $P_1(t)$ de los tiempos del primer retorno al origen, a tiempos grandes. Esto se obtiene analizando la ecuación (18) para s pequeño. En $d = 1$, $P(\mathbf{0}, t) \simeq (4t)^{-1/2}$ implicando que $P(\mathbf{0}, s) \simeq 4s^{-1/2}$ cuando $s \rightarrow 0$. Deducimos de (18) que $P_1(\mathbf{0}, s) \simeq 1 - \sqrt{s}$, que es la transformada de Laplace de la función:

$$P_1(t) \simeq \frac{1}{\sqrt{4\pi}} t^{-3/2}, \quad t \gg 1 \quad (d = 1). \quad (20)$$

Esta distribución es una ley de potencia decreciente: es más probable que el caminante regrese al origen (o a cualquier punto previamente visitado) en un plazo corto que después de mucho tiempo. Sin embargo, el decaimiento de $P_1(t)$ con t es lento, por ejemplo, cuando se compara con una exponencial, incluso algunas excursiones son muy largas, tan largas que el tiempo medio de primer retorno al origen diverge: $\langle t \rangle = \int_0^\infty t P_1(t) dt \propto \int_1^\infty t^{-1/2} dt = \infty$. Los caminantes siempre regresan pero en promedio se tardan un tiempo infinito para hacerlo... En $d = 2$, para s pequeño $P(\mathbf{0}, s) \simeq \int_1^\infty (4\pi t)^{-1} e^{-st} dt \simeq -1/(4\pi) \ln s$. Usando (18), obtenemos $P_1(s) = 1 + \pi \ln s$, que es la transformada de Laplace de

$$P_1(t) \simeq \frac{\pi}{t(\ln t)^2}, \quad t \gg 1 \quad (d = 2). \quad (21)$$

Esta distribución decae aún más lentamente con t que en el caso $1d$ y su tiempo medio de primer retorno $\langle t \rangle$ es infinito también.

2.3 Número de sitios distintos visitados y procesos de aniquilación

También es de particular interés el número $S(t)$ de sitios distintos visitados en promedio durante un intervalo de tiempo t [5]. Esta cantidad representa el territorio cubierto o “conocido” por el caminante. Dado que el caminante puede visitar varias veces un mismo sitio, $S(t)$ no puede ser mayor a $\mathcal{O}(t)$. La probabilidad de que \mathbf{r} haya sido visitado durante $[0, t]$ es $\int_0^t P_1(\mathbf{r}, t') dt'$, entonces

$$S(t) = 1 + \sum_{\mathbf{r} \neq \mathbf{0}} \int_0^t P_1(\mathbf{r}, t') dt'. \quad (22)$$

Usando (18), obtenemos para la transformada de Laplace de $S(t)$: $S(s) \simeq 1/[s^2 P(\mathbf{0}, s)]$ cuando $s \rightarrow 0$. Del comportamiento de $P(\mathbf{0}, s)$ para s pequeño (ver arriba) deducimos que:

$$S(t) \sim \begin{cases} t^{1/2}, & d = 1 \\ t/\ln t, & d = 2. \\ t & d > 2 \end{cases} \quad (23)$$

De acuerdo con el argumento heurístico anterior, la dimensión 2 marca una transición recurrencia/no-recurrencia arriba de la cual las propiedades de $S(t)$ no cambian fundamentalmente. Por esta razón, la dimensión $d_c = 2$ se denomina dimensión crítica para el número de sitios visitados. Como suele pasar en otros fenómenos críticos, aquí también aparecen correcciones logarítmicas en d_c , ver la ecuación (23). Un sitio visitado durante $[0, t]$ se visita en promedio $t/S(t)$ veces. En $d = 2$ este número crece de manera logarítmica pero es una propiedad difícil de predecir heurísticamente.

Los resultados mostrados tienen aplicaciones a problemas de cinética de reacciones químicas, entre otros [2]. Consideremos un conjunto de partículas que se difunden libremente en un medio. Supongamos que cuando 2 partículas se encuentran en un mismo sitio se aniquilan, es decir, desaparecen: $A + A \rightarrow \emptyset$. Dado un número inicial de partículas $N_0 \gg 1$ en un volumen L^d (densidad $\rho_0 = N_0/L^d$), en promedio ¿cuál es el número de partículas $N(t)$ sobrevivientes al instante t ?

En este problema con muchos cuerpos interactuando entre sí, un argumento de tipo cualitativo puede ser muy útil. Consideremos al tiempo t una partícula sobreviviente y mantengamos las otras $N(t) - 1 \simeq N(t)$ fijas en sus posiciones. Al tiempo $t + \tau$, el número medio n_e de partículas encontradas por la partícula focal es proporcional al número de sitios distintos que visitó durante el tiempo τ y a la densidad media en $[t, t + \tau]$: $S(\tau)\bar{\rho} = S(\tau)\tau^{-1} \int_t^{t+\tau} dt' N(t')/L^d$. Suponiendo que $N(t)$ decae como una ley de potencia, $N(t) \sim t^{-\alpha}$, entonces $n_e \propto S(\tau)[(t + \tau)^{-\alpha+1} - t^{-\alpha+1}]/\tau$. Cuando n_e es de orden 1, la probabilidad de que la partícula focal desaparezca es bastante alta, independientemente de t . Eso es posible si $\tau = \text{const} \times t$ y $S(t) \times t^{-\alpha} = \text{const}$. Concluimos que:

$$N(t) \propto 1/S(t). \quad (24)$$

Dado (23), $N(t)$ decae asintóticamente como $1/t$ si $d > 2$, como $\ln t/t$ en $d = 2$ y como $1/\sqrt{t}$ en $d = 1$. Es interesante observar que estos resultados se comprueban en simulaciones numéricas.

Podemos comparar estos resultados con lo que predice una ecuación cinética de tipo campo medio que ignora las fluctuaciones espaciales de densidad. Tradicionalmente, la evolución de las densidades medias en reacciones químicas se describe con ecuaciones diferenciales ordinarias. Para el proceso de aniquilación considerado aquí:

$$\frac{d\rho}{dt} = -2K\rho^2, \quad (25)$$

que tiene como solución $\rho(t) \sim 1/t$. Este comportamiento es correcto solamente en $d > 2$, ya que en dimensiones bajas, la recurrencia de las caminatas aleatorias modifica de manera importante el resultado de campo medio. Estos argumentos también se pueden aplicar a otros tipos de reacciones. Para procesos de aniquilación entre dos especies distintas, $A + B \rightarrow \emptyset$, se obtiene $d_c = 4$: los efectos de la difusión son aún más fuertes y siguen presentes en el caso $d = 3$ de mayor relevancia física [2].

3 Movilidad de organismos complejos: perspectivas presentes y futuras

Más de un siglo después de los experimentos clásicos de Jean Perrin, quien rastreaba con precisión el movimiento browniano de partículas microscópicas de látex, los avances tecnológicos de hoy en día permiten realizar experimentos similares y por largos periodos de tiempos (típicamente meses) sobre tiburones, babuinos, tortugas, albatros, roedores, etc... sin olvidar a los humanos. Desde hace apenas algunos años, el interés por el estudio de la movilidad de los organismos vivos ha conocido un crecimiento exponencial. Es sorprendente constatar que se sabe más sobre la difusión de proteínas en una célula que sobre las propiedades estadísticas de los desplazamientos de un humano en una ciudad o de un animal en un ecosistema silvestre. Sin embargo, la movilidad individual es de gran importancia en problemas de dinámica poblacional, de tránsito o de propagación de enfermedades. Dado su carácter interdisciplinario, cada vez es más frecuente que físicos colaboren con biólogos, ecólogos o antropólogos en estos temas.

3.1 Más allá de la difusión simple

Desde un punto de vista teórico, el movimiento de organismos complejos en ambientes no menos complejos plantea varios retos matemáticos y computacionales. Los modelos de caminatas aleatorias antes mencionados son relativamente fáciles de resolver porque se trata de procesos markovianos o sin memoria: la evolución de t a $t + 1$ depende del estado del sistema a t . Sin embargo, muchos procesos en la naturaleza no son markovianos. Por ejemplo, los mamíferos y vertebrados en general tienen memoria y la capacidad de crear mapas mentales, es decir representaciones espacio-temporales de mayor alcance que la vecindad de la posición \mathbf{r} y del tiempo t actuales. Aunque las caminatas aleatorias han sido muy útiles para describir el movimiento animal en escalas temporales cortas [6], el uso de las capacidades cognitivas debería tener repercusiones importantes en varias propiedades dinámicas asintóticas y en lo que los ecólogos llaman “el uso del espacio”. Desafortunadamente estas repercusiones han sido poco exploradas.

Una limitación de las caminatas aleatorias en la naturaleza se manifiesta, por ejemplo, en el poco entendido fenómeno del *ámbito hogareño*. Los animales no se difunden libremente y su desplazamiento cuadrático medio no crece indefinidamente según una ley de Smoluchowski-Einstein (10). Más bien, los individuos en general se quedan confinados en ciertas áreas relativamente pequeñas (“territorios”), a menudo sin fronteras físicas visibles. La probabilidad de ver al animal fuera de esa área es prácticamente nula (ver Figura 1-derecha). De manera similar, muchos humanos deben ir al trabajo y regresar a casa tarde o temprano, lo cual limita los lugares que puede visitar un individuo. En otras palabras, somos *muy* recurrentes. Datos obtenidos de animales equipados de collares con GPS [7]

o de humanos usuarios de teléfonos celulares [8] muestran que el desplazamiento cuadrático medio de un individuo $\langle \mathbf{r}^2 \rangle(t)$ tiende asintóticamente a un valor constante o crece muy lentamente, típicamente con una ley logarítmica de t .

3.1.1 Visitas preferenciales

En [9], se discute cualitativamente en contextos ecológicos un modelo donde un ámbito hogareño emerge a partir de reglas estocásticas simples. Sea un parámetro $0 \leq p \leq 1$. A cada paso de tiempo $t \rightarrow t + 1$, un caminante con posición discreta elige con probabilidad $1 - p$ moverse a uno de los sitios vecinos más cercanos (cada uno con igual probabilidad), o con probabilidad complementaria p , elige un tiempo al azar t' de manera uniforme en $[0, t]$ y salta directamente al sitio donde estaba al instante t' . Por lo tanto, el movimiento combina una exploración local aleatoria estándar y pasos que pueden ser muchos más largos (pero menos frecuentes si $p \ll 1$). Dicha combinación se observa empíricamente en muchas especies y a veces se le denomina movimiento *intermitente* [10]. El presente modelo supone que estos pasos hacia lugares más lejanos se deben al uso intermitente de la memoria (la cual es ilimitada porque todas las posiciones pasadas se recuerdan). Un ejemplo de una trayectoria generada por estas reglas se muestra en la Figura 1b. En una dimensión espacial, la ecuación de recurrencia generaliza la ecuación (1):

$$P_t(n) = \frac{1-p}{2}P_{t-1}(n-1) + \frac{1-p}{2}P_{t-1}(n+1) + \frac{p}{t} \sum_{t'=0}^{t-1} P_{t'}(n). \quad (26)$$

El último término de memoria indica que el sitio n puede ser visitado desde cualquier lugar si ha sido visitado antes. La regla de memoria tiene una interpretación simple: la probabilidad de que el caminante elija un sitio particular es proporcional al número de veces que este sitio fue visitado antes. Un sitio familiar tiene mayor posibilidades (tiempos t') de ser elegido para visitas futuras que un sitio visitado pocas veces. Es similar al principio “el rico se vuelve más rico” o de “vinculación preferencial” muy popular actualmente en la ciencia de las redes complejas. La transformada de Fourier-Laplace discreta⁵ de $P_t(n)$, definida como

$$P(z, \lambda) = \sum_{n=-\infty}^{\infty} \sum_{t=0}^{\infty} z^n \lambda^t P_t(n), \quad (27)$$

se puede obtener analíticamente a partir de (26):

$$P(z, \lambda) = (1 - \lambda)^{-a} [1 - u(1 - p)\lambda]^{-b/u(1-p)}, \quad (28)$$

donde $u = (z + z^{-1})/2$, $a = p/[1 - u(1 - p)]$ y $b = 1 + u(1 - p) - a$. Esta relación es difícil de invertir pero el desplazamiento cuadrático medio $\langle x^2 \rangle(t) \equiv \sum_{-\infty}^{\infty} n^2 P_t(n)$

⁵En las notaciones anteriores, $z = e^{ik}$ y $\lambda = e^{-s}$.

se puede obtener sin conocer $P_t(n)$. Notamos que la transformada de Laplace $\langle x^2 \rangle(\lambda)$ obedece a la relación general $\langle x^2 \rangle(\lambda) = \frac{\partial}{\partial z} (z \frac{\partial}{\partial z} P(z, \lambda))|_{z=1}$. El análisis del límite $\lambda \rightarrow 1$ (equivalente al $s \rightarrow 0$ de la sección anterior) permite determinar el comportamiento asintótico $t \rightarrow \infty$:

$$\langle x^2 \rangle(\lambda) \simeq -\frac{(1-p)}{p(1-\lambda)} [\ln(1-\lambda) - \ln p] \implies \langle x^2 \rangle(t) \simeq \frac{1-p}{p} [\ln(pt) + \gamma], \quad (29)$$

con $\gamma = 0.57721\dots$. Concluimos que la difusión es logarítmicamente lenta en este modelo. La memoria convierte al caminante en muy recurrente, a tal punto de modificar drásticamente, para todo $p > 0$, la ley estándar (10) lineal en t . Un modelo similar basado en visitas preferenciales pudo reproducir patrones de movilidad humana observados recientemente [8].

Una consecuencia importante del principio “el rico se vuelve más rico” es la de generar mucha heterogeneidad entre sitios o entre nodos en caso de redes: hay muchos “pobres” (sitios poco populares/poco visitados) y pocos “ricos” (sitios muy populares/de uso muy rutinario). Un sitio “rico” suele ser mucho más visitado que un sitio típico elegido al azar. Frecuentemente, esta disparidad creada por la asociación preferencial se manifiesta por la emergencia de distribuciones en *leyes de potencias* o *libres de escala*⁶. Sea un sitio elegido al azar dentro de los sitios visitados mínimo una vez por un caminante durante un cierto intervalo de tiempo. En los modelos preferenciales discutidos arriba, la probabilidad $P(k)$ de que este sitio haya sido visitado exactamente k veces sigue una ley de potencia:

$$P(k) \propto k^{-\gamma}, \quad (30)$$

con $\gamma > 1$ un exponente que puede depender de los parámetros del modelo considerado. Es bastante probable que el sitio haya sido visitado solamente 1 vez, pero como la función $x^{-\gamma}$ decae lentamente con x , no es imposible que algunos sitios hayan recibido 10, 100 o más visitas. En el mundo real, dicho uso muy heterogéneo del espacio se observa efectivamente tanto en humanos como en otros animales. Subdividiendo el espacio continuo (por ejemplo una ciudad) en celdas y contando el número de visitas a cada celda por un mismo individuo durante un periodo dado (6 meses, típicamente), la probabilidad $P(k)$ de popularidad definida arriba se aproxima bien por (30) con $\gamma \approx 1.8$ en humanos en EU [8] y $\gamma \approx 1.2$ en monos capuchinos en una selva de Panamá [7]. Unos pocos sitios reciben una enorme cantidad de visitas (la casa y el lugar de trabajo para los humanos). En la Figura 1-derecha, se aprecia cómo un mono capuchino también usa el espacio de manera muy heterogénea. A pesar de la elegancia y éxito del principio de visitas preferenciales para reproducir datos empíricos, no queda muy claro que dicho principio estocástico gobierne exclusivamente el movimiento de organismos complejos.

⁶Vimos un ejemplo de dicha distribución en la sección 2.

3.1.2 Caminatas reforzadas

Otra clase de procesos en donde los sitios visitados en el pasado tienden a ser revisitados preferentemente son las caminatas reforzadas. A diferencia del modelo anterior, los pasos siempre se dan hacia los vecinos más cercanos, por lo tanto no hay pasos largos. Esta propiedad complica el análisis porque la ecuación maestra se convierte en *no lineal*. Lo interesante es que por efectos de no linealidad se pueden presentar fenómenos nuevos como *transiciones de fase* entre regímenes dinámicos muy diferentes cuando algún parámetro rebasa un valor crítico.

Un modelo propuesto recientemente [11] muestra una transición de fase entre un régimen en donde el caminante se localiza estrictamente y un régimen en donde el caminante se difunde sin límites. A cada paso de tiempo con probabilidad $1 - p$ el caminante ubicado en el sitio n salta a un sitio vecino más cercano $m = n \pm 1$ elegido al azar con probabilidad $w_{n \rightarrow m} = 1/2$ (en una dimensión). Con probabilidad p el caminante salta a m con probabilidad $w_{n \rightarrow m} \propto [v_m(t) + 1]^\alpha$ donde $v_m(t)$ es el número de visitas recibidas por el sitio m en el intervalo $[0, t]$ y α una constante. Si $\alpha > 0$, cuanto más visitas haya recibido un sitio, se vuelve más atractivo cuando el caminante pasa cerca⁷; p representa la frecuencia con la cual se toman decisiones basadas en este mecanismo. La ecuación maestra en su forma más general

$$P_{t+1}(n) = \sum_{m \text{ vecino de } n} w_{m \rightarrow n} P_t(m). \quad (31)$$

es ahora no lineal dado que $w_{m \rightarrow n}$ depende del número de visitas pasadas y por lo tanto de P . Tomando el límite continuo y usando la relación $v_x(t) = \int_0^t P(x, t') dt'$, se obtiene la siguiente ecuación de difusión para la densidad de probabilidad de presencia $P(x, t)$:

$$\frac{\partial P(x, t)}{\partial t} = \frac{1}{2} \frac{\partial^2 P(x, t)}{\partial x^2} - \alpha p \frac{\partial}{\partial x} \left\{ P(x, t) \frac{\partial}{\partial x} \ln \left[1 + \int_0^t P(x, t') dt' \right] \right\}. \quad (32)$$

La densidad de probabilidad de una partícula que se difunde libremente en un medio infinito se diluye cuando $t \rightarrow \infty$, es decir $P(x, t) \rightarrow 0$ para todo x , como en (15). A pesar de la complejidad de (32), se pueden buscar condiciones para la existencia de soluciones estacionarias espacialmente localizadas en el límite $t \rightarrow \infty$, es decir un patrón territorial permanente. Sustituyendo en el lado derecho de (32) $P(x, t)$ por una función de x , $P(x) \neq 0$, pero manteniendo en primera aproximación la dependencia temporal del lado izquierdo, se obtiene:

$$\frac{\partial P}{\partial t} \approx \tilde{D} \frac{\partial^2 P}{\partial x^2}, \quad \text{con } \tilde{D} = \frac{1}{2} - \alpha p. \quad (33)$$

Si $\tilde{D} > 0$, las soluciones de (33) son difusivas, ver (32), y se contradice la suposición $P(x, t = \infty) = P(x) \neq 0$. Sin embargo, si $\tilde{D} < 0$, se observa un fenómeno inverso a la difusión⁸: típicamente la partícula se va concentrando en

⁷Si $\alpha < 0$, un caso no discutido aquí, los sitios nunca visitados antes son los más atractivos.

⁸Equivalente a mantener $\tilde{D} > 0$ e invertir el tiempo, $t \rightarrow -t$.

algún punto x_0 a un cierto tiempo, y para tiempos mayores $P(x) = \delta(x - x_0)$. Por lo tanto, fijando α , existe una frecuencia crítica de uso de la memoria $p_c = 1/(2\alpha)$ tal que si $p > p_c$ el caminante se convierte en sedentario en tiempos suficientemente grandes mientras se difunde si $p < p_c$. Esta transición es posible si $p_c < 1$, es decir si $\alpha > 1/2$. Estos resultados se confirman por simulaciones numéricas en dos dimensiones. Algo nuevo mostrado por las simulaciones es que el caminante parece restringir sus visitas a unos cuantos sitios y no a uno sólo [11]. En resumen, en este modelo una región limitada de presencia se auto-organiza dinámicamente debido a la interacción del caminante con su propia trayectoria, por arriba de un umbral crítico de memoria.

3.1.3 Exclusión y efectos de muchos cuerpos

Existen otros mecanismos relevantes para los organismos vivos que pueden causar difusión restringida. Con el propósito de tener un acceso exclusivo a ciertos recursos, muchos animales marcan los sitios que visitan con olores para excluir a otros individuos de la misma especie. Esta interacción repulsiva entre animales tiene conexiones con una clase de procesos muy estudiados en física estadística fuera de equilibrio: los procesos de exclusión. El modelo más sencillo de este tipo consiste en una línea uni-dimensional discreta donde N caminantes aleatorios se difunden sin poder ocupar un mismo sitio (ver Figura 2, arriba). Cuando un caminante se encuentra en un sitio vecino de otro caminante, el único cambio permitido de t a $t+1$ es retroceder en la dirección opuesta (o quedarse inmóvil si está bloqueado entre dos caminantes). En este sistema de muchas partículas en interacción⁹ los efectos de exclusión frenan de manera drástica la difusión individual. Contrariamente al caso una partícula libre, el desplazamiento cuadrático medio no crece linealmente con el tiempo como en (10) sino mucho más lentamente. La ley asintótica exacta fue derivada por Arratia [12]:

$$\langle [x(t) - x(0)]^2 \rangle \simeq \frac{1 - \rho}{\rho} \sqrt{\frac{2t}{\pi}}, \quad t \rightarrow \infty, \quad (34)$$

donde $x(t)$ es la posición de algún caminante y ρ la densidad de caminantes (N dividido por el número de sitios). Este es un resultado notable porque indica que, debido a efectos colectivos, la distancia típica entre un caminante y su punto de origen, $l(t) = \sqrt{\langle [x(t) - x(0)]^2 \rangle}$, crece como $t^{1/4}$ en lugar de $t^{1/2}$ en la difusión normal. Este resultado se puede entender de manera cualitativa en altas densidades ($\rho \sim 1$), donde los caminantes son a menudo inmóviles. En este límite, las vacancias (sitios desocupados, en densidad $1 - \rho$) recorren caminatas aleatorias normales e independientes; dos vacancias se cruzan sin interactuar. Cada vez que una vacancia alcanza desde la izquierda a un caminante (inicialmente ubicado en

⁹O procesos de Markov interactuando en Z .

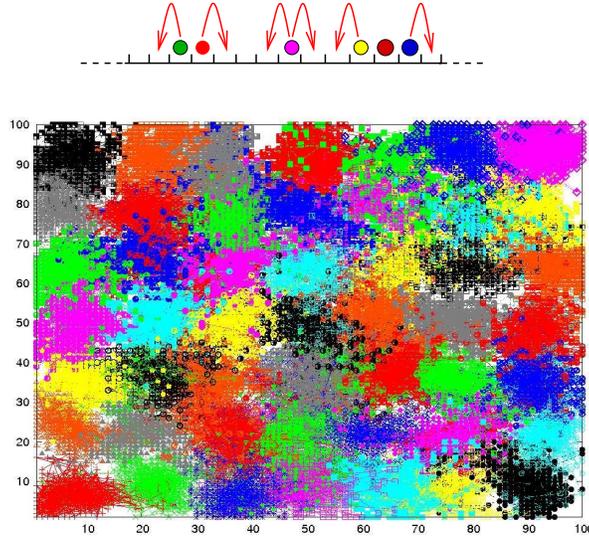


Figure 2: **Arriba:** Proceso de exclusión simétrica entre N caminantes en $1d$. Cada caminante puede saltar hacia uno de sus dos sitios más cercanos con igual probabilidad, siempre y cuando el sitio este desocupado. **Abajo:** Patrón territorial numérico formado por 49 caminantes con tendencia a excluirse (ver texto). Cada mancha de color ilustra el área ocupada por un caminante.

el origen), éste puede dar un paso hacia la izquierda. En un tiempo t , las vacancias recorren típicamente una longitud $l_v \sim t^{1/2}$. Entonces, son $N_v \sim 2(1 - \rho)l_v$ vacancias que pueden alcanzar al caminante focal desde la izquierda o derecha en $[0, t]$. El desplazamiento $x(t) - x(0)$ del caminante se puede ver como una caminata aleatoria de N_v pasos independientes, $x(t) - x(0) \approx \sum_{i=1}^{N_v} \epsilon_i$, con $\epsilon_i = \pm 1$. De (14), deducimos $\langle [x(t) - x(0)]^2 \rangle \approx N_v \langle \epsilon_i^2 \rangle \sim (1 - \rho)t^{1/2}$, en acuerdo cualitativo con (34).

Sin embargo, se trata de un crecimiento algebraico con t , que sigue siendo mucho más rápido que una ley logarítmica como en (29). Se pueden considerar reglas de exclusión más severas y también más realistas. En una extensión de este modelo [13], cada caminante deposita en los sitios discretos que visita un olor que tarda un tiempo T_{AS} en desaparecer. Si un caminante llega a un sitio marcado, donde el olor de otro está todavía activo, éste retrocede. Entonces, las zonas de exclusión pueden ser grandes y no solamente limitadas a la posición actual de los animales. Si el parámetro T_{AS} es $\gg 1$, se forma un patrón espacial de olores activos (es decir, ámbitos hogareños) similares a los mostrados en la Figura 2 (abajo) en un espacio bidimensional. En el caso más simple unidimensional, se muestra que, aunque este mecanismo no elimina por completo la difusión en tiempos muy largos, disminuye fuertemente su amplitud. Se muestra en [13] que las fronteras que separan ámbitos hogareños vecinos se difunden con la ley (34)

modificada con un prefactor exponencialmente pequeño:

$$\langle [x_f(t) - x_f(0)]^2 \rangle \propto e^{-\frac{1}{2}\rho^2 T_{AB}} \sqrt{t}, \quad (35)$$

en donde x_f denota la posición de una frontera. Los dominios se vuelven casi estacionarios si T_{AB} es muy grande. Un efecto cualitativamente similar se observa en 2d. Este modelo con valores realistas de T_{AB} ha sido aplicado a la descripción de la distribución de zorros en el área urbana de Bristol, Reino Unido.

3.2 Biología y modelos basados en agentes

Los modelos expuestos desde el inicio de este capítulo generan trayectorias individuales (posición, tiempo) a partir de reglas dinámicas. Una ventaja de este enfoque es que las trayectorias se pueden simular y analizar de la misma manera que los datos reales, lo cual permite *a posteriori* evaluar la relevancia de las reglas propuestas. Este es el principio de los modelos basados en agentes. Debido a que el uso de las computadoras se está generalizando, en un futuro cercano la complejidad de la reglas a modelar se podrá incrementar prácticamente *ad libitum*, dependiendo del grado de realismo que se desee alcanzar en el estudio de un sistema específico o de los mecanismos de movimiento de interés. Aunque una solución matemática esté fuera de alcance, un modelo computacional puede ser valioso para identificar mecanismos biológicos en un patrón observado y/o para predecir algún fenómeno nuevo.

Una pregunta que no se planteó en las secciones anteriores es ¿ por qué los organismos vivos hacen lo que hacen? Según la teoría de la evolución, si un animal tiene cierto comportamiento al recorrer su ambiente, es que este comportamiento ha evolucionado (por ejemplo con el desarrollo de ciertas capacidades cognitivas) para mejorar la adecuación del animal con su medio y su facultad para sobrevivir. Una limitación de los modelos presentados en las secciones anteriores es que en gran parte ignoran la existencia de un medio ambiente con el cual el animal interactúa. Tomemos el ejemplo de las *búsquedas aleatorias*[14, 10, 15]: consideremos un ambiente con puntos de alimentos distribuidos al azar y un animal que desconoce la ubicación de estos puntos. Supongamos que la alimentación es un factor limitante y que el animal adopta una estrategia que le permite encontrar más lugares de alimentos por unidad de tiempo comparado con lo que haría con otra estrategia (por ejemplo una caminata aleatoria simple), esta estrategia tendrá una ventaja evolutiva. Muchas teorías biológicas recientes (por ejemplo [16]) deberían ser comprobadas o invalidadas con datos de campo y modelos basados en agentes, hacerlo significa un gran reto. A continuación mencionaremos esfuerzos recientes en esta dirección.

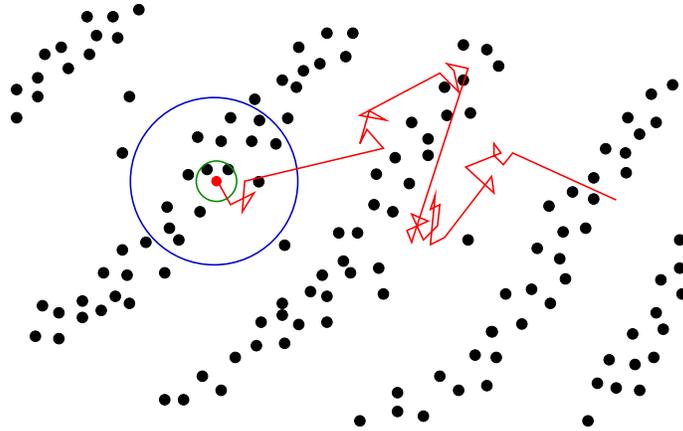


Figure 3: Cada agente de esta población, por ejemplo el señalado en rojo, decide en cada intervalo de tiempo dar un paso aleatorio o quedarse inmóvil. Se supone que la probabilidad de quedarse inmóvil crece cuando hay más vecinos a corta distancia (dentro del círculo pequeño) porque la agregación disminuye los riesgos de mortalidad. Pero si hay demasiados agentes a distancias mayores (dentro del círculo grande), aumenta la probabilidad de que el agente decida moverse para no competir por alimentos. Por estas dos reglas antagonistas, el sistema se auto-organiza rápidamente en bandas, parecidas a las observadas en colonias de mejillones en el mar. La velocidad a la cual se forma este patrón es máxima si los caminantes se mueven combinando pasos cortos con pasos largos menos frecuentes (caminatas de Lévy, como lo muestra la línea roja). Además, la adecuación media de los individuos es también máxima para este tipo de movimiento [17].

3.2.1 ¿Agruparse o andar solo?

Los mejillones, como muchos otros animales, se enfrentan a un dilema: si se agrupan densamente sobre rocas en el mar pueden resistir más fácilmente a factores externos adversos, como corrientes y depredadores, a los cuales un individuo aislado es más vulnerable. Sin embargo, en grupos deben competir por recursos limitados (alimentos), mientras que un mejillón sin vecinos cercanos tiene más recursos para él solo. La Figura 3 ilustra un modelo basado en agentes propuesto recientemente en donde cada mejillón simplemente combina periodos de movimiento aleatorio con periodos de inmovilidad [17]. Se encontró que cierto tipo de caminatas aleatorias (llamadas de Lévy [14]), permiten *maximizar* la adecuación media (o probabilidad de sobrevivencia) de los mejillones. Además, esta estrategia óptima individual hace emerger por efectos de retroalimentación patrones colectivos de agrupación espacialmente *no-uniformes*. Estos patrones están caracterizados por la alternancia periódica de zonas pobladas y poco pobladas en forma de rayas. Finalmente, si se permiten mutaciones aleatorias en el comportamiento, cualquier estado sub-óptimo forzosamente evoluciona en el tiempo hacia el estado óptimo, siendo por lo tanto un punto atractor estable [17].

La importancia de este modelo es que ha sido capaz de explicar con un mismo mecanismo motivado biológicamente, la función de dos observaciones empíricas *a priori* sin conexión: los movimientos de Lévy de mejillones individuales, que han sido analizados cuidadosamente en el laboratorio, y sus patrones de agrupación en el mar en forma de rayas. Además, el modelo hizo la predicción de que esta especie tenía que haber evolucionado hacia un estado óptimo, detalle que no es nada trivial y raramente es considerado en estudios del movimiento animal.

3.2.2 ¿Usar su memoria o explorar al azar?

Ahora consideremos a un animal que se mueve solo en un ambiente con fuentes de alimentos (árboles con fruta), ver la Figura 4. Supongamos, como se observa en muchas selvas, que los árboles no son todos del mismo tamaño: la mayoría son pequeños/poco productivos, pero existen unos cuantos árboles muy grandes/productivos, es decir la distribución espacial de los recursos es heterogénea. Además, aunque siempre hay frutos en el ambiente, un árbol no tiene frutos todo el año sino durante un periodo relativamente corto, que varía entre árboles. Cuando fructifica un árbol, produce cierta cantidad de fruta cada día. Después de un tiempo, los frutos no consumidos se caen y se pudren. Supongamos finalmente que el caminante tiene la capacidad de construirse *mapas mentales*: al moverse, recuerda las posiciones, tamaños y estados de fructificación de los árboles que visita. ¿Cuál es la mejor estrategia a seguir para maximizar la cantidad de frutos ingeridos por unidad de distancia recorrida?

Este es un problema muy complejo de optimización combinatoria (aún más difícil que el Problema del Agente Viajero). Para simplificar, podemos tratar de optimizar las siguientes reglas: a cada paso el caminante se pregunta: ¿Me muevo hacia una dirección al azar o voy hacia un árbol que conozco y que tiene frutos? Si siempre camina al azar, el animal no saca provecho de su memoria: en lugar de dirigirse directamente a un buen árbol ya visitado, lo revisitaría solamente si lo encuentra otra vez por azar. Un resultado poco intuitivo es que tampoco un uso excesivo de la memoria es eficiente, porque en este caso el caminante se limita a una misma zona reducida, ignorando otras regiones con recursos disponibles [18]. Como se muestra en la Figura 4 (derecha), la estrategia óptima combina tanto el azar como el uso de la memoria.¹⁰ Una dosis de exploración al azar es valiosa sobretodo cuando los recursos son efímeros, es decir cuando el caminante necesita actualizar constantemente sus conocimientos. También se demuestra que la memoria es de mayor utilidad en ambientes más heterogéneos espacialmente, lo cual sugiere que ecosistemas complejos pudieron haber propiciado el desarrollo de la memoria en los animales [18].

Debido al uso de mapas mentales, este modelo genera caminatas poco aleatorias y con un alto grado de determinismo. Además de formar un ambito hogareño,

¹⁰Calquiera puede haber experimentado este dilema a la hora de salir a comer y tener que elegir un restaurante.

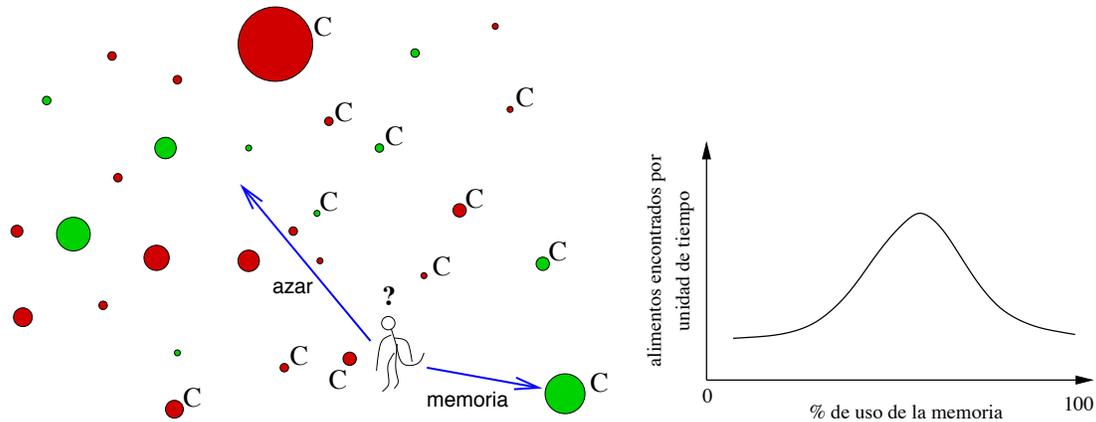


Figure 4: **Izquierda:** Un agente (mono) conoce ciertos árboles (marcados con una “C”) por haberlos visitado en el pasado. Puede decidir usar este mapa mental para visitar un árbol conocido que esté fructificando (color verde), o bien dar un paso al azar, sin certeza de encontrar recursos pero con la posibilidad de incrementar su conocimiento del bosque. Si los árboles fructifican por poco tiempo y de manera estocástica, los recursos son efímeros y el medio es poco predecible. En tal caso, si el agente usa solamente su memoria, puede ser que muchos de los árboles que conoce ya no estén fructificando (de color café) y que ignore por falta de exploración árboles que ya están siendo productivos. **Derecha:** Existe un balance óptimo entre el uso del azar y de la memoria que permite maximizar la cantidad de alimentos ingeridos por unidad de tiempo. (Adaptado de [18]).

las trayectorias son bastante *predecibles* o de baja entropía: el caminante visita frecuentemente secuencias de lugares en el mismo orden. Esta propiedad también caracteriza los trayectos humanos [19].

El tipo de estrategia descrita arriba, que consiste en combinar *explotación* de información disponible con *exploración* para adquirir información nueva, tiene aplicaciones para organismos muy sencillos en otros contextos. Consideremos una bacteria móvil buscando una fuente de alimento fija en un punto. Supongamos que esta fuente emite moléculas (“olores”) que se difunden en el medio y que representan índices que proporcionan información sobre la ubicación de la fuente cuando la bacteria encuentra estas moléculas. Si la fuente emite muy pocas moléculas por unidad de tiempo, la información disponible para la bacteria es muy escasa y desconectada espacialmente. En este caso, la bacteria no puede seguir los gradientes de concentración¹¹ porque éstos son muy pequeños y fragmentados. Una estrategia adecuada para este caso llamada *infotaxis*, ha sido planteada recientemente [20]. A medida que encuentra moléculas, el caminante se va construyendo con el tiempo una idea sobre dónde puede estar la fuente de alimento. Esta información, a un tiempo t , está contenida en una densidad de probabilidad esperada $P(\mathbf{x}, t)$ de que la fuente esté en la posición \mathbf{x} . A esta distribución se le puede asociar una entropía de Shannon, $S = - \int d\mathbf{x} P(\mathbf{x}, t) \ln P(\mathbf{x}, t)$. Cuando más localizada es $P(\mathbf{x}, t)$, menos incertidumbre hay sobre la ubicación de la fuente y más baja es la entropía S . Cada vez que se mueve de un paso (exploración) y encuentre o no una molécula, el caminante usa un modelo de inferencia estadística para actualizar la densidad de probabilidad esperada (y por lo tanto S). La *infotaxis* consiste en moverse en la dirección hacia donde se maximiza la disminución de la entropía S , es decir, se maximiza la ganancia en información a partir de la información previa disponible (explotación). Esta estrategia permite al organismo tener tiempos de búsqueda bastante cortos, aún en ambientes escasos.

4 Conclusiones

En años recientes, el estudio del movimiento de organismos vivos ha permitido ampliar el marco de las caminatas aleatorias, mediante el planteamiento de nuevas preguntas teóricas y la introducción de nuevos modelos con propiedades físicas y matemáticas inéditas. Paralelamente, se están acumulando enormes cantidades de datos de movilidad, que hace unos años eran imposibles adquirir. No cabe duda que las tecnologías de rastreo empleadas en estudios de campo van a progresar aún más. Sin embargo, aún con estos datos, el proceso de validación de las teorías clásicas y modelos conocidos se está apenas iniciando. Solamente el análisis conjunto de modelos y datos permitirá determinar cuáles son los procesos generadores de movimiento y entender su origen. Los datos empíricos motivarán nuevos en-

¹¹Tal estrategia es adecuada cuando las concentraciones son altas y se le conoce como *quimiotaxis*.

foques teóricos y estos últimos permitirán extraer más información cuantitativa de los datos. Una pregunta abierta importante, es si existen principios universales que gobiernen el movimiento de todos los organismos vivos, de bacterias a elefantes. Existen opiniones divergentes al respecto [16, 8, 7], pero esta pregunta tuvo el mérito a nutrir el crecimiento de una nueva rama de las ciencias naturales en la cual la Física puede aportar mucho: la Ecología del Movimiento [16].

Varias de las consideraciones expuestas se pueden aplicar a los Sistemas Complejos en general. Si bien es difícil predecir dónde estaremos en el futuro, podemos afirmar que la investigación en muchos temas ha cambiado significativamente durante las últimas dos décadas. Entre otros aspectos nuevos, existe ahora un mar de datos, muchos de los cuales están disponibles libremente en el Internet y a menudo los podemos encontrar sin haber sido nunca analizados por un científico. De alguna manera, esperan que alguien decifre los secretos que encierran. Entender matemáticamente un modelo puede ser muy satisfactorio y enriquecedor, sin embargo esto no significa necesariamente entender el mundo real. Por citar dos ejemplos entre muchos posibles, los estudios [17] y [21] son fuentes de inspiración y esperanza para futuras investigaciones.

Nota: Debido a la falta de espacio, muchos temas de interés actual y relacionados con la difusión no se pudieron abordar en este capítulo. Simplemente hemos mencionado brevemente los procesos de búsqueda aleatoria y su aplicación al forrajeo animal. Para más detalles sobre las búsquedas de Lévy, intermitentes o de reseteo se recomiendan las referencias [14],[10] y [15], respectivamente. Las búsquedas bayesianas se exponen en [22]. El estudio de los efectos de la movilidad de individuos sobre la propagación de enfermedades en humanos y animales es un tema de intensa investigación actual [23]. Modelos epidemiológicos basados en agentes móviles pueden ayudar a diseñar políticas de conservación de primates en parques de África [24]. En [21] se estudia con un enfoque similar, la lenta propagación de palabras nuevas en un país (Japón).

References

- [1] Chandrasekhar S (1943) Stochastic problems in physics and astronomy. *Reviews of modern physics* 15: 1–89.
- [2] Krapivsky P, Redner S, Ben-Naim E (2010) *A kinetic view of statistical physics*. Cambridge University Press.
- [3] Hughes B, Shlesinger M, Montroll E (1981) Random walks with self-similar clusters. *Proceedings of the National Academy of Sciences* 78: 3287–3291.
- [4] Redner S (2001) *A guide to first-passage processes*. Cambridge University Press.

- [5] Hughes B (1995) *Random Walks and Random Environments: Volume 1: Random Walks*. Clarendon Press, Oxford.
- [6] Smouse P, Focardi S, Moorcroft P, Kie J, Forester J, et al. (2010) Stochastic modelling of animal movement. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 2201–2211.
- [7] Boyer D, Crofoot M, Walsh P (2012) Non-random walks in monkeys and humans. *Journal of The Royal Society Interface* 9: 842–847.
- [8] Song C, Koren T, Wang P, Barabási A (2010) Modelling the scaling properties of human mobility. *Nature Physics* 6: 818–823.
- [9] Börger L, Dalziel B, Fryxell J (2008) Are there general mechanisms of animal home range behaviour? a review and prospects for future research. *Ecology Letters* 11: 637–650.
- [10] Bénichou O, Loverdo C, Moreau M, Voituriez R (2011) Intermittent search strategies. *Rev Mod Phys* 83: 81–129.
- [11] Choi J, Sohn J, Goh K, Kim I (2012) Modeling the mobility with memory. *EPL (Europhysics Letters)* 99: 50001.
- [12] Arratia R (1983) The motion of a tagged particle in the simple symmetric exclusion system on Z^1 . *The Annals of Probability* 11: 362–373.
- [13] Giuggioli L, Potts J, Harris S (2011) Animal interactions and the emergence of territoriality. *PLoS Computational Biology* 7: e1002008.
- [14] Viswanathan G, Raposo E, Da Luz M (2008) Lévy flights and superdiffusion in the context of biological encounters and random searches. *Physics of Life Reviews* 5: 133–150.
- [15] Evans M, Majumdar S (2011) Diffusion with stochastic resetting. *Physical Review Letters* 106: 160601.
- [16] Nathan R, Getz W, Revilla E, Holyoak M, Kadmon R, et al. (2008) A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences* 105: 19052–19059.
- [17] de Jager M, Weissing F, Herman P, Nolet B, van de Koppel J (2011) Lévy walks evolve through interaction between movement and environmental complexity. *Science* 332: 1551–1553.
- [18] Boyer D, Walsh P (2010) Modelling the mobility of living organisms in heterogeneous landscapes: does memory improve foraging success? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368: 5645–5659.

- [19] Song C, Qu Z, Blumm N, Barabási A (2010) Limits of predictability in human mobility. *Science* 327: 1018–1021.
- [20] Vergassola M, Villermaux E, Shraiman B (2007) Infotaxis as a strategy for searching without gradients. *Nature* 445: 406–409.
- [21] Lizana L, Mitarai N, Sneppen K, Nakanishi H (2011) Modeling the spatial dynamics of culture spreading in the presence of cultural strongholds. *Physical Review E* 83: 066116.
- [22] Hein A, McKinley S (2012) Sensing and decision-making in random search. *Proceedings of the National Academy of Sciences* 109: 12070–12074.
- [23] Belik V, Geisel T, Brockmann D (2011) Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X* 1: 011001.
- [24] Bonnell T, Sengupta R, Chapman C, Goldberg T (2010) An agent-based model of red colobus resources and disease dynamics implicates key resource sites as hot spots of disease transmission. *Ecological Modelling* 221: 2491–2500.

Propiedades de los plasmas

F. Castillo¹, H. Martínez¹, O. Flores¹, B. Campillo²

¹Laboratorio de Espectroscopia, Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Apartado Postal 48-3, 62191 Cuernavaca, Morelos, México. hm@fis.unam.mx

²Facultad de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, México, D.F.

1.1.- Ionización en gases.

Los materiales aislantes eléctricos (o dieléctricos) son materiales en los que los campos electrostáticos pueden permanecer casi indefinidamente. Estos materiales ofrecen así una resistencia muy alta al paso de corrientes continuas. Sin embargo, no pueden soportar un voltaje infinitamente grande. Cuando el voltaje aplicado a través del dieléctrico excede un valor crítico el aislamiento se daña. Los dieléctricos pueden ser gaseosos, líquidos o sólidos. Los dieléctricos gaseosos en la práctica no están libres de partículas cargadas eléctricamente, como los electrones libres. Los electrones, que pueden ser producidos por irradiación o por emisión de campo, pueden conducir al inicio del proceso de rompimiento. Estos electrones libres, a pesar de ser producidos por la aplicación de un campo eléctrico son acelerados desde el cátodo hasta el ánodo por el campo eléctrico que aplica una fuerza sobre ellos. Ellos adquieren una energía cinética ($\frac{1}{2} m v^2$) a medida que avanzan por el campo. La energía se expresa generalmente como un voltaje (en electrón-voltios, eV, donde e es la carga de un electrón) como las energías implicadas son extremadamente pequeñas. [la energía $E_i = eV_i$ se expresa en electrón-volts. $1\text{eV}=1,6 \times 10^{-19}$ J]. Estos electrones libres, que se mueven hacia el ánodo chocan con las moléculas de gas que se encuentran entre los electrodos. En estas colisiones, parte de la energía cinética de los electrones se pierde y parte se transmite a la molécula neutra. Si esta molécula gana energía suficiente (más de la E_i energía necesaria para que se produzca la ionización), puede ionizar por colisión. El número (promedio) de colisiones ionizantes por un electrón por cada unidad de deriva a través del espacio entre los electrodos no es una constante, sino está sujeta a fluctuaciones estadísticas. El electrón recién liberado y el electrón que incidente son acelerados por el campo y esto establece una avalancha de electrones. La ionización aumenta rápidamente con el voltaje, una vez que estos procesos secundarios tienen lugar, hasta que finalmente tiene lugar la ruptura.

Vale la pena señalar que en campos uniformes, la ionización presente a voltajes por debajo de la ruptura normalmente es demasiado pequeña, como para afectar a las aplicaciones en ingeniería. Para campos no uniformes, sin embargo, una considerable ionización puede estar presente en la región de alto campo eléctrico, con voltajes muy por debajo de la ruptura, constituyendo la descarga bien conocida como corona.

1.1.1 Los procesos de ionización en emisiones de gases.

La ruptura eléctrica de un gas se produce por diversos procesos de ionización. Se trata de procesos que implican la colisión de electrones, iones y fotones con moléculas de gas, y los procesos de electrodos que tienen lugar en o cerca de la superficie del electrodo [electrones pueden ser emitidos desde el cátodo si el campo eléctrico es alrededor de 100 - 1000 kV/cm debido a emisión de campo].

Ionización es el proceso mediante el cual se le quita un electrón a un átomo, dejando al átomo con una carga positiva (iones positivos). Dado que un electrón en la órbita más externa está sujeto a la menor fuerza atractiva del núcleo, es el más fácil de retirar por cualquiera de los procesos de

colisión. La energía requerida para separar un electrón exterior completamente de su estado normal en el átomo a una distancia más allá del núcleo se denomina primer potencial de ionización.

El proceso recíproco de que un electrón desde una gran distancia caiga a la órbita más baja desocupada, también es posible. En este caso, un fotón es emitido con la misma energía que previamente fue absorbida.

1.1.2 Procesos de ionización

(i) Ionización por colisión

Cuando la energía cinética de un electrón ($\frac{1}{2} mu^2$), en colisión con una molécula del gas neutro excede la energía de ionización ($E_i = eV_i$) de la molécula, entonces puede ocurrir ionización. (es decir, cuando la media $\frac{1}{2}mu^2 > E_i$)



en general, resultará un ion positivo y 2 electrones que se mueven lentamente. La probabilidad de este proceso es cero para energías de los electrones igual a la energía de ionización E_i , pero aumenta casi linealmente al principio, y luego gradualmente con la energía de los electrones hasta un máximo.

Cuando moléculas del gas se bombardean con electrones, los electrones unidos a las moléculas pueden ser liberados por el choque con los electrones de alta energía. La razón entre los electrones producidos por colisión y los electrones incidentes, dependerá principalmente de la energía de los electrones incidentes. Esta será máxima a energías de electrones incidentes de aproximadamente 200 a 500 eV. Para los valores más pequeños de estas energías, la energía transferida puede no ser suficiente para causar que los electrones escapen de las moléculas, y por lo tanto la probabilidad de ionización será pequeña. Para valores mucho más grandes de energía incidente, la energía del electrón que colisiona sería suficiente para que este electrón pueda penetrar la superficie más profundamente en la molécula, por lo que de nuevo la posibilidad de escape de otros electrones disminuya.

Así, la variación de la probabilidad de ionización en el aire con respecto a la energía incidente de los electrones es como se muestra en la figura 1.

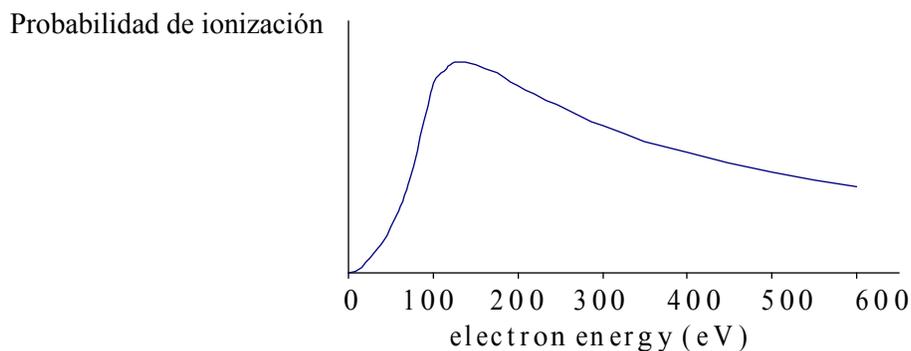


Figure 1 – Curva de probabilidad de ionización de aire por impacto electrónico.

(ii) Excitación

En el caso de una colisión, la molécula de gas neutro no siempre se ioniza por impacto electrónico. En tales casos, la molécula se queda en un estado excitado M^* , con una energía E_e .



Esta molécula excitada posteriormente puede emitir un fotón de frecuencia ν con una energía emitida de $h\nu$. La energía se da cuando el electrón salta de una órbita a otra.



donde h es la constante de Planck = $6,624 \times 10^{-34}$ J-s.

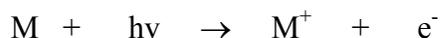
(iii) Ionización por doble impacto electrónico

Si una molécula de gas esta en un estado excitado (con energía E_e) debido a colisión anterior con un electrón, a continuación, esta molécula puede ionizarse por una colisión con un electrón relativamente lento. Este electrón necesitaría menos energía que la energía de ionización, pero la energía deberá superar la energía adicional necesaria para alcanzar la energía de ionización (Es decir, $\frac{1}{2} \mu v^2 > E_i - E_e$)



(iv) Foto-ionización

Una molécula en el estado fundamental puede ser ionizada por un fotón de frecuencia ν , siempre y cuando la energía emitida $h\nu$ (energía para producir un salto de electrones desde una órbita a otra), sea mayor que la energía de ionización de la molécula (Es decir, $h\nu > E_i$).



(v) Formación de iones negativos

Si una molécula de gas tiene niveles de energía desocupados en su órbita más externa, entonces uno de los electrones incidentes pudiera ocupar uno de estos niveles, convirtiendo la molécula en un ión negativo M^- .



El ion negativo así formado estaría en un estado excitado, producido por el exceso de energía.

Nota: los electrones disminuye el número de los electrones libre, a diferencia de ionización que aumenta los electrones libres.

(vi) Desajo electrónico

Esto ocurre cuando un ion negativo cede su electrón extra, y se convierte en una molécula neutra.



(vii) Otros procesos

Los procesos anteriores son los más importantes en relación con los fenómenos de descarga. Otros procesos posibles incluye colisiones ion-átomo, colisiones átomo excitado-molécula, y colisiones átomo-átomo. Cabe señalar que las colisiones entre iones y átomos raramente resultar en ionización, debido al relativamente bajo tiempo de interacción, lo que permite al movimiento

interno del sistema atómico ajustarse el mismo gradualmente a la condición de cambiar sin que ocurra ningún tipo de transición de energía.

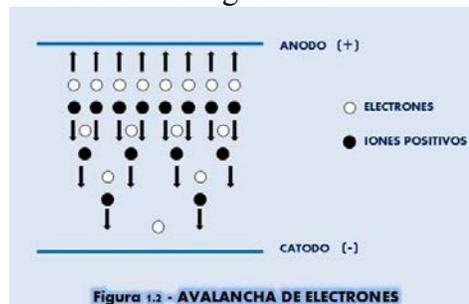
Para causar la ionización de un átomo neutro en el estado base de su propio tipo, un ion positivo debe poseer una energía de al menos 2 eV. Normalmente, los iones y los átomos que tienen estas energías se encuentran sólo en arcos de alta corriente y descargas termonucleares.

1.2 Características de una ruptura en gases.

Se conocen dos mecanismos de ruptura en gases. Estos son la avalancha y los mecanismos de serpiente.

1.2.1 Mecanismo de avalancha de electrones (Proceso de ruptura Townsend).

Uno de los procesos que se consideran en la ruptura es el mecanismo de ruptura Townsend. Está basado en la generación de avalanchas sucesivas secundarias para producir la ruptura. Supóngase que existe un electrón libre (producido por algún efecto externo, tal como la radioactividad o la radiación cósmica) en un gas, donde existe un campo eléctrico. Si la intensidad de campo es suficientemente grande, entonces es probable que se pueda ionizar una molécula de gas por colisión, dando como resultado 2 electrones libres y un ion positivo. Estos 2 electrones serán capaces de causar más ionizaciones por la colisión, produciendo en general, a 4 electrones y 3 iones positivos. El proceso es acumulativo, y el número de electrones libres seguirá aumentando a medida que continúan moviéndose bajo la acción del campo eléctrico. El enjambre de electrones e iones positivos producidos de esta forma se llama avalancha de electrones. En el espacio de unos cuantos milímetros, puede crecer hasta que contener muchos millones de electrones. Este mecanismo se muestra en la Figura 1.2.



1.2.1.1 Análisis Matemático

Cuando el voltaje aplicado a través de un par de electrodos se incrementa, la corriente entre los electrodos aumenta lentamente, como los electrones emitidos desde el cátodo se mueven a través del gas con una velocidad media determinada por su movilidad por la intensidad de campo existente para un valor particular de voltaje. La ionización por impacto electrónico es probablemente el proceso más importante en la ruptura de los gases, pero este solo proceso no es suficiente para producir una rompimiento.

Tomando

n_0 = Número de electrones/segundo emitido desde el cátodo,

n_x = número de electrones/segundo que se mueven una distancia x del cátodo

[$n_x > n_0$, debido a las colisiones ionizantes entre los electrodos]

α = número de colisiones ionizantes, en promedio, producidas por un electrón por la unidad de deriva en la dirección del campo. [primer coeficiente de ionización de Townsend]

luego $1/\alpha$ = distancia media recorrida en la dirección del campo entre las colisiones ionizantes.

Considerando una lámina de espesor dx a una distancia x del cátodo. Los n_x electrones que entran a la lamina, atravesaran esta en la presencia del campo aplicado E . Las colisiones ionizantes generadas en el gas entre los electrodos serán proporcionales tanto a dx como n_x . Así

$$\begin{aligned} dn_x &\propto n_x \\ &\propto dx \end{aligned}$$

Por lo tanto $dn_x = \alpha n_x dx$ (De la definición de α)

Re-arreglando e integración se tiene

$$\int_{n_0}^{n_x} \frac{dn_x}{n_x} = \alpha \int_0^x dx$$

$$\ln\left(\frac{n_x}{n_0}\right) = \alpha x$$

$$n_x = n_0 e^{\alpha x}$$

Si el ánodo está a una distancia $x = d$ del cátodo, entonces el número de electrones n_d golpeando el ánodo por segundo viene dada por

$$n_d = n_0 e^{\alpha d}$$

Por lo tanto, en promedio, cada electrón emitido en el cátodo produce $(n_d - n_0)/n_0$ nuevos electrones (y los correspondientes iones positivos) entre los electrodos.

En el estado estacionario, el número de iones positivos que llegan a él cátodo/segundo debe ser exactamente igual al número de electrones recién formados que llegan a él ánodo. Así, la corriente del circuito estará dado por

$$I = I_0 e^{\alpha d}$$

donde I_0 es la corriente foto-eléctrica en el cátodo.

En el proceso de rompimiento, la ionización por impacto de electrones es atendido por procesos secundarios en el cátodo, que reponen el espacio entre electrodos con electrones libres, con cada avalancha recién formado superando la anterior en el número de electrones.

Considérese ahora las ecuaciones de crecimiento con el mecanismo secundario también presente.

Sea

γ = número de electrones secundarios (en promedio) producido en el cátodo por colisiones ionizantes en el espacio entre electrodos. [Segundo coeficiente de ionización de Townsend]

n_0 = número inicial de foto-electrones/segundo emitidos desde el cátodo.

n_0' = número de electrones secundarios/segundo producidos en el cátodo.

n_0'' = número total de electrones/segundo dejando el cátodo.

Entonces $n_0' = n_0 + n_0''$

En promedio, cada electrón dejando el cátodo produce $[e^{\alpha d} - 1]$ colisiones en el espacio entre

electrodos, dando el número de colisiones ionizantes por segundo en el espacio entre electrodos como $n_0''(e^{\alpha d}-1)$. Así, por definición

$$\gamma = \frac{n_0'}{n_0''(e^{\alpha d}-1)}$$

dando, $n_0' = \gamma n_0''(e^{\alpha d}-1)$, pero como $n_0'' = n_0 + n_0'$, de modo que $n_0'' = n_0 + \gamma n_0''(e^{\alpha d}-1)$ Esto da el resultado

$$n_0'' = \frac{n_0}{1-\gamma(e^{\alpha d}-1)}$$

Similar para el caso del proceso primario (con solo α), se tiene

$$n_d = n_0'' e^{\alpha d} = \frac{n_0 e^{\alpha d}}{1-\gamma(e^{\alpha d}-1)}$$

Así, en estado estacionario, la corriente del circuito (I) será

$$I = \frac{I_0 e^{\alpha d}}{1-\gamma(e^{\alpha d}-1)}$$

Esta ecuación describe el crecimiento de la corriente promedio entre los electrodos antes de que ocurra la chispa de ruptura.

A medida que el voltaje aplicado aumenta, $e^{\alpha d}$ y $\gamma e^{\alpha d}$ aumentan hasta $\gamma e^{\alpha d} \rightarrow 1$, cuando el denominador de la expresión del circuito de corriente llega a cero y la corriente $I \rightarrow \infty$. En este caso, la corriente, en la práctica, se limita solamente por la resistencia de la fuente de alimentación y la conducción de gas.

Esta condición por lo tanto se puede definir como la ruptura y puede escribirse como

$$\gamma(e^{\alpha d}-1)=1$$

Esta condición se conoce como criterio de Townsend para la chispa de ruptura.

La avalancha de la ruptura se desarrolla durante períodos relativamente largos de tiempo, por lo general sobre 1 μ s y no suele ocurrir con voltajes impulsivos.

1.2.1.2 Determinación de los coeficientes de Townsend α y γ .

Los coeficientes de Townsend son determinados en una cámara de ionización, que primero se evacua a un vacío muy alto del orden de 10^{-4} y 10^{-6} Torr antes de llenar con el gas deseado a una presión de unos pocos Torr. El voltaje directo aplicado es aproximadamente entre 2 y 10 kV, y el sistema de electrodos consta de un electrodo plano de alto voltaje y un electrodo de bajo voltaje rodeado por un electrodo de protección para mantener un campo uniforme. El electrodo de bajo voltaje está conectado a tierra a través de un electrómetro capaz de medir las corrientes en un intervalo de 0,01 pA a 10 nA. El cátodo se irradia utilizando una lámpara ultravioleta desde el exterior para producir la iniciación de electrones. Las curvas características de voltaje-corriente se obtienen a diferentes separaciones de los electrodos. A bajos voltajes el crecimiento de la

corriente no es constante. Después, el proceso de Townsend constante se desarrolla como se muestra en la figura 1.3.



De el mecanismo de Townsend, la corriente de descarga está dada por

$$I = \frac{I_0 e^{\alpha d}}{1 - \gamma(e^{\alpha d} - 1)} \approx I_0 e^{\alpha d} \quad \text{cuando} \quad \alpha d \gg 1$$

Esto se puede escribir en forma logarítmica como

$$\ln(I) = \alpha d + \ln(I_0)$$

$$y = m x + c$$

De una gráfica de $\ln(I)$.vs. d , las constantes I_0 y α , se pueden determinar a partir de la pendiente y la intersección respectivamente, como se muestra en la figura 1.4.



Figura 1.4 - $\ln(I)$ en función de d

Una vez que I_0 y α sean conocidos, γ puede ser determinada a partir de puntos de la región ascendente de la curva. El experimento se puede repetir a diferentes presiones, si es necesario.

1.2.1.3 Ruptura en gases electronegativos

En el análisis anterior, la formación de moléculas y/o átomos negativos no se consideró. La formación de iones negativos disminuye electrones libres y por lo tanto da gases de resistencias

dieléctricas muy altas. Los gases que producen iones negativos, se les conoce como gases electro-negativos.

Se puede definir un coeficiente de **fijación** η , de forma análoga a α , como el número de fijaciones por electrón por unidad de deriva en la dirección del campo. Bajo estas condiciones, la ecuación para el crecimiento de corriente media en un campo uniforme está dada por

$$I = I_0 \frac{\left[\frac{\alpha}{\alpha - \eta} e^{(\alpha - \eta)d} - \frac{\eta}{\alpha - \eta} \right]}{I - \gamma \frac{\alpha}{\alpha - \eta} \left[e^{(\alpha - \eta)d} - 1 \right]}$$

Los criterios correspondientes para la distribución de chispa son

$$\gamma \frac{\alpha}{\alpha - \eta} \left[e^{(\alpha - \eta)d} - 1 \right] = 1$$

1.2.2 Ley de Paschen

Cuando los electrones y los iones se mueven a través de un gas en un campo uniforme (E) y a una presión (p) del gas, sus energías medias alcanzan los valores de equilibrio dependiendo de la relación E / P; o más precisamente

$$\alpha / p = f_1(E/p) \quad \text{and} \quad \gamma = f_2(E/p)$$

Para una campo uniforme, el campo eléctrico esta dado por $E=V/d$. Así, aplicando los criterios de Townsend para la chispa de rompimiento en gases da

$$\gamma (e^{\alpha d} - 1) = 1$$

puede escribirse en términos de las funciones como

$$f_2 \left(\frac{V}{pd} \right) \left[e^{pd f_1 \left(\frac{V}{pd} \right)} - 1 \right] = 1$$

Esta ecuación muestra que el voltaje de ruptura V es una función implícita del producto de p (presión del gas) y la separación (d) entre electrodos. Esto es $V = f(p.d)$

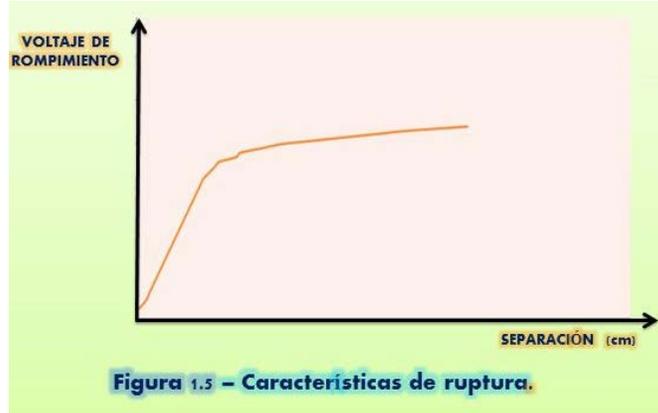
En la derivación anterior el efecto de la temperatura sobre el voltaje de ruptura no se tomo en cuenta. Usando la ecuación de los gas: (presión)(volumen) = (masa)(temperatura absoluta)R, se tiene: presión = (densidad) R (temperatura absoluta). Así, la ecuación correcta de la expresión anterior es: $V = f(p.d)$, donde ρ es la densidad del gas. Esta es la ecuación de la Ley de Paschen. Bajo condiciones atmosféricas constantes, se encontró que experimentalmente el voltaje de ruptura en un campo uniforme, se puede expresar en la forma

$$V = Ad + B\sqrt{d}, \quad \text{donde } d \text{ es la separación de los electrodos.}$$

Para el aire, bajo condiciones normales, $A=24,4 \text{ kV/cm}$ y $B=6,29 \text{ kV/cm}^{1/2}$. [El gradiente voltaje de ruptura es de aproximadamente 30 kV/cm en campos uniformes en

separaciones entre electrodos del orden de 1 cm, pero se reduce a alrededor de 6 kV/cm para separaciones del orden de varios metros]

La Figura 1.5 muestra la relación entre el voltaje de ruptura y la separación de los electrodos.



Esta variación del voltaje de ruptura con la separación de los electrodos se puede modificar usando la Ley de Paschen para incluir la variación con la densidad del gas de la siguiente manera.

$$V = \frac{A}{\rho_0} \rho d + \frac{B}{\rho_0^{1/2}} (\rho d)^{1/2} = A(\delta d) + B(\delta d)^{1/2}$$

donde δ es la densidad relativa (o factor de corrección de la densidad del gas).

Esta ecuación se aplica a separación de más de 0,1 mm. Sin embargo, a productos muy pequeños de (ρd), se produce un voltaje mínimo de ruptura, conocido como mínimo el Paschen. Esto puede explicarse de la siguiente manera.

Considérese una separación fija, y permita que la presión disminuya (un valor a la derecha del mínimo, ver Figura 1.6).



La densidad por lo tanto, disminuye y un electrón que se mueve en el campo por consiguiente, hace menos colisiones con las moléculas de gas a medida que viaja hacia el ánodo. Puesto que cada colisión resulta en una pérdida de energía, se deduce que basta con un campo eléctrico pequeño para producir electrones con una energía cinética ($\frac{1}{2} m u^2$), energía requerida para la ionización por electrónico.

Gas	$V_{\min}(\text{V})$	Pd (Torr-cm) a V_{\min}
Aire	327	0.57
Argón	137	0.90
Hidrogeno	273	1.15
Helio	156	4.00
Dióxido de carbono	420	0.51
Nitrógeno	251	0.67
Oxido nitroso	418	0.50
Oxigeno	450	0.70
Dióxido de azufre	457	0.33
Hidrogeno sulfuroso	414	0.60

Tabla 1.1. Voltajes mínimos de ruptura.

Cerca del mínimo, la densidad es baja y hay relativamente pocas colisiones. Es necesario ahora tomar en cuenta el hecho de que un electrón no necesariamente ioniza una molécula en colisión con ella, incluso si la energía del electrón supera la energía de ionización. Los electrones tienen una probabilidad finita para producir una ionización, que depende de su energía. Si la densidad y por lo tanto el número de colisiones se reduce, la ruptura sólo puede ocurrir si la probabilidad de producir una ionización se incrementa, y esto explica el aumento del voltaje a la izquierda del mínimo.

Vale la pena señalar que si la densidad se fija, la ruptura a la izquierda del mínimo se produce más fácilmente (es decir, a menor voltaje) a distancias más largas. Típicamente, el voltaje mínimo es de 300 V y se produce a pd igual 5 Torr-mm, o a una separación de electrodos de aproximadamente 0,06 mm.

A presiones muy bajas, y a presiones muy altas (comparadas con la atmosférica), la Ley de Paschen falla. Además, la Ley de Paschen es válida para temperaturas inferiores a 1100 °C. Un aumento de la temperatura se traduce en el fracaso de la Ley de Paschen, debido a la ionización térmica.

Se puede observar a partir de la característica ruptura, que para una separación de electrodos constante, el voltaje de ruptura, y por tanto el campo eléctrico de ruptura necesario, es mucho mayor que a condiciones atmosféricas, bajo condiciones de una presión muy alta y bajo presiones muy bajas (alto vacío). Así, tanto a alto vacío, como a alta presión de gas se puede utilizar como medios aislantes. [La región de ruptura de vacío es aquel en el que el voltaje de ruptura se hace independiente de la presión del gas] Bajo condiciones normales, el funcionamiento está bien a la derecha del mínimo de Paschen.

La tabla 1.1 presenta el potencial de chispa mínimo de varios gases.

1.2.3 Mecanismo de serpiente

Este tipo de ruptura se deriva principalmente por el efecto añadido del campo de espacio-carga de una avalancha y de ionización fotoeléctrica en el volumen de gas. Aunque el mecanismo Townsend predice una forma muy difusa de descarga, en la práctica muchas descargas se han encontrado que son filamentosas e irregulares. La teoría de serpiente predice el desarrollo de

una descarga de chispa directamente de una simple avalancha. El espacio-carga producido en la avalancha provoca una distorsión suficiente del campo eléctrico que hace que los electrones libres se mueven hacia la cabeza avalancha, y al hacerlo así generar más avalanchas en un proceso que rápidamente se convierte en acumulativo. Como los electrones avanzan rápidamente, los iones positivos son dejados atrás en una cola moviéndose relativamente lenta. El campo se incrementara en frente de la cabeza. Justo detrás de la cabeza del campo entre los electrones y los iones positivos es en la dirección opuesta al campo aplicado y, por tanto la intensidad de campo resultante será menor. De nuevo entre la cola y el cátodo el campo se verá incrementado. (Figura 1.7)

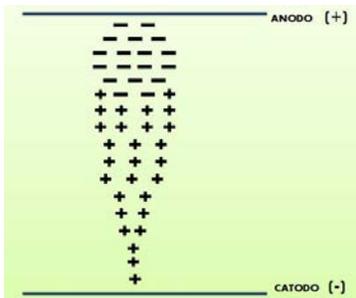


Figura 1.7- Mecanismos de serpiente.

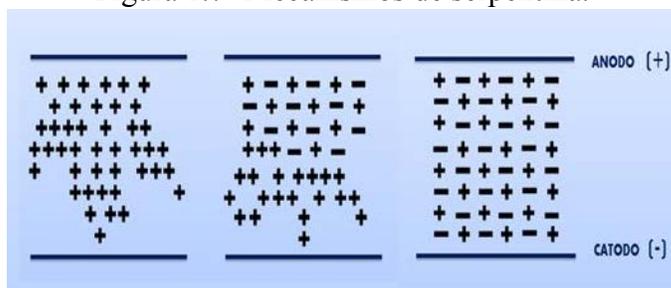


Figura 1.8 - Rompimiento serpiente.

Debido al aumento del campo entre la cabeza y el ánodo, el espacio-carga aumenta, causando una mejora adicional del campo alrededor del ánodo. El proceso es muy rápido y el espacio-carga positivo se extiende hacia el cátodo muy rápidamente dando como resultado la formación de una serpiente. La Figura 1.8 muestra el proceso de rompimiento.

1.2.4 Factores que afectan el voltaje de ruptura

Vacío es idealmente el mejor aislante, con campos eléctricos de rompimiento del orden de 10^4 kV/c. El voltaje de ruptura de un alto vacío es el voltaje el cual cuando aumenta en una pequeña cantidad producirá la ruptura del sistema, el cual se mantuvo a ese potencial durante un tiempo infinito. Sin embargo, esta definición no siempre es posible ya que la ruptura está afectada por muchos factores.

(i) Separación de los electrodos

Para separaciones de electrodos menores que aproximadamente 1 mm, el voltaje de ruptura es aproximadamente proporcional a la longitud, todos los demás parámetros permanecen constantes. Esto da un campo eléctrico de ruptura constante. Para estas pequeñas separaciones, el campo eléctrico de rompimiento es relativamente grande, siendo del orden de 1 MV/cm. La emisión de campo de electrones probablemente juega un papel importante en el proceso de ruptura.

$$V = kd \quad \text{para} \quad d < 1 \text{ mm}$$

Para separaciones mayores que aproximadamente 1 mm, el voltaje de ruptura no aumenta a una tasa igual y el campo eléctrico de ruptura aparente para separaciones mas grandes es mucho mas reducido, siendo aproximadamente de 10kV/cm a una distancia de 10 cm. [El campo eléctrico aparente se define como el voltaje requerido para causar la ruptura dividido por la distancia entre los electrodos]. Figura 1.9.

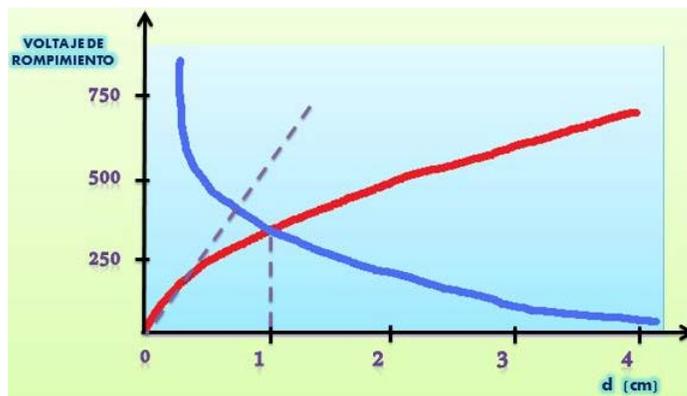


Figura 1.9 - Desglose de característica.

Cranberg ha demostrado teóricamente que para separaciones grandes, es el producto de voltaje de ruptura y el campo eléctrico los que se mantiene constante.

$$VE = k_1, \quad \text{para } d > 1 \text{ mm.}$$

donde la constante k_1 depende del material y condiciones de la superficie de los electrodos.

Para un campo eléctrico uniforme, $E=V/d$, de modo que $V = k^2 d^{1/2}$

O en una forma más general, ambas regiones se pueden expresar por la ecuación

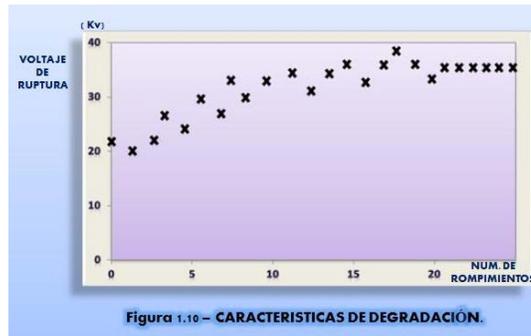
$$V = k d^x, \quad \text{donde } x = 0,5 \quad \text{para grandes gaps } > 1 \text{ mm}$$

$$x = 1 \quad \text{para gaps } < 1 \text{ mm}$$

(ii) Efecto de los electrodos

Acondicionamiento

El voltaje de ruptura de un sistema aumenta en descargas disruptivas sucesivas, hasta que se alcanza un valor constante. Se dice entonces que los electrodos están acondicionados. Este aumento en el voltaje se atribuye a la quema por chispas de irregularidades microscópicas o impurezas que puedan existir en los electrodos. Al investigar el efecto de diversos factores en la ruptura, los electrodos primero debe estar acondicionado de tal manera que se puedan obtener resultados reproducibles.



El efecto de acondicionamiento se muestra en la figura 1,10. El voltaje de ruptura de electrodos acondicionados o separaciones es mucho más reproducible que otra cosa, y por lo tanto los valores de ruptura que son normalmente obtenidos experimentalmente son los de electrodos acondicionados. Electrodos no condicionados pueden tener valores de ruptura tan bajos como 50% del voltaje de ruptura de electrodos acondicionado.

(iii) El material y el acabado superficial.

Material de los electrodos	Voltaje (kV) a través de una separación de 1 mm
Acero	122
Acero inoxidable	120
Niquel	96
Metal monel	60
Aluminio	41
cobre	37

Tabla 1.2 - Efecto de material de electrodo sobre la ruptura.

Las superficies de los electrodos forman los límites físicos entre las cuales la ruptura finalmente se lleva a cabo. Así es que no es sorprendente encontrar que el campo eléctrico de ruptura de una separación de tamaño dado es fuertemente dependiente del material de los electrodos. En general, mientras más suave es el acabado de la superficie, mayor es el voltaje de ruptura.

(iv) Superficie contaminada.

La presencia de contaminación en los sistemas de prueba reduce el voltaje de ruptura, algunas veces tanto como un 50% del valor que se obtiene con electrodos limpios.

(v) Configuración de electrodos y área.

El aumento de la superficie de los electrodos hace que sea más difícil mantener un voltaje de ruptura dado. Así voltaje de ruptura disminuye ligeramente con aumento de la superficie. Por ejemplo, los electrodos de 20 cm² da el voltaje de ruptura a través de una separación de 1 mm de 40 kV, mientras que los electrodos del mismo material con una superficie de 1000 cm² da un voltaje de ruptura a través de la misma separación de 1 mm de 25 kV.

Hasta separación de 1 mm, los electrodos más convexos tienen una mayor voltaje de ruptura que el de electrodos planos aun más a el mismo voltaje ellos tienen un campo eléctrico más elevado en la superficie.

(vi) Temperatura

La variación del voltaje de ruptura con la temperatura es muy pequeña, y para electrodos de níquel y hierro, el campo eléctrico se mantiene sin cambios para temperaturas tan altas como 500 °C. El enfriamiento de los electrodos a la temperatura de nitrógeno líquido aumenta el voltaje de ruptura.

(vii) La frecuencia de la tensión aplicada.

Se sabe que a una separación dada se encuentra un voltaje impulsivo más grande que un voltaje de corriente alterna, y un voltaje de corriente alterna más grande que el de una tensión continua. Sin embargo, se ha demostrado que para una pequeña separación (2 mm) no hay dependencia del voltaje de ruptura con la frecuencia en el rango de 50 Hz a 50 kHz.

(viii) La presión.

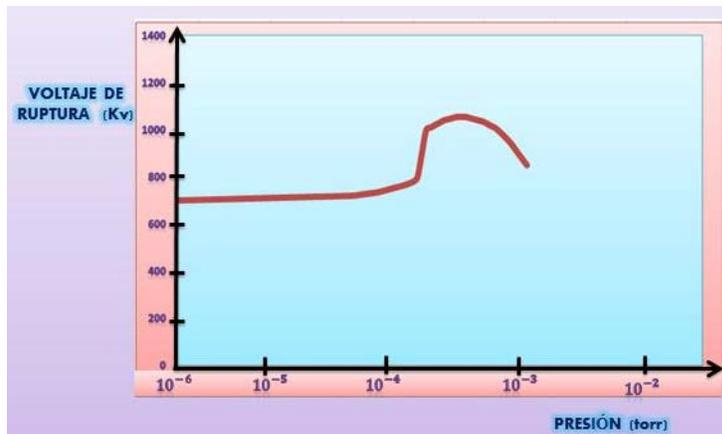


Figura 1.11 – Características del rompimiento.

Para separaciones pequeñas, aumentando el grado de vacío aumenta el voltaje de ruptura hasta por debajo de una cierta presión donde ya no se observa ningún cambio. La región de ruptura es la región en la que el voltaje de ruptura se hace independiente de la naturaleza de la presión de la separación existente entre los electrodos.

Sin embargo, para grandes separaciones (aproximadamente 200 mm de separación) se ha encontrado que por debajo de una cierta presión el voltaje de ruptura comienza a disminuir de nuevo, de modo que el campo eléctrico de ruptura en esta etapa de hecho, podría ser mejorada empeorando el vacío.

Por ejemplo, para una esfera de 1,6 mm de diámetro opuesto a un cátodo plano y separado unos 200 mm, se obtuvieron las características que se muestran en la figura 1,11. El voltaje de ruptura fue constante para presiones menores de 5×10^{-4} Torr, pero aumentó al aumentar la presión hasta un máximo en 5×10^{-4} Torr, y con un aumento adicional de la presión el voltaje se redujo significativamente.

¿Cómo se mezclan los fluidos?

Sergio Cuevas García¹ y Aldo Figueroa Lara²

¹Instituto de Energías Renovables, UNAM, Temixco, Morelos, 62580 México

²Facultad de Ciencias, UAEM, Cuernavaca, Morelos, 62209 México

1. Introducción

El mezclado de fluidos es uno de los procesos físicos más comunes en la Naturaleza y de mayor utilidad en diversas aplicaciones tecnológicas. Se presenta en una multitud de escalas de longitud que abarcan desde escalas muy grandes, como es el caso del mezclado en el interior de las estrellas, o bien a escala planetaria, en la atmósfera o el océano, hasta escalas muy pequeñas, milimétricas o micrométricas, como el mezclado que tiene lugar en los diminutos vasos sanguíneos o en los dispositivos *microfluidicos*. A su vez, los diferentes procesos de mezclado involucran una gran cantidad de escalas temporales que van de los miles o millones de años, por ejemplo en los procesos geológicos, a las fracciones de segundo, como en los procesos de combustión. No obstante esta variedad de escalas, existen mecanismos que son comunes en estos procesos.

Comprender el mezclado de fluidos es en general bastante difícil y requiere, a la par, de herramientas teóricas y experimentales elaboradas. Un ejemplo cotidiano que ilustra en esencia el proceso de mezclado en fluidos podemos encontrarlo cuando vertemos una gota de leche o crema en una taza de café negro [1]. Si agitamos suavemente con una cuachara, podemos observar que en un inicio se forman patrones estriados hasta que eventualmente la crema y el café se homogeneizan formando una mezcla uniforme de color marrón. Lo que está en juego en este proceso son los mecanismos fundamentales del mezclado, es decir, la advección y la difusión. La difusión se presenta aun en ausencia de movimiento del fluido. Por ejemplo, cuando vertemos la gota de leche en el café en reposo, suponiendo que la perturbación introducida por ésta es despreciable, la leche se mezcla debido al mecanismo de difusión molecular. Sabemos por experiencia que en este caso el mezclado tomará mucho más tiempo que cuando utilizamos la cuchara y agitamos el fluido, activando entonces el mecanismo advectivo que lleva a una homogeneización mucho más rápida.

Antes de presentar cómo podemos describirlo y entenderlo, veamos qué es lo que comunmente entendemos por mezclado. Mezclar implica manipular un sistema físico heterogéneo con la intención de volverlo más homogéneo [2]. Normalmente al mezclar introducimos desorden en un sistema inicialmente ordenado. De hecho, el objetivo del mezclado es crear *aleatoriedad* en la distribución de una cierta cantidad física, por ejemplo,

la concentración de una sustancia. En el ejemplo anterior, la gota de crema introducida en el café negro tiene inicialmente una concentración muy alta en una región espacial muy pequeña, pero después de agitar el fluido la crema se distribuye de manera homogénea logrando una mezcla en donde la concentración de esta sustancia es uniformemente baja. El mezclado entonces conlleva el transporte de una cantidad física, en este caso la masa, caracterizada por la concentración. Un rasgo fundamental de los procesos de mezclado es su naturaleza irreversible.

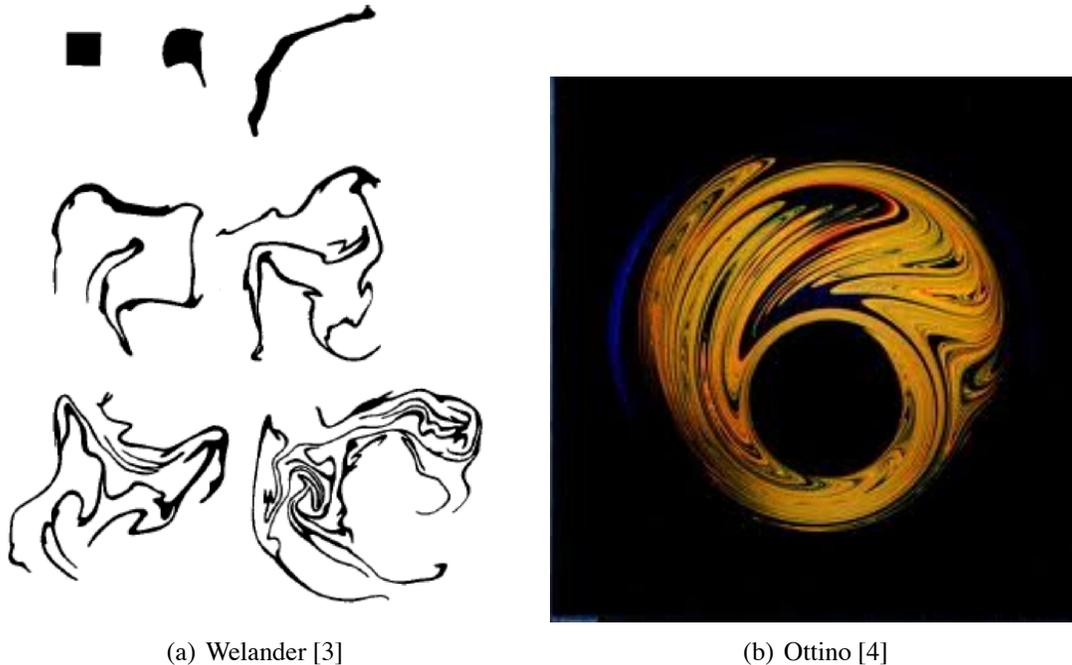
En muchas aplicaciones tecnológicas o cotidianas, como en los sistemas de combustión, la elaboración de pinturas o la preparación de alimentos o bebidas, se busca llevar a cabo el proceso de mezclado de manera *óptima*. Por ejemplo, al mezclar, el objetivo puede ser producir la máxima cantidad de área interfacial entre dos fluidos inicialmente segregados, utilizando para ello el menor tiempo posible o usando la mínima cantidad de energía. El aumentar el área interfacial trae consigo un mayor contacto entre los fluidos y por tanto un mejor transporte. Por otra parte, en ocasiones se desea evitar o controlar el mezclado tanto como sea posible, como ocurre en los derrames accidentales de petróleo en el océano. En cualquier caso, el entendimiento profundo de este fenómeno es de fundamental importancia para poder manipularlo.

2. Mecanismos del mezclado

Las características del mezclado dependen en gran medida de régimen de flujo en el que tiene lugar, es decir, laminar o turbulento. Comúnmente dicho régimen se determina a partir del valor del número de Reynolds, $Re = UL/\nu$, donde U y L representan escalas características de velocidad y longitud y ν es la difusividad viscosa del fluido, también conocida como viscosidad cinemática. El número de Reynolds se interpreta como el cociente de las fuerzas inerciales y las viscosas y toma valores muy grandes ($Re \gg 1$) cuando el flujo es turbulento; en tal caso, dominan las fuerzas inerciales sobre las viscosas. A su vez, los flujos a bajos números de Reynolds ($Re \ll 1$) son laminares. Sin embargo, ésta no es una definición precisa y los límites que determinan la transición entre un flujo laminar y uno turbulento dependen del flujo en particular. En este trabajo consideraremos únicamente el mezclado que tiene lugar en flujos laminares. Aunque es sabido que los flujos turbulentos producen un mezclado eficiente, veremos que bajo ciertas condiciones los flujos laminares pueden llevar también a un buen mezclado.

Veamos ahora con más detalle los mecanismos involucrados en el mezclado laminar. El primero es el mecanismo advectivo o convectivo que se presenta cuando se agita un fluido. La agitación (*stirring*, en inglés) ocasiona que las interfaces entre los materiales se estiren, aumentando el área interfacial entre ellos y creando estructuras estriadas o filamentarias. Así, los filamentos de fluido generados por el movimiento del medio se

estiran y doblan consecutivamente incrementando su longitud con el tiempo. De hecho, la escala de longitud del proceso se determina por el grado de estiramiento de los filamentos. Este fenómeno es lo que en inglés se conoce como *stretching and folding*, que podríamos traducir como *estiramiento y plegado*. Lo que en un inicio era una gota de un fluido en otro, al cabo de un tiempo se convierte en filamentos que se estiran y doblan y cuyos grosores no son todos iguales.



(a) Welander [3]

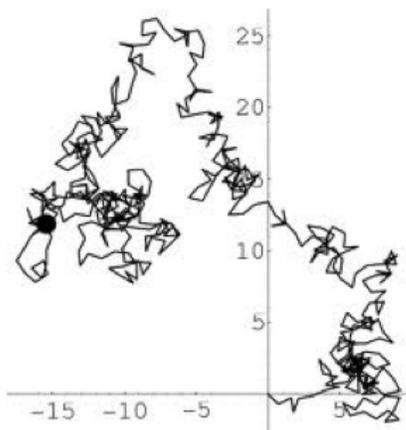
(b) Ottino [4]

Figura 1: (a) Experimento realizado por Welander [3] donde se muestra la evolución de un colorante a partir de una forma cuadrada debido a un movimiento del fluido esencialmente bidimensional. (b) Experimento realizado por Ottino [4] donde se muestran las estructuras estriadas de un colorante en un fluido formadas por el estiramiento y plegado ocasionado por la rotación de dos cilindros excéntricos.

La figura 1(a) muestra uno de los experimentos pioneros, realizado por Welander [3], donde se muestra este proceso. Welander utilizó un tanque de agua que rotaba lentamente de modo que se establecía un movimiento en el fluido cercanamente bidimensional. Mediante un colorante que se difundía muy lentamente, marcó un cuadrado que dejó evolucionar libremente de acuerdo al flujo impuesto. En la figura se muestra el colorante a distintos tiempos, observando que el cuadrado inicial evoluciona en un patrón espacial altamente ramificado, donde se forman filamentos alargados con una estructura muy fi-

na. Es importante resaltar que a pesar de lo intrincado de los patrones, el flujo que los produce es laminar, no turbulento. La figura 1(b) muestra otro experimento famoso realizado por J. Ottino [4] donde se coloca un colorante o marcador dentro de otro fluido contenido entre dos cilindros excéntricos los cuales pueden rotar. El movimiento de los cilindros lleva a la formación de los filamentos que se estiran y doblan, generando zonas donde hay un buen mezclado y otras donde el colorante está ausente, llamadas *islas*. Al crear y estirar los filamentos de fluido, la agitación tiende a incrementar el valor medio de cualquier gradiente inicial [5], por ejemplo, el gradiente de concentración de la sustancia. En ocasiones, si el filamento ha sido suficientemente estirado, las diferencias en la tensión interfacial en los lados opuestos de una interfaz pueden ocasionar que los filamentos se rompan en gotas aisladas, reduciendo la escala de longitud [1]. En líquidos muy viscosos o cuando los materiales son muy similares este rompimiento puede no tener lugar.

El otro aspecto importante en el mezclado es la difusión que es un proceso molecular. Si los fluidos son miscibles, el movimiento Browniano de las moléculas individuales de fluido ocasionado por las fluctuaciones térmicas, actúa con el fin de homogeneizar al fluido a escala molecular [1]. En la figura 2(a) se muestra una simulación de la trayectoria aleatoria que sigue una partícula sujeta a un movimiento Browniano. A su vez, en la figura 2(b) se observa el efecto de la difusión cuando una gota de colorante se introduce en un recipiente con agua en reposo. De hecho, el efecto de la difusión es disminuir el valor medio de cualquier gradiente inicial [5]. Por supuesto, la difusión no tiene lugar si los fluidos son incompatibles.



(a) Trayectoria aleatoria



(b) Difusión de tina en agua

Figura 2: (a) Trayectoria aleatoria seguida por una partícula sujeta a un movimiento similar al que se presenta en el movimiento Browniano ocasionado por la agitación térmica molecular. (b) Difusión de una gota de tina en un recipiente con agua en reposo.

Resumiendo, mientras que la advección es un proceso mecánico, la difusión es un proceso molecular. La advección y la difusión no necesariamente se presentan juntas en una determinada situación. En las etapas iniciales del mezclado, los efectos convectivos (estiramiento y plegado) son los que dominan, llevando a la creación de estructuras filamentosas. El grosor de los filamentos decrece exponencialmente con el tiempo hasta que se alcanza un equilibrio en el cual la difusión empieza a ser significativa, deteniendo el adelgazamiento de los filamentos y propiciando la homogeneización de la concentración. Por su parte, la viscosidad tiende a detener la agitación antes de que se obtenga un mezclado apreciable, a menos de que se contrarreste mediante otros medios, por ejemplo, mediante un forzamiento permanente del flujo. Comúnmente, los campos de velocidad estacionarios llevan a un mezclado pobre, es decir un mezclado localizado solo en una región específica. Por su parte, los campos de velocidad dependientes del tiempo ocasionan un mezclado extensivo, es decir, un buen mezclado.

Una manera de caracterizar los mecanismos físicos involucrados en el proceso de mezclado, es a través de un parámetro adimensional conocido como el número de Péclet, definido como

$$Pe = \frac{UL}{\mathcal{D}},$$

donde U y L son escalas características de velocidad y longitud, respectivamente, y \mathcal{D} es la difusividad de masa del material, cuyas unidades, como todas las difusividades, son m^2/s . Este número también es utilizado en procesos de transferencia de calor, en donde el coeficiente \mathcal{D} se intercambia por la difusividad térmica. El número de Péclet, puede interpretarse como una estimación de la magnitud del transporte advectivo entre el transporte difusivo. Por tanto, cuando tenemos un proceso en donde $Pe \gg 1$, la advección domina sobre la difusión, mientras que cuando $Pe \ll 1$, la difusión es el mecanismo de transporte dominante. Otra interpretación útil del número de Péclet se obtiene a partir de las escalas temporales que caracterizan el proceso de mezclado. De esta forma, Pe se calcula como el cociente del *tiempo difusivo*, caracterizado por una escala de orden L^2/\mathcal{D} , entre el *tiempo inercial* o advectivo, dado por L/U , entonces $Pe = (L^2/\mathcal{D})/(L/U) = UL/\mathcal{D}$. En fluidos, comúnmente el número de Péclet es muy grande por lo que se tienen valores en el rango $Pe \approx 10^6 - 10^{10}$.

Otro parámetro adimensional de utilidad en los procesos de mezclado es el número de Schmidt que está dado como el cociente de los números de Reynolds y Péclet, es decir,

$$Sc = \frac{Re}{Pe} = \frac{\nu}{\mathcal{D}}.$$

Vemos entonces que el número de Schmidt depende sólo de las propiedades del fluido, es decir, del cociente de las difusividades viscosa y de masa. La *longitud de penetración* del

momentum en un proceso difusivo puede estimarse de la forma $\delta_v \approx (vt)^{1/2}$, mientras que la longitud de penetración de la concentración puede estimarse como $\delta_c \approx (\mathcal{D}t)^{1/2}$. Por tanto, la raíz cuadrada del número de Schmidt puede interpretarse como el cociente de la rapidez de propagación del momentum (i.e. la rapidez con la que el movimiento se difunde o muere) entre la rapidez de propagación de la concentración (i.e. la rapidez con la que la concentración se suaviza) [1], es decir,

$$Sc^{1/2} = \frac{d\delta_v/dt}{d\delta_c/dt}.$$

En líquidos comúnmente se tiene que $Sc \gg 1$, lo que implica que las fluctuaciones de concentración sobreviven sin ser borradas por el mezclado mecánico o advectivo hasta tiempos largos en el proceso [1].

3. Enfoques para el análisis del mezclado

Cuando deseamos analizar la homogeneización de tinta o de una impureza en un campo de flujo prescrito, hablamos de mezclado escalar. En tal caso, consideramos lo que se conoce como una cantidad escalar *pasiva*, por ejemplo, la concentración de la tinta, que es transportada por el flujo. El adjetivo *pasiva* denota que la sustancia transportada no altera el campo de flujo original, sino que actúa pasivamente dejándose llevar hacia donde dicte el flujo impuesto. Esencialmente existen dos enfoques para el estudio del mezclado que se asocian a las dos maneras utilizadas para describir el movimiento en la dinámica de fluidos: los enfoques Lagrangiano y Euleriano [6].

Enfoque Lagrangiano

El enfoque Lagrangiano consiste en determinar la trayectoria de las partículas o parcelas de fluido inmersas en el flujo. Este enfoque es muy similar al enfoque cinemático utilizado al describir la trayectoria de las partículas sólidas en la mecánica clásica. Lo que se hace es *etiquetar* o marcar a las partículas pasivas mediante condiciones iniciales dadas y perseguirlas durante su movimiento, impuesto por un campo de velocidad determinado. Como mencionamos antes, una partícula pasiva (o inerte) es aquella que no afecta el campo de velocidad y que simplemente sigue el movimiento del fluido. La suposición básica del enfoque Lagrangiano es que la velocidad de la partícula es igual a la velocidad del fluido [5], es decir,

$$\mathbf{u}_{particula} = \mathbf{u}_{fluido}. \quad (1)$$

De la mecánica clásica, sabemos que la velocidad de la partícula está dada en la forma

$$\mathbf{u}_{particula} = \frac{d\mathbf{x}}{dt} = \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right), \quad (2)$$

mientras que el campo de velocidad, como veremos más adelante, se obtiene de la solución de las ecuaciones de balance o bien del experimento, y puede expresarse en la forma

$$\mathbf{u}_{fluido} = [u(x, y, z, t), v(x, y, z, t), w(x, y, z, t)], \quad (3)$$

donde u , v y w son las componentes del campo de velocidad, en general dependientes de la posición y del tiempo. La condición expresada en la ecuación (1), lleva entonces a las ecuaciones advectivas:

$$\frac{dx}{dt} = u(x, y, z, t), \quad \frac{dy}{dt} = v(x, y, z, t), \quad \frac{dz}{dt} = w(x, y, z, t). \quad (4)$$

Aplicando condiciones iniciales dadas, es posible determinar la trayectoria de las partículas individuales resolviendo las ecuaciones (4). La determinación de las trayectorias de las partículas también se conoce como *seguimiento Lagrangiano* ya que lo que se está haciendo es perseguir a las partículas en su movimiento impuesto por el campo de flujo. Aref [7] hace notar que las ecuaciones (4) reducen el problema advectivo a un sistema dinámico de dimensión finita. Cuando se considera advección turbulenta, el sistema está gobernado por ecuaciones de movimiento estocásticas, mientras que para un flujo laminar el sistema (4) es determinista. Dichas ecuaciones, evidentemente, no consideran el transporte difusivo, es decir, son puramente advectivas.

Enfoque Euleriano

El enfoque Euleriano es una descripción de *campo*, es decir, una descripción basada en la evolución de una cantidad física, en este caso el campo escalar (pasivo) de concentración de una sustancia, c , que toma valores en cada punto del espacio y a cada tiempo. Dada una distribución espacio-temporal de la concentración $c(\mathbf{x}, t)$, el flujo de masa \mathbf{j}_m está dado por una ley fenomenológica conocida como la ley de Fick,

$$\mathbf{j}_m = -\mathcal{D}\nabla c, \quad (5)$$

que establece la proporcionalidad del flujo de masa y el gradiente de concentraciones, siendo la difusividad de masa, \mathcal{D} , el coeficiente de proporcionalidad. Si utilizamos la ley de Fick y la ley de conservación de la masa, en ausencia de movimiento de fluido, obtenemos la ecuación de difusión para la concentración, es decir,

$$\frac{\partial c}{\partial t} = \mathcal{D}\nabla^2 c, \quad (6)$$

donde se ha supuesto que el coeficiente \mathcal{D} permanece constante, lo que es una aproximación común. Si existe movimiento de fluido, entonces la concentración satisface ahora la ecuación de advección-difusión dada por

$$\frac{\partial c}{\partial t} + (\mathbf{u} \cdot \nabla)c = \mathcal{D}\nabla^2 c, \quad (7)$$

donde el segundo término del lado izquierdo denota el transporte advectivo, siendo $\mathbf{u}(\mathbf{x}, t)$ un campo de velocidad conocido. La ecuación (7) puede describir el transporte laminar o turbulento, de acuerdo con la naturaleza del campo de velocidad \mathbf{u} . Comúnmente, al analizar el transporte laminar se introduce un campo de flujo determinista, mientras que para el transporte turbulento el campo de flujo \mathbf{u} se especifica probabilísticamente [7].

Determinación del campo de velocidad

Un requisito indispensable para llevar a cabo la descripción del mezclado escalar, tanto en el enfoque Lagrangiano (basado en determinar las trayectorias de las partículas de fluido) como en el Euleriano (basado en determinar la evolución del campo de concentración), es el conocimiento del campo de velocidad del fluido. Dicho campo puede obtenerse experimental o teóricamente y, en ciertos casos, su determinación podría involucrar dificultades mayores. Si se considera un fluido Newtoniano incompresible, el campo de velocidad puede obtenerse teóricamente resolviendo la ecuación de conservación de la masa o ecuación de continuidad, y la ecuación de balance de la cantidad de movimiento o ecuación de Navier-Stokes, que se expresan, respectivamente, en la forma

$$\nabla \cdot \mathbf{u} = 0, \quad (8)$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} + \frac{1}{\rho} \mathbf{f}, \quad (9)$$

donde ρ es la densidad de masa, p es el campo de presión y \mathbf{f} es una fuerza externa específica por unidad de volumen (fuerza de cuerpo) que actúa sobre el fluido. Con excepción de algunos casos sencillos, comúnmente el problema dinámico que implica encontrar el campo de velocidad y el campo de presión a partir de la solución de las ecuaciones (8) y (9), requiere de un tratamiento numérico.

En la siguiente sección describiremos brevemente una manera no intrusiva de agitar fluidos eléctricamente conductores utilizando fuerzas electromagnéticas. Presentaremos algunos resultados experimentales y soluciones numéricas que permiten reproducir adecuadamente los campos de velocidad requeridos para realizar la descripción del mezclado escalar, utilizando tanto el enfoque Lagrangiano como el Euleriano.

4. Agitación electromagnética

Existe una gran variedad de métodos para agitar un fluido. Los más comunes hacen uso de métodos mecánicos para lograr el movimiento del medio, por ejemplo, utilizando aspas, como en una licuadora, o bien moviendo las paredes del recipiente que lo contiene [8]. También es posible agitar un fluido imponiendo gradientes térmicos de modo que se

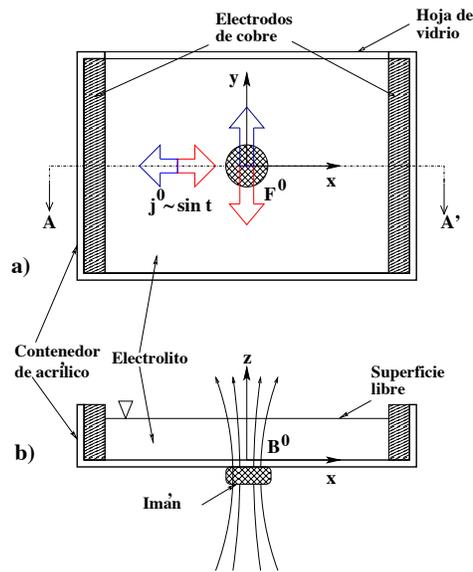


Figura 3: Dispositivo experimental. a) Vista de planta. b) Vista lateral, corte transversal, plano AA'. F^0 denota la dirección principal de la fuerza de Lorentz de acuerdo a la dirección de la corriente aplicada.

produzca convección natural [9]. Una manera conveniente de lograr agitación en fluidos eléctricamente conductores, tales como los electrolitos y los metales líquidos, es utilizando fuerzas electromagnéticas. La idea básica es introducir una corriente eléctrica dentro del medio la cual interacciona con un campo magnético aplicado, dando lugar a una fuerza de cuerpo rotacional, la fuerza de Lorentz, que es capaz de agitar el fluido. Este método tiene la ventaja de ser no intrusivo y fácilmente realizable en laboratorio [10, 11].

Dispositivo y observaciones experimentales

La figura 3 muestra el dispositivo de laboratorio utilizado en el Instituto de Energías Renovables de la UNAM, para llevar a cabo experimentos de agitación electromagnética en electrolitos [11]. Esencialmente, el dispositivo experimental consiste de un contenedor rectangular de acrílico parcialmente lleno con una delgada capa de una solución de bicarbonato de sodio (NaHCO_3) a una concentración de 8.6% en masa. El volumen total del electrolito es 400 cm^3 , teniendo la capa una altura de 4 mm. En dos de las paredes opuestas del contenedor se colocan electrodos de cobre a través de los cuales se aplica una corriente eléctrica alterna al fluido por medio de un generador de funciones. Debajo del contenedor se colocan uno o varios imanes permanentes que generan un campo magnético no homogéneo. La interacción de la corriente alterna y el campo magnético da lugar a

una fuerza de Lorentz periódica que agita el fluido, produciendo un flujo oscilatorio [12]. Si la corriente aplicada es directa, la fuerza de Lorentz no depende del tiempo y da lugar a un flujo estacionario [10]. El movimiento generado se visualiza al introducir trazadores (colorantes de alimentos o fluoresceína) en la superficie del fluido, mientras que la obtención cuantitativa del campo de velocidad de los flujos se lleva a cabo con la técnica de Velocimetría por Imágenes de Partículas (o PIV por sus siglas en Inglés) [11]. La figura 4 muestra el campo vectorial de velocidad a distintas fases de oscilación del flujo producido por la interacción de una corriente alterna y un campo magnético dipolar. Dado que la corriente se aplica en la dirección x y la componente dominante del campo magnético está en dirección z , la fuerza oscila en dirección vertical (y), creando un vórtice dipolar oscilatorio. En esta figura se pueden apreciar los puntos elípticos (marcados con asteriscos) e hiperbólicos (marcados con cruces) presentes en el flujo, además de las trayectorias que siguen estos puntos al evolucionar el flujo; asimismo, se muestra el punto de máxima velocidad (instantánea) dentro del campo de velocidad.

Simulación numérica y comparación con resultados experimentales

La agitación electromagnética en electrolitos puede modelarse mediante la solución de las ecuaciones de continuidad (8) y de Navier-Stokes (9), tomando a la fuerza de Lorentz como término fuente [10]. Como se mencionó previamente, dicha fuerza se crea por la interacción del campo magnético tridimensional $\mathbf{B}^0 = \mathbf{B}^0(x, y, z)$ y la corriente eléctrica aplicada al sistema, que en general es una función del tiempo $\mathbf{j}^0 = \mathbf{j}^0(t)$, es decir $\mathbf{f} = \mathbf{j}^0 \times \mathbf{B}^0$. La figura 5 muestra la comparación de los perfiles de velocidad de la componente en dirección y ($v(x, y, t)$) obtenidos experimentalmente, con los resultados obtenidos de la solución numérica de las ecuaciones de balance. Con el fin de determinar la importancia de los efectos tridimensionales en el flujo, se realizaron simulaciones numéricas utilizando un modelo cuasi-bidimensional (que se justifica dado el pequeño espesor de la capa de fluido), así como un modelo completamente tridimensional [11, 12]. En la figura se puede observar que los perfiles de velocidad $v(x, 0, t)$ son simétricos, mientras que los perfiles $v(0, y, t)$ muestran una marcada asimetría dictada por la dirección principal del flujo originada por la fuerza aplicada. Tanto la predicción numérica cuasi-bidimensional como la tridimensional, presentan un acuerdo cuantitativo muy bueno con los resultados experimentales para diversos arreglos de imanes en un extenso rango de amplitud y frecuencia de corrientes. Esto indica que para las condiciones de flujo, los efectos tridimensionales son poco importantes y el movimiento se aproxima a la cuasi-bidimensionalidad. Podemos afirmar que la solución numérica de las ecuaciones de balance reproduce satisfactoriamente los principales efectos físicos observados experimentalmente.

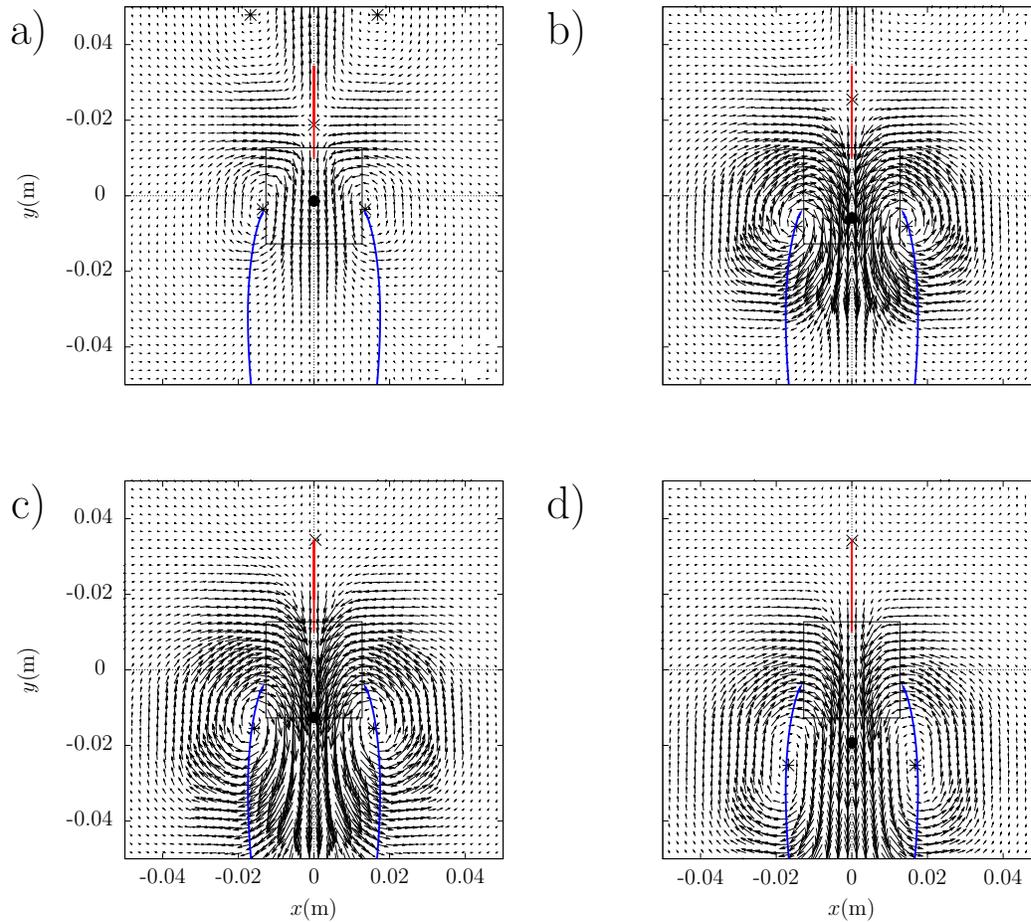


Figura 4: Campo de velocidad experimental para el flujo producido por la interacción de una corriente alterna y en campo magnético dipolar. El punto negro \bullet corresponde al punto de velocidad máxima instantánea. Los asteriscos $*$ y cruces \times denotan los puntos elípticos e hiperbólicos en el flujo, respectivamente, mientras que las líneas azules y rojas denotan sus trayectorias, respectivamente. Las fases de oscilación en las distintas gráficas son: a) $\varphi = 2/10 \pi$, b) $\varphi = 4/10 \pi$, c) $\varphi = 6/10 \pi$ and d) $\varphi = 8/10 \pi$. El cuadrado en la figura denota la localización del imán.

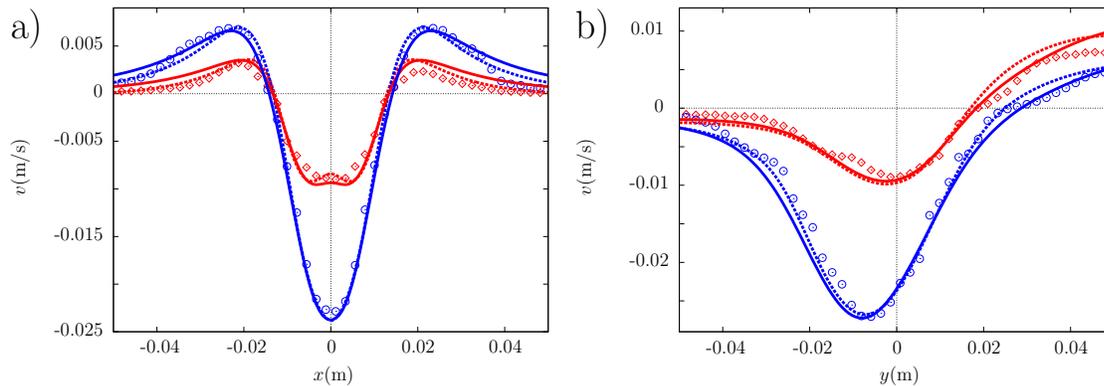
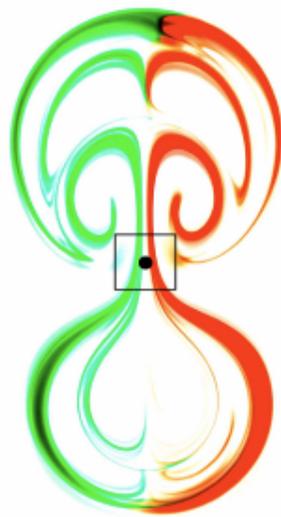


Figura 5: Componente vertical de la velocidad como función de la posición para el flujo producido por la interacción de una corriente alterna y un campo magnético dipolar. Los datos experimentales y numéricos están representados por símbolos y líneas, respectivamente, correspondientes a distintas fases (φ). Diamante rojo \diamond y líneas rojas: $\varphi = 2/10 \pi$; círculo azul \odot y líneas azules: $\varphi = 4/10 \pi$. a) $v(x, 0)$; b) $v(0, y)$. Líneas continuas: modelo cuasi-2D; líneas punteadas: modelo 3D.

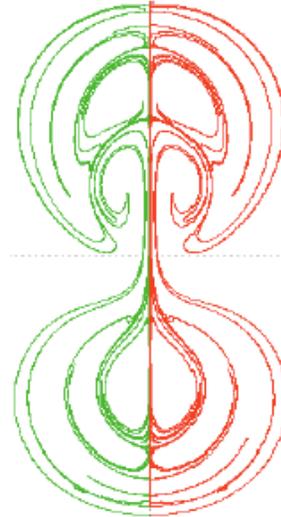
5. Trayectorias Lagrangianas en flujos producidos por agitación electromagnética

Una vez conocido el campo de velocidad, es posible integrar las ecuaciones advectivas (4) para encontrar las trayectorias Lagrangianas. A continuación se muestra la comparación de observaciones experimentales con las trayectorias Lagrangianas calculadas numéricamente para los flujos cuasi-bidimensionales generados mediante fuerzas electromagnéticas oscilatorias, presentados en la sección anterior [11].

En la figura 6 se muestra el caso de un flujo producido por una fuerza electromagnética oscilatoria que apunta en dirección vertical, generada por un campo magnético dipolar. En el experimento (figura 6(a)) se colocaron dos gotas de colorante, una verde y una roja, cerca de la región donde se localiza el imán permanente (marcado con un cuadrado en la figura 6), a cada lado de la línea vertical de simetría y se dejó evolucionar el flujo. Para realizar el seguimiento Lagrangiano numéricamente, como condición inicial se marcó una colección de partículas dispuestas sobre un círculo cuyo centro coincidía con el del imán permanente; en la mitad izquierda del círculo se colocaron partículas verdes y en la mitad derecha partículas rojas. Al evolucionar el flujo, las partículas describen las trayectorias mostradas en la figura 6(b). Con el fin de obtener trayectorias suaves, cuando dos partículas se separan más de una cierta distancia, numéricamente se *inyectan* partículas



(a) Visualización experimental



(b) Simulación numérica: seguimiento Lagrangiano

Figura 6: Flujo oscilatorio producido en una capa delgada de electrolito por una fuerza electromagnética que oscila en dirección vertical. La fuerza es producida por una corriente eléctrica alterna en dirección horizontal y el campo magnético dipolar, esencialmente normal al plano de la imagen, generado por un imán permanente marcado por un cuadrado en la figura 6(a).

adicionales. Puede observarse que la simulación Lagrangiana reproduce adecuadamente las características del flujo experimental, particularmente la simetría respecto al eje vertical que muestra la distribución de los trazadores. Dicha simetría, que impide el transporte de masa entre los lados izquierdo y derecho, es ocasionada por la distribución simétrica de campo magnético producida por el dipolo. La diferencia principal con el experimento se debe a que numéricamente, el seguimiento Lagrangiano no considera la difusión y por tanto no captura el engrosamiento de los filamentos de fluido observados experimentalmente.

La figura 7 muestra el flujo resultante cuando el campo magnético utilizado es cuadrupolar, es decir, producido por dos imanes permanentes cuyos centros se localizan sobre el eje vertical y están separados una distancia $2L$, donde L es la longitud del lado del imán. Al inicio del experimento se colocaron cuatro gotas de colorante de distinto color (verde, amarillo, azul y rojo) en cada cuadrante, mientras que numéricamente se dividió el círculo de partículas pasivas en los cuatro cuadrantes, cada uno con el color correspondiente al experimento. Observamos que la evolución del flujo da lugar a un mezclado por cuadrante.

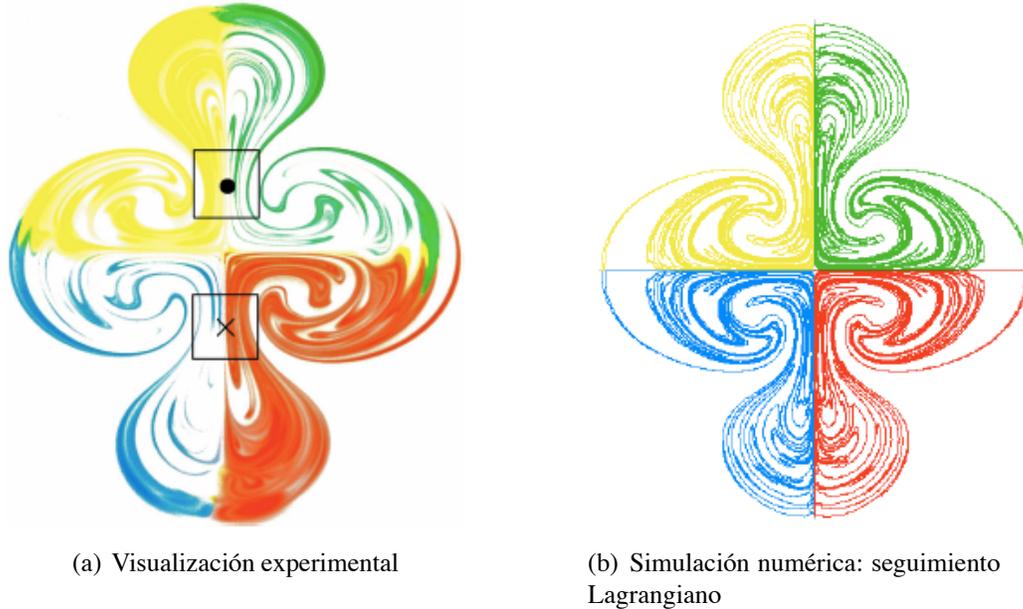


Figura 7: Flujo oscilatorio generado en una capa delgada de electrolito por una fuerza electromagnética que oscila verticalmente. La fuerza es producida por una corriente alterna horizontal y un campo magnético cuadrupolar, generado por dos imanes permanentes con polaridades opuestas (marcados en la figura 7(a)), cuyos centros se localizan sobre el eje vertical.

tes, en donde el transporte de masa entre cuadrantes se inhibe debido a la distribución simétrica de campo magnético. Nuevamente, la simulación Lagrangiana reproduce adecuadamente las características esenciales del flujo. Si se incrementa a cuatro el número de imanes, se produce un flujo octopolar oscilatorio donde se mantienen líneas de simetría que inhiben la homogeneidad del mezclado.

Los ejemplos anteriores muestran claramente que la existencia de simetrías en el flujo inhibe la homogeneización completa de los colorantes, dando lugar a un mezclado localizado. Una manera sencilla de romper las simetrías y propiciar un mejor mezclado es utilizando una distribución irregular de campo magnético, lo que se logra disponiendo un conjunto de imanes de manera desordenada y con polaridad no uniforme [11], como se muestra en la figura 8. En este caso se observa el transporte de masa en todas direcciones de modo que en unos pocos ciclos puede lograrse la homogeneización del colorante, lo que depende también, evidentemente, del número de imanes y de su distribución específica. Se ha mostrado experimentalmente que un arreglo aleatorio de imanes da buenos resultados [13].

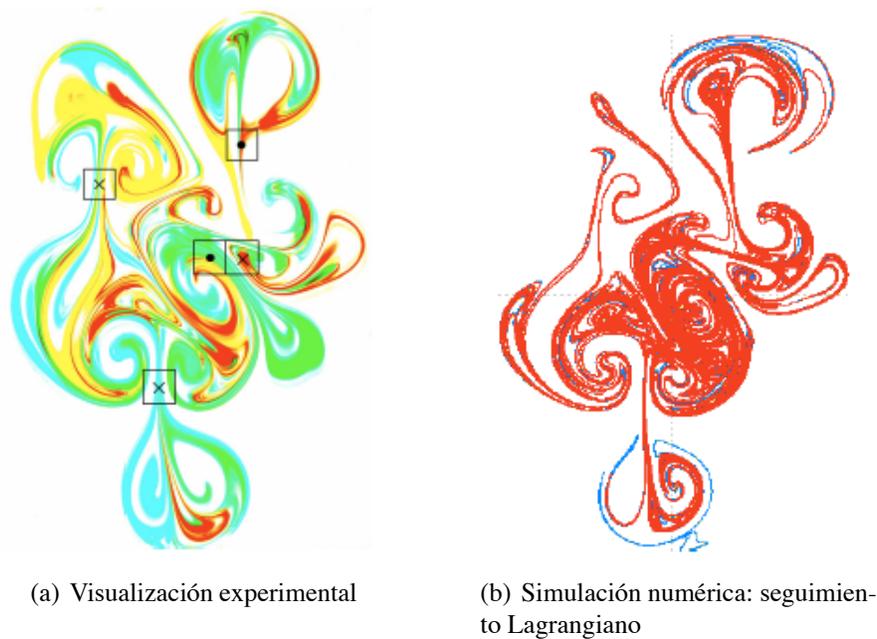


Figura 8: Flujo oscilatorio producido en una capa delgada de electrolito por una fuerza electromagnética que oscila verticalmente. La fuerza es producida por una corriente alterna horizontal y el campo magnético generado por cinco imanes permanentes cuya distribución espacial y polaridad son irregulares. La localización de los imanes está marcada por cuadrados en la figura 8(a)).

Existen otras maneras de romper las simetrías en el flujo y lograr un mezclado más homogéneo sin necesidad de utilizar un flujo turbulento. Esto es importante porque hay situaciones físicas en donde la generación de turbulencia es prácticamente imposible, como ocurre en los dispositivos *microfluidicos*, es decir, dispositivos cuyo tamaño es del orden de los micrómetros, fabricados gracias a las técnicas de la microelectrónica [14]. Estos dispositivos tienen muchas aplicaciones médicas, biológicas y químicas, por ejemplo, en el análisis de compuestos o en la preparación de productos farmacéuticos a partir de una mínima cantidad de materia, en donde se requiere obtener rápidamente una mezcla homogénea de distintas componentes. La microfluidica se caracteriza por tratar siempre con flujos laminares dominados por la viscosidad, de manera que el número de Reynolds es mucho menor que la unidad. La manera de lograr un mezclado eficiente en estas condiciones es propiciando un flujo caótico, o más precisamente, un flujo en donde las partículas de fluido sigan trayectorias caóticas. Curiosamente esto es posible aun en campos de flujo sencillos y aparentemente inofensivos los cuales pueden dar lugar a patrones de advección

altamente complejos. A continuación presentaremos brevemente algunas ideas acerca de este tema.

Advección caótica

En un artículo clásico, Aref [7] demostró que aun campos de velocidad completamente deterministas y que desde el enfoque Euleriano aparecen como flujos simples y regulares, son capaces de dar lugar a una respuesta esencialmente estocástica en las trayectorias Lagrangianas producidas por la advección de un trazador pasivo. Aref denominó a esta situación o régimen *advección caótica*.

En su análisis, Aref consideró campos de velocidad \mathbf{u} deterministas, bidimensionales, incompresibles y dependientes del tiempo. En un flujo incompresible en dos dimensiones, la ecuación de continuidad se expresa en la forma

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \quad (10)$$

lo cual implica que existe una función ψ , conocida como la función de corriente, que permite calcular las componentes de la velocidad como

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x}, \quad (11)$$

de manera que se satisface la ecuación de continuidad. Si ahora escribimos las ecuaciones advectivas (4) en términos de la función ψ utilizando las ecuaciones (11), obtenemos

$$\frac{dx}{dt} = \frac{\partial \psi}{\partial y}, \quad \frac{dy}{dt} = -\frac{\partial \psi}{\partial x}. \quad (12)$$

Nótese que estas ecuaciones tienen la misma estructura que las ecuaciones canónicas de Hamilton para un sistema de un grado de libertad, donde la función de corriente ψ juega el papel del Hamiltoniano. Es decir, la cinemática de la advección en dos dimensiones en un flujo incompresible es equivalente a la dinámica Hamiltoniana de un sistema de un grado de libertad [5]. Este hecho es independiente de si el flujo es viscoso o inviscido (*i.e.* disipativo o no disipativo) ya que la naturaleza Hamiltoniana de la cinemática proviene únicamente de la condición de incompresibilidad. Lo que descubrió Aref [7] fue que aun en un campo de velocidad laminar y aparentemente sencillo, las trayectorias de las partículas pueden ser muy complejas y tener un comportamiento caótico. En otras palabras, vistas como un sistema dinámico, las ecuaciones advectivas pueden llevar a una dinámica caótica. Si se escoge el campo de flujo de tal manera que el sistema dinámico sea no integrable, lo cual en un flujo incompresible bidimensional es posible si el campo es dependiente del tiempo, los movimientos de las partículas individuales pueden producir

trayectorias caóticas. Parafraseando a Aref, podemos decir que el mezclado de un fluido es la representación visual del comportamiento de un sistema caótico Hamiltoniano [5].

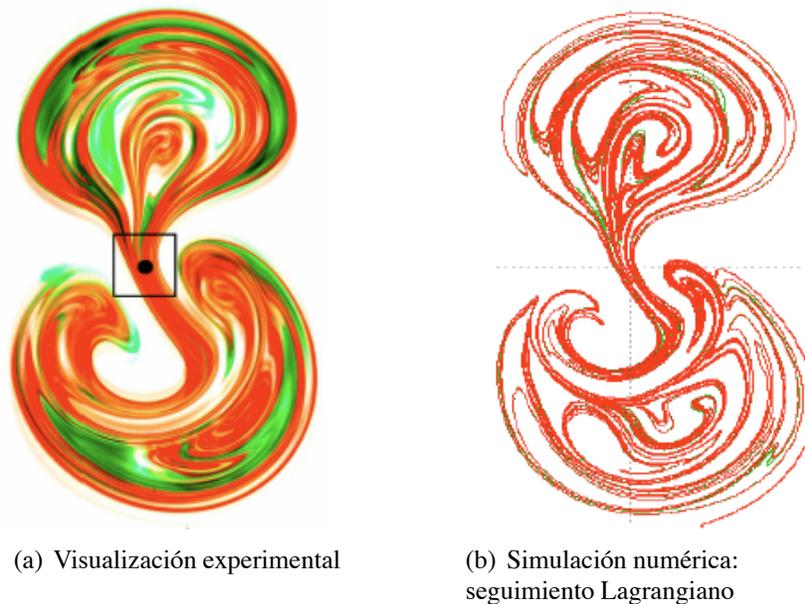
Para mostrar los rasgos esenciales de la advección caótica, Aref [7] estudió un modelo idealizado del movimiento de un fluido en un contenedor producido por un agitador. El fluido se supone inviscido e incompresible y el movimiento es completamente bidimensional. El agitador se modela como dos vórtices puntuales en una posición fija los cuales se “prenden y apagan” de manera periódica y alternada, dando lugar a un flujo dependiente del tiempo. Con dicho modelo, que se conoce como “vórtices parpadenates” (*blinking vortex*), Aref calculó numéricamente las trayectorias de las partículas y demostró su naturaleza caótica, logrando un mezclado eficiente.

Flujo generado por dos fuerzas de Lorentz ortogonales

Veamos ahora otra manera de romper las simetrías en los flujos generados por fuerzas electromagnéticas permitiendo, bajo ciertas condiciones, llegar a un régimen de advección caótica. Partiendo del dispositivo mostrado en la figura 3, la idea es colocar dos electrodos adicionales en las paredes donde antes no existían, evitando un corto circuito con el otro par de electrodos. De esta forma es posible aplicar corrientes alternas en direcciones ortogonales, es decir, vertical y horizontalmente, de modo que si existe un campo magnético aplicado se generarán fuerzas de Lorentz independientes en ambas direcciones.

En la figura 9 se muestra la comparación entre la visualización experimental y el cálculo numérico de las trayectorias Lagrangianas en un flujo producido por la acción de dos fuerzas de Lorentz periódicas y ortogonales que tienen un cociente de frecuencias igual a 2 y fase relativa cero, cuando el campo magnético aplicado es generado por un solo dipolo magnético. Al comparar con el caso en donde existe solo una fuerza en dirección vertical (ver figura 6), observamos que la acción de dos fuerzas ortogonales permite el transporte de masa tanto en dirección vertical como horizontal, lo que da lugar al rompimiento de las simetrías y consecuentemente a un mejor mezclado.

Este sistema tiene propiedades interesantes que pueden explorarse tanto analítica como numéricamente [15]. Primero, si consideramos flujos en el régimen lineal, es decir, donde los efectos convectivos son despreciables, y nos enfocamos en la trayectoria de una sola partícula colocada en el centro geométrico del sistema como condición inicial, puede demostrarse que dependiendo de los parámetros asociados a las corrientes aplicadas en direcciones horizontal y vertical (amplitud, frecuencia y fase), la partícula traza trayectorias que asemejan a las conocidas *figuras de Lissajous*. Estas figuras fueron observadas en 1855 por Jules Antoine Lissajous quien diseñó un sistema óptico simple para estudiar vibraciones compuestas. De esta forma encontró patrones complejos formados cuando se superponen dos vibraciones armónicas a lo largo de líneas perpendiculares. Cuando el cociente de las frecuencias de las dos señales armónicas es racional, se obtienen trayectorias cerradas que dan lugar a movimientos periódicos, mientras que cuando el cociente de



(a) Visualización experimental

(b) Simulación numérica:
seguimiento Lagrangiano

Figura 9: Flujo oscilatorio producido en una capa delgada de electrolito por dos fuerza electromagnéticas oscilatorias en fase que actúan en direcciones ortogonales, con un cociente de frecuencias igual a 2. Las fuerzas son producidas por dos corrientes eléctricas alternas en direcciones vertical y horizontal y el campo magnético generado por un imán marcado por un cuadro en la figura (a).

las frecuencias es irracional, las trayectorias son abiertas, resultando en un movimiento no periódico. En tal caso, al transcurrir el tiempo el sistema visita todos los puntos en un plano limitado por la amplitud de la señal, de manera que la curva llenará gradualmente todo el dominio. De hecho, en 1871, L. Boltzmann (1844-1906) hizo por primera vez la analogía entre las trayectorias físicas en un espacio bidimensional y las figuras de Lissajous, lo que parece ser la primera descripción de Boltzmann de la hipótesis ergódica [16].

Mientras que las figuras de Lissajous se pueden entender desde un punto de vista puramente cinemático y no disipativo, las trayectorias Lagrangianas resultantes de la acción de dos fuerzas ortogonales se originan en un sistema disipativo no lineal, es decir, en un flujo viscoso forzado electromagnéticamente. En tanto que en el régimen lineal las trayectorias Lagrangianas reproducen las propiedades típicas de las figuras de Lissajous relacionadas con la periodicidad y no periodicidad del movimiento para cocientes de frecuencias racionales e irracionales, respectivamente, en el régimen ligeramente no lineal, es decir, cuando los efectos convectivos son pequeños, aún si la razón de frecuencias es

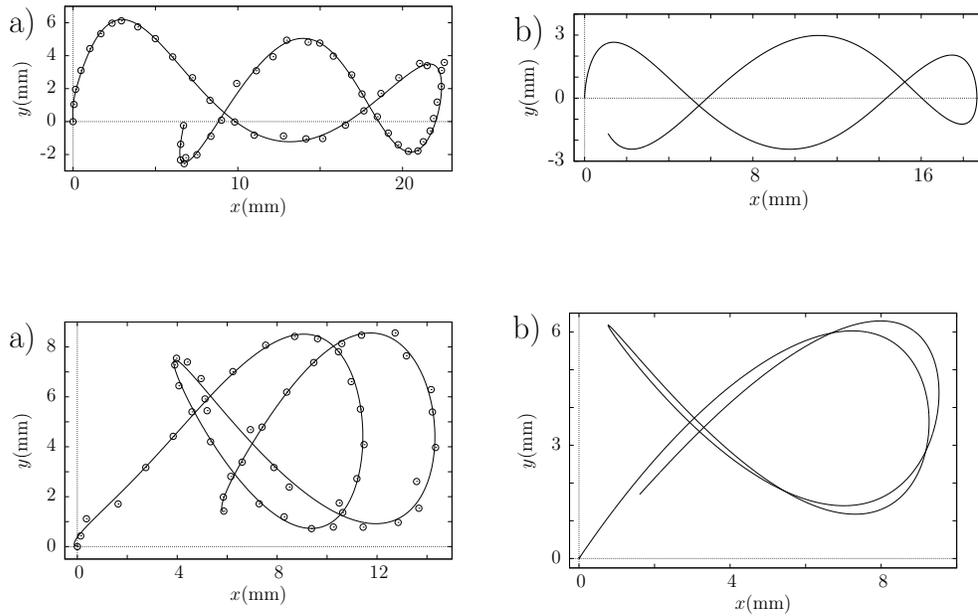


Figura 10: Trayectorias Lagrangianas de una partícula sujeta a dos fuerzas de Lorentz armónicas y ortogonales. Primera columna: observación experimental. Segunda columna: cálculo numérico. Primera fila: razón de frecuencias $\Omega=3$ y fase relativa $\phi = \frac{\pi}{2}$. Segunda fila: razón de frecuencias $\Omega = 3/2$ y fase relativa $\phi = 0$.

racional, la partícula no regresa a su punto inicial. La figura 10 muestra la comparación de las trayectorias Lagrangianas obtenidas experimental y numéricamente de una partícula sujeta a dos fuerzas de Lorentz ortogonales que oscilan armónicamente, en el caso en que los efectos convectivos no son despreciables. Este hecho lleva a trayectorias Lagrangianas no cerradas.

Al incrementarse los efectos no lineales, las trayectorias Lagrangianas son cada vez más complejas y describir su comportamiento se vuelve muy complicado, sobre todo para tiempos largos. Una manera alternativa conveniente para analizar la dinámica del sistema es calcular la sección de Poincaré [17]. Dicha sección es particularmente útil para el análisis de sistemas sujetos a fuerzas periódicas. De manera muy breve, podemos decir que la sección de Poincaré se calcula de la siguiente manera. Al tiempo inicial marcamos la posición de la partícula en el plano $x - y$, la dejamos evolucionar y volvemos a registrar su posición cuando el forzante ha completado un periodo, lo que se continúa haciendo durante un determinado número de periodos. Así, la sección de Poincaré se forma mediante las intersecciones a cada periodo en el plano $x - y$ de las órbitas seguidas

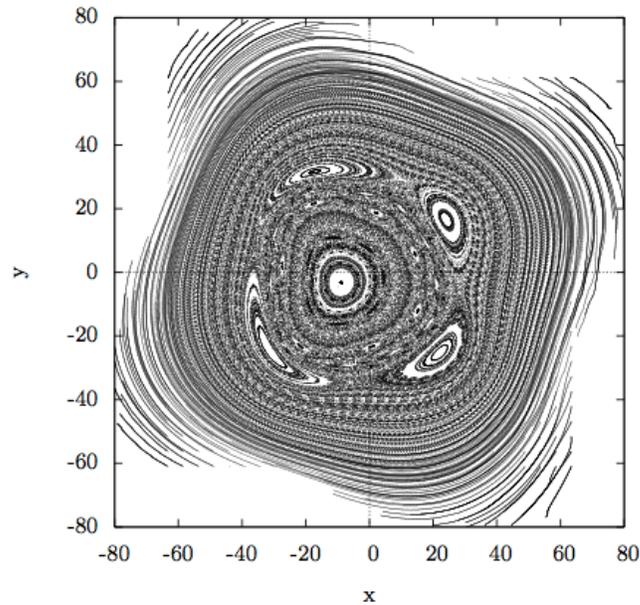


Figura 11: Sección de Poincaré obtenida para 400 periodos a partir de las trayectorias de una partícula dentro de un flujo producido por dos fuerzas de Lorentz ortogonales periódicas. El cociente de las frecuencias de la fuerzas es $\Omega=1$ y la fase relativa $\phi = \frac{\pi}{2}$.

por la partícula. Si la trayectoria de la partícula es cerrada, la sección de Poincaré será un punto, mientras que si no es cerrada, en principio los puntos podrían ir ocupando todo el espacio disponible. La figura 11 muestra la sección de Poincaré correspondiente a una partícula en un flujo generado por dos fuerzas de Lorentz ortogonales periódicas, con un cociente de frecuencias racional igual a la unidad y desfasadas $\pi/2$. Puede observarse que en la sección de Poicaré aparecen *islas*, es decir, regiones no visitadas por la partícula y que son indicativas de zonas dentro del dominio donde no ocurre un buen mezclado. Estas islas aparecen en muchos otros sistemas de mezclado. De hecho, este tipo de sección de Poincaré parece indicar la presencia de un régimen de advección caótica.

6. La ecuación de advección-difusión: el método DSM

Veamos ahora la manera de tratar el mezclado a partir del enfoque Euleriano, es decir, resolviendo la ecuación de advección-difusión:

$$\frac{\partial c}{\partial t} + (\mathbf{u} \cdot \nabla)c = \mathcal{D}\nabla^2 c. \quad (13)$$

Como se mencionó previamente, el objetivo es obtener el campo escalar de concentración de la sustancia como función de la posición y del tiempo, es decir, $c(\mathbf{x}, t)$. Recordemos que el mecanismo advectivo, mediante el estiramiento y plegado, propicia la creación de filamentos de fluido que se alargan al transcurrir el tiempo, disminuyendo su grosor hasta alcanzar una cierta escala en donde la difusión entra en acción. Dicha escala se conoce como la escala de Batchelor y se define como la escala más pequeña que pueden alcanzar las fluctuaciones en la concentración escalar antes de que el fenómeno sea dominado por la difusión molecular [18]. Esta escala es similar a la escala de Kolmogorov que corresponde a la escala más pequeña que pueden alcanzar los vórtices o remolinos en un flujo turbulento antes de que domine la viscosidad. Cuando el número de Schmidt es mayor que la unidad ($Sc > 1$), lo que es común en muchos líquidos, la escala de Batchelor es más pequeña que la escala de Kolmogorov. Esto significa que el transporte escalar tiene lugar a escalas menores que el remolino más pequeño que sobrevive en un flujo turbulento antes de que sea disipado por viscosidad [18].

Si deseamos simular adecuadamente el proceso de advección-difusión, en principio la malla requerida para encontrar la solución numérica de la ecuación (13), dado un campo de velocidad \mathbf{u} , debe ser lo suficientemente fina para resolver al menos la escala de Batchelor. Para ejemplificar la situación, consideramos un problema bidimensional. Si suponemos que el grosor inicial del filamento es s_0 , la escala de Batchelor es del orden de $s = s_0 Pe^{-1/2}$ [19]. Esto quiere decir que si consideramos un dominio cuadrado con $L = 10$ cm de lado y tomamos $s_0 = 1$ cm, suponiendo que $Pe = 10^{10}$ (que es un valor común para flujo de líquidos), encontramos $s \approx 10^{-7}$ cm. Entonces una malla uniforme ($dx = dy$) capaz de resolver la escala de Batchelor debe cumplir

$$dx = \frac{L}{N} \approx s,$$

donde N es el número de nodos. Tomando en cuenta los valores anteriores, encontramos que $N \approx 10^6$, es decir, la malla numérica debe ser de al menos $N \times N \approx 10^{12}$, que es una malla extremadamente fina que considera apenas un dominio de 10 cm^2 . Por tanto, en general, la solución numérica directa de la ecuación (13) es muy costosa computacionalmente ya que requiere mallas espaciales muy finas para resolver adecuadamente los gradientes de concentración.

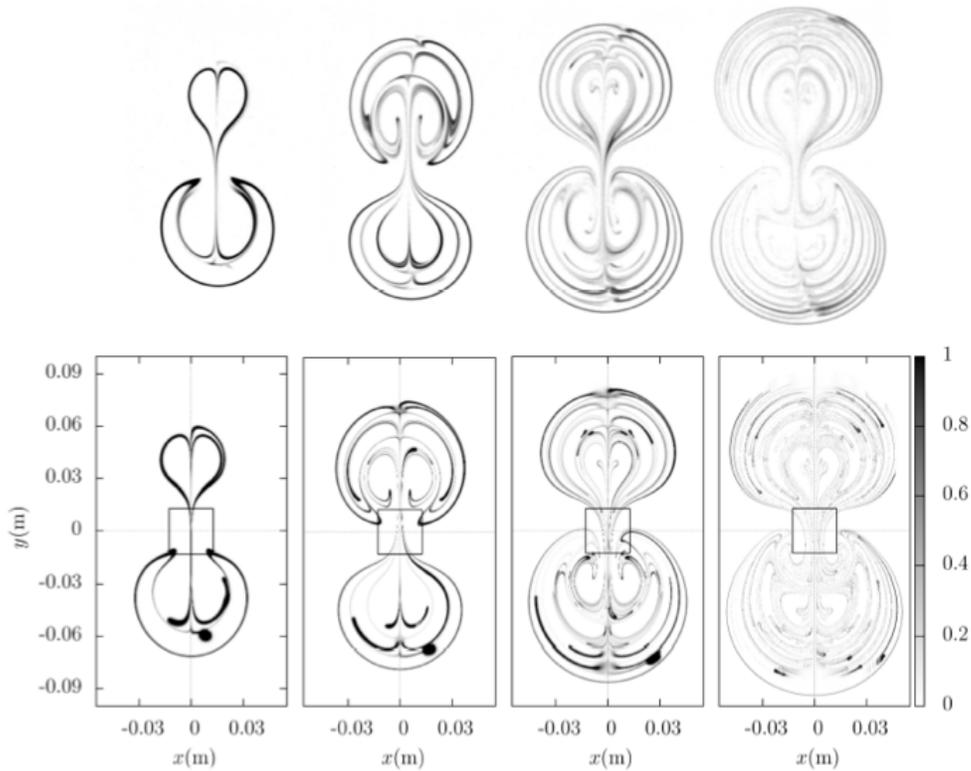


Figura 12: Vórtice dipolar oscilatorio producido por una fuerza electromagnética que actúa en dirección vertical. La primera fila presenta la visualización experimental, mientras que la segunda muestra los resultados de la simulación numérica usando el método DSM. Primera columna, $t=30$ s, segunda columna, $t=60$ s, tercera columna, $t=90$ s, cuarta columna, $t=150$ s. El cuadrado denota la localización del imán permanente.

Una alternativa para determinar la advección-difusión de un escalar en un campo de velocidad incompresible en dos dimensiones para números de Péclet grandes, es utilizar el método propuesto por Meunier & Villermaux [19], conocido como *Diffusive Strip Method* o bien DSM, que combina los enfoques Lagrangiano y Euleriano. Esencialmente, en este método el escalar se introduce como una tira material representada por un arreglo de trazadores pasivos Lagrangianos cuyas posiciones se calculan integrando las ecuaciones advectivas (4). Se elige entonces un nuevo sistema de referencia montado en la i -ésima partícula advectiva de la tira material, y haciendo un cambio de variables apropiado, la ecuación de advección-difusión se transforma en una simple ecuación de difusión en una dimensión, la cual tiene solución analítica para el campo de concentración. Escogiendo

las coordenadas apropiadamente y suponiendo que el grosor de la tira difusiva es menor que su radio local de curvatura, el método DSM resuelve el problema asociado de advección-difusión calculando la tasa local de estiramiento a lo largo de la tira. Así, el campo de concentración se reconstruye sumando pequeñas elipses Gaussianas, dadas por la solución analítica, centradas en cada trazador. La sucesión de elipses forma un filamento que representa la distribución escalar. Una explicación detallada del procedimiento numérico para la reconstrucción del campo escalar de concentración como función del tiempo, puede encontrarse en el artículo de Meunier & Villermaux [19]. La figura 12 muestra la comparación a distintos tiempos de la visualización experimental y la solución numérica del campo de concentración calculada con el método DSM para un flujo agitado electromagnéticamente utilizando un campo magnético dipolar [11, 20]. Podemos observar que cualitativamente la solución numérica reproduce razonablemente bien la evolución de la distribución experimental del trazador escalar. Nótese el engrosamiento de los filamentos que hace patente la tendencia a la homogeneización de la concentración la cual está ausente cuando se considera únicamente el seguimiento Lagrangiano. Debe mencionarse que también se ha efectuado una comparación cuantitativa calculando la función de densidad de probabilidad [19] de los niveles de la concentración c , encontrando un acuerdo satisfactorio con el experimento [20].

7. Comentarios finales

En este escrito hemos presentado los mecanismos físicos básicos que intervienen en el mezclado escalar en régimen laminar, así como los conceptos y enfoques fundamentales utilizados para describirlo. A partir de los flujos generados por fuerzas electromagnéticas periódicas en capas delgadas de electrolito, se ilustraron los dos enfoques característicos del mezclado, a saber, el Lagrangiano y el Euleriano. Primero, usando los campos de velocidad calculados numéricamente con un modelo cuasi-bidimensional validado experimentalmente, se comparó el seguimiento Lagrangiano numérico de distintos flujos de vórtices oscilatorios con la visualización experimental, encontrando un acuerdo razonable dado que este método no considera la difusión. Asimismo, se introdujeron las ideas básicas de la advección caótica y se presentó una manera de obtenerla en flujos agitados electromagnéticamente. Posteriormente, el transporte advectivo-difusivo de un escalar fue ilustrado implementando el método DSM, que resulta conveniente para números de Péclet grandes [19]. Este método numérico establece el vínculo entre los métodos de simulación estandar de la dinámica de fluidos limitados a bajos números de Péclet y los métodos de seguimiento Lagrangiano que no modelan la difusión escalar. Se ha encontrado que el método DSM captura correctamente las principales características físicas del mezclado escalar en flujos agitados electromagnéticamente.

Es importante enfatizar que en el trabajo presentado aquí se ha realizado un esfuerzo para validar los resultados de las simulaciones numéricas con las observaciones experimentales ya que, al igual que en cualquier rama de la física, este es un requisito indispensable para tener certeza de la validez de las herramientas teóricas utilizadas.

Agradecimiento. Este trabajo fue apoyado por CONACYT mediante el proyecto de Ciencia Básica 131399.

Referencias

- [1] Sturman, R., Ottino, J. M. y Wiggins, S. *The mathematical foundations of mixing*, Cambridge, 2006.
- [2] [http://en.wikipedia.org/wiki/Mixing\(processengineering\)](http://en.wikipedia.org/wiki/Mixing(processengineering))
- [3] Welander, P. Studies of the general development of motion in a two-dimensional, ideal fluid, *Tellus*, **7**, 265 (1955).
- [4] Ottino, J.M., The mixing of fluids, *Scientific American*, **260**, 5667 (1988).
- [5] Aref, H., The development of chaotic advection, *Phys. Fluids*, **14**, 1315-1325 (2002).
- [6] Ottino, J. M., *The kinematics of mixing: stretching, chaos, and transport*, Cambridge, 1989.
- [7] Aref, H., Stirring by chaotic advection, *J. Fluid Mech.*, **14**, 1315-1325 (1984).
- [8] Chien, W.L., Rising, H., and Ottino, J. M., Laminar mixing and chaotic mixing in several cavity flows, *J. Fluid Mech.* **170**, 355 (1986).
- [9] de la Cruz, L. M. y Ramos, E. Mixing with time-dependent natural convection, *Int. Comm. Heat Mass Transfer*, **33**, 191-198 (2006).
- [10] Figueroa, A., Demiaux, F., Cuevas, S., Ramos, E., Electrically driven vortices in a weak dipolar magnetic field in a shallow electrolytic layer, *J. Fluid Mech.*, **641**, 245–261 (2009).
- [11] Figueroa, A., Dinámica de vórtices generados por fuerzas electromagnéticas, Tesis de Doctorado, Centro de Investigación en Energía, Universidad Nacional Autónoma de México (2010).
- [12] Figueroa, A., Cuevas, S., Ramos, E., Electromagnetically driven oscillatory shallow layer flow, *Phys. Fluids*, **23**, 013601 (2011).

- [13] Voth, G. A, Saint, T. C., Dobler, G. and Gollubb, J. P. *Phys. Fluids*, **15**, 2560-2566 (2003).
- [14] Guyon, E., Hulin, J. P., Petit, L., *Ce que disent les fluides*, Belin: Pour la Science, 2011.
- [15] Figueroa, A., Cuevas, S., Ramos, E., Lissajous Lagrangian trajectories in electromagnetically driven vortices. En preparación (2013).
- [16] Nolte, D. D., The tangled tale of phase space, *Phys. Today*, **63**(4), 33-44 (2010).
- [17] José, J. V., Saletan, E. J. *Classical dynamics: A contemporary approach*, Cambridge University Press, 1998.
- [18] Paul, E. L., Atiemo-Obeng, V. A., Kresta, S. M. *Handbook of industrial mixing. Science and practice*, Wiley, New York, 2004.
- [19] Meunier, P. and Villiermaux, E. The Diffusive Strip Method for scalar mixing in two-dimensions, *J. Fluid Mech.* **662**, 134-172 (2010).
- [20] Figueroa, A., Meunier, P., Cuevas, S., Villiermaux, E. and Ramos, E., Chaotic advection at large Péclet number: electromagnetically driven experiments, numerical simulations and theoretical predictions. En preparación (2013).

Estimación y cuantificación de la adaptabilidad homeostática en sistemas fisiológicas

Rubén Fossión

Instituto Nacional de Geriátría, Institutos Nacionales de Salud

Centro de Ciencias de la Complejidad (C3), UNAM

Facultad de Ciencias, UNAM

Con el envejecimiento de la población en muchos países en el mundo, y con el aumento de la contaminación del medio ambiente, se puede notar una mayor prevalencia en las llamadas enfermedades complejas, es decir, enfermedades que afectan a muchos órganos y funciones biológicas a la vez y que requieren un enfoque multidisciplinario en su tratamiento. Tales enfermedades complejas parecen no siempre estar relacionadas a daños estructurales, y posiblemente son causadas por una disregulación en las interacciones del sistema complejo adaptativo. Por esta razón, se las llama también enfermedades dinámicas. Nuevas tecnologías permiten monitorear de una manera no-invasiva la fluctuación en el tiempo de muchos observables fisiológicos, la cual en la física se llama “serie de tiempo”, y que refleja cómo la función biológica asociada se adapta a cambios en el medio externo y interno. La serie de tiempo fisiológica mejor estudiada es la serie de intervalos de tiempo entre latidos cardíacos sucesivos, y se ha demostrado que el índice de la Variabilidad del Ritmo Cardíaco predice exitosamente diferentes morbilidades y mortalidades. Recientemente, se ha encontrado un comportamiento estadístico muy parecido en series de tiempo de la mayoría de las funciones biológicas del cuerpo humano, p.ej. en el monitoreo de la acidez del esófago durante 24 horas. Es de interés desarrollar un índice de “complejidad” de fluctuación de series de tiempo fisiológicas en general.

Enfermedades complejas y pérdida de homeostasis

La Medicina moderna occidental ha sido enormemente exitosa en el diagnóstico y tratamiento de enfermedades causadas por daños estructurales, que se pueden visualizar con p.ej. un escán. Sin embargo, según Martínez-Lavín et al. (2008), este paradigma lleva al prejuicio de que un síndrome clínico no existe o pertenece a la psiquiatría cuando a este síndrome (el efecto) no corresponde una alteración anatómica o por lo menos una alteración serológica (la causa). Esto dificulta el entendimiento de padecimientos como la fragilidad relacionada al envejecimiento, la fibromialgia, el

síndrome de la guerra del golfo, la fatiga crónica, entre otras, porque parecen no estar relacionadas con daños estructurales. A los padecimientos mencionados se les llama también *enfermedades complejas* porque afectan a muchos órganos o funciones biológicas a la vez, y en su tratamiento requiere un esfuerzo multidisciplinario de varias disciplinas médicas diferentes. Los mismos autores sugieren redefinir el concepto de la enfermedad como una *disfunción*. En esta aproximación, daños estructurales sin disfunción no corresponden con una enfermedad, mientras que disfunción con o sin daños estructurales sí corresponde con una enfermedad.

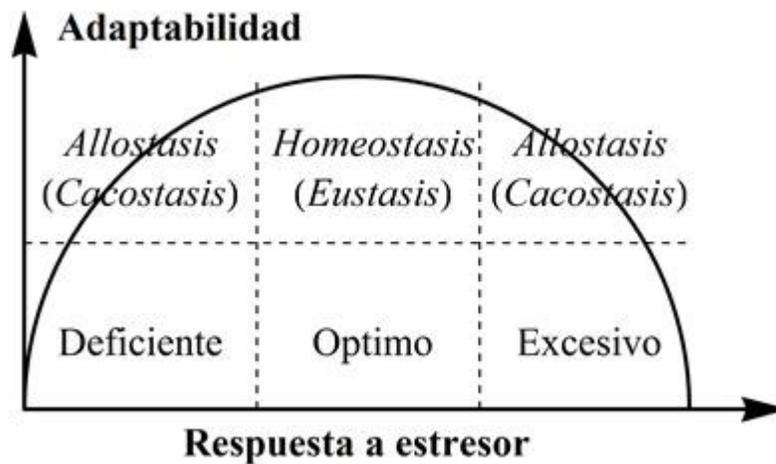


Figura 1. El cuerpo humano se adapta a cambios (estresores) del medio ambiente externo e interno. Para lograr una buena adaptabilidad (homeostasis o eustasis), la fuerza de la respuesta a estresores debe estar en un rango adecuado (justa medianía), entre los dos extremos de una respuesta deficiente o excesiva, que no llevan a una buena adaptabilidad (alostasis o cacostasis).

El fisiólogo Chrousos (2009) logra explicar una gran variedad de enfermedades complejas dentro de un mismo marco teórico como la consecuencia de una mala adaptabilidad del cuerpo humano a condiciones cambiables del medio externo e interno. Tales cambios son “estresores”, que requieren una respuesta adecuada de los sistemas de adaptación para lograr el *homeostasis*. Cuando la respuesta consecuentemente tiende a ser deficiente o excesiva, el homeostasis se degenera hacia un *alostasis* o *cacostasis*, lo cual con el tiempo puede resultar en una enfermedad. Resulta una curva característica con forma de “U” invertida, con el homeostasis como una justa medianía entre dos estados patológicos opuestos (véase la Fig. 1). Buchman (2002) concuerda que la estabilidad dinámica de la salud se obtiene a través de mecanismos de retroalimentación positivos y negativos en redes de órganos y funciones biológicas interconectadas. Fried y colaboradores hipotizan que la fragilidad asociada al envejecimiento resulta por una disregulación de interacciones en el sistema complejo adaptativo;

proponen un modelo matemático que demuestra que la pérdida de capacidad de recuperación del sistema homeostático se puede interpretar como un incremento en el tiempo para regresar a un estado de equilibrio después de una perturbación (Varadhan et al., 2008). Llama la atención que uno de los conceptos más importantes de la Medicina Tradicional China (MTC) es la energía antipatógena (Zheng Qi) que sirve para restablecer y mantener el cuerpo humano en una *justa medianía* (González González, 2012).

Series de tiempo fisiológicas para monitorear la capacidad de adaptación

Uno de los ejemplos más ilustrativos de un sistema adaptable es el sistema de calefacción central automático instalado en casas en países con clima frío: el sistema se enciende cuando la temperatura cae por debajo de cierto umbral preestablecido, luego, la temperatura sube hasta llegar a otro umbral de temperatura más elevado, y el sistema se apaga. Este sistema induce cambios cíclicos en el observable físico de la temperatura de la casa, véase la Fig. 2. En las ciencias físicas, la evolución en el tiempo de un observable específico se llama *serie de tiempo*.

Un descubrimiento de las últimas dos decenias es que el valor de cualquier observable fisiológico fluctúa en el tiempo, como el resultado de la adaptación de su función biológica a cambios en el medio externo o a fluctuaciones en otras funciones biológicas relacionadas (medio interno). Ejemplos son el ritmo cardíaco (véanse las Figs. 3 y 4, y también la Ref. Goldberger et al., 1990), electroencefalografías (Ferree y Hwa, 2003), la secuencia de pasos en la marcha (Hausdorff et al., 1999), el equilibrio (Duarte y Zatsiorsky, 2001), pHmetría del esófago y del estómago durante 24 horas (véase la Fig. 5 y la Ref. Gardner et al., 2005), etc. Fluctuaciones en estos ritmos fisiológicos ocurren a muchas escalas de tiempo (horas a minutos a segundos) con diferentes intensidades que también son autosimilares, así que constituyen un patrón temporal *fractal* (Goldberger et al., 1990). En el caso de la fisiología, el sistema automático que induce las fluctuaciones en todos estos observables fisiológicos es el *sistema nervioso autónomo* con sus dos componentes antagonistas de (i) el *sistema simpático* que responde en situaciones de pelea o de huida (lo que en inglés se puede recordar fácilmente con la mnemotecnica “fight or flight”), y (ii) el *sistema parasimpático* cuyo función es la relajación y la digestión (“rest and digest”). Cabe mencionar que el efecto de estos dos sistemas no es el mismo en todos los órganos: donde el sistema simpático acelera el ritmo cardíaco y decelera la digestión, el sistema parasimpático estimula el resultado opuesto. La competencia entre los sistemas simpático y parasimpático se puede

comparar con el juego de la cuerda o juego de “estira y afloja”, y en estado de salud hay un equilibrio dinámico entre los dos sistemas. Del otro lado, en el caso patológico tiende a predominar uno de los dos sistemas, lo cual lleva a alostasis, y se habla de *disautonomía* (Martínez Lavín y Vargas, 2009). Por la razón de que adaptabilidad significa dinámica y stasis significa muerte, algunos autores han optado por rebautizar el homeostasis como *homeodinámica* o *homeocinética* (Lloyd et al., 2001 y Yates, 2008).

El estudio de series de tiempo fisiológicas permite desarrollar *biomarcadores no-sintomáticos* (Fossion, 2010a). La filosofía es la siguiente: cuando un observable fisiológico rebase los límites de su buen funcionamiento (p.ej. 50 lpm o 100 lpm en el caso del ritmo cardíaco), el paciente muestra síntomas (bradicardia o taquicardia, respectivamente) y su salud está en un peligro inmediato (se puede desmayar o padecer un ataque cardíaco, respectivamente). Del otro lado, cambios en la variabilidad o en la complejidad del patrón de fluctuaciones de una serie de tiempo pueden alertar sobre el estado de salud del órgano que genera la serie, antes todavía de la aparición de síntomas.

Otra ventaja del estudio de series de tiempo es que aporta nuevos datos fisiológicos de índole dinámico, anteriormente inaccesibles, para apoyar al diagnóstico médico de enfermedades complejas. En el caso de pacientes con fibromialgia, el ritmo cardíaco promedio no difiere mucho del ritmo de personas sanas, pero se ha encontrado una predominancia crónica del sistema simpático durante las horas nocturnas que impide el descanso (Martínez-Lavín et al., 1995). Por la presunta causa dinámica de muchas de las enfermedades complejas, a veces se las llama también *enfermedades dinámicas* (Mackey y Milton, 1987).

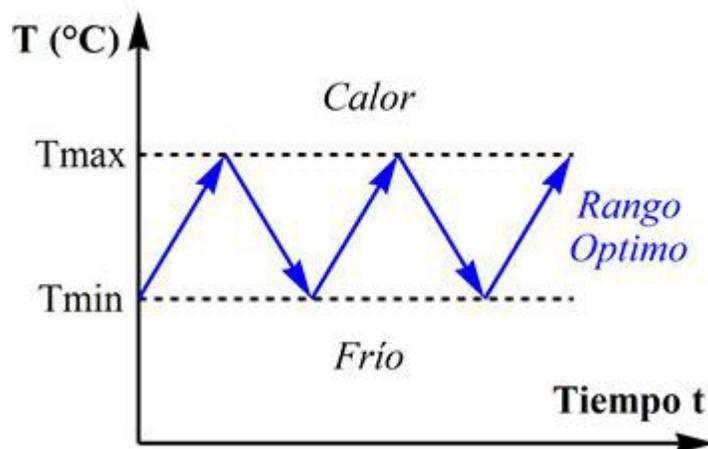


Figura 2. El sistema de calefacción central automático como ejemplo de un sistema adaptable. La temperatura oscila en un rango óptimo entre los límites de temperatura preestablecidos T_{min} y T_{max} .

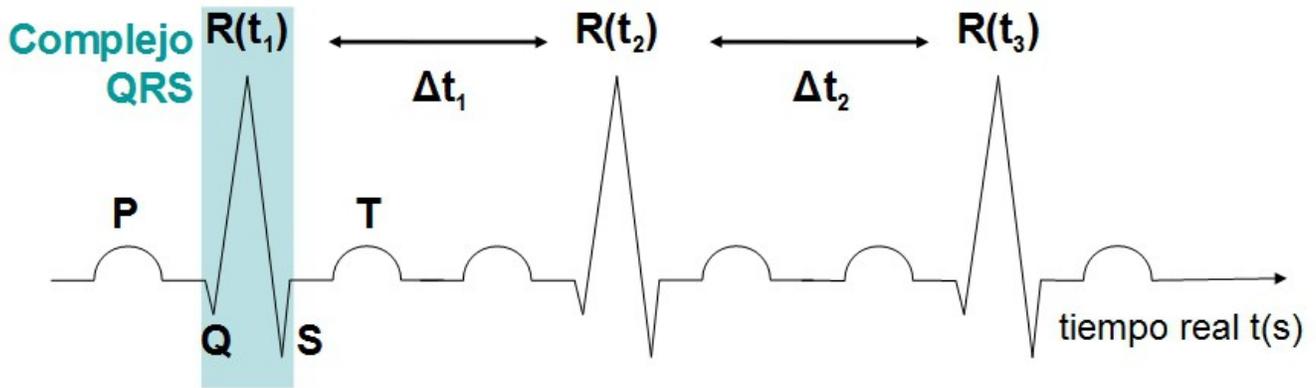


Figura 3. Representación esquemática de tres latidos sucesivos del corazón, como resultan cuando se monitorea la actividad eléctrica del corazón con un electrocardiograma (ECG) o un holter. El complejo PQRST constituye un latido cardíaco, y corresponde con la secuencia particular del llenado y vaciado de las aurículas y ventrículos con sangre. Para estudiar fluctuaciones en el ritmo cardíaco, se considera la serie de intervalos Δt_{RR} entre los picos R sucesivos, véase la Fig. 4.

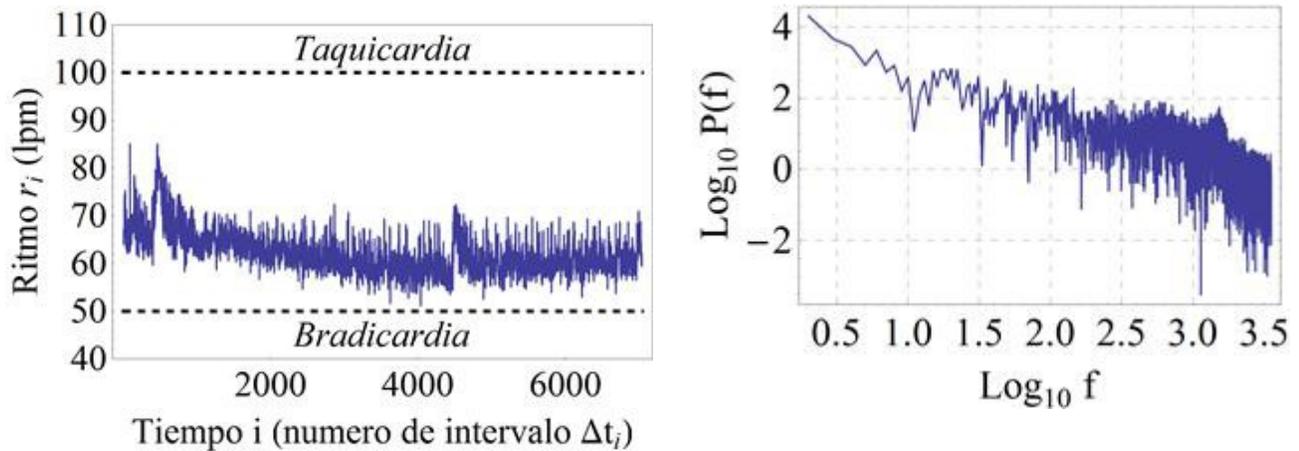


Figura 4. (a) Ritmo cardíaco instantáneo $r_i = 1/\Delta t_i$ en unidades de número de latidos por minuto (lpm). El ritmo instantáneo es el recíproco del intervalo Δt entre dos latidos sucesivos. El monitoreo actual es de un sujeto en reposo, durante dos horas. El ritmo fluctúa continuamente, con un promedio entre los límites de bradicardia (50lpm) y taquicardia (100lpm). (b) Espectro de potencias que muestra las frecuencias presentes en la serie de tiempo (en unidades de número de oscilaciones durante la duración de la serie de tiempo). Se registran pocas periodicidades (pocas frecuencias dominantes), y más bien se muestra una coexistencia de muchas escalas temporales diferentes según una ley de potencias $P(f) \sim 1/f$

^{1.80}. Base de datos “Fantasia” de Physionet (Goldberger et al., 2000).

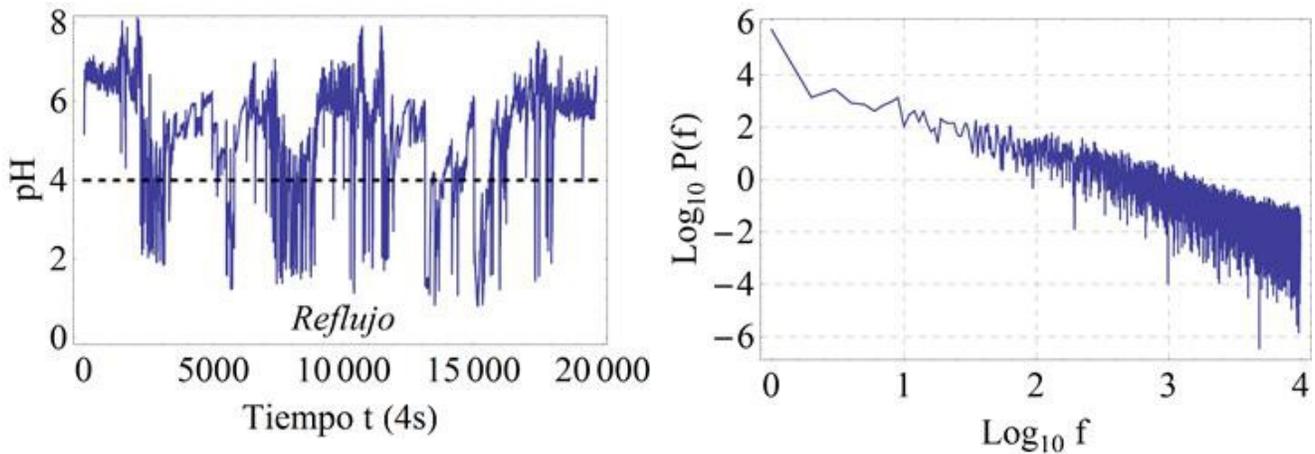


Figura 5. (a) Monitoreo de la acidez del esófago durante 24 horas de un paciente con reflujo en escala logarítmica (pH), donde pH = 7 significa neutralidad y pH = 4 se interpreta como la acidez máxima permitida para el esófago. (b) El espectro de potencias de la serie de tiempo indica una ausencia de periodicidades y una coexistencia de muchas escalas temporales de segundos a 24 horas según una ley de potencias $P(f) \sim 1/f^{1.65}$. Los datos son una cortesía del Dr. Valdovinos del Departamento de Gastroenterología del Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubiran

Variabilidad del ritmo cardíaco

La serie de tiempo fisiológica mejor estudiada es la del ritmo cardíaco, por la razón de que el ritmo cardíaco es un observable relativamente fácil y no-invasivo de monitorear (p.ej. con ECG o holter) y también por el importante papel que juega el corazón en la salud. El desarrollo de índices de la *Variabilidad del Ritmo Cardíaco* (HRV, por sus siglas en inglés) es un campo de *estudio multidisciplinario* que incluye ingenieros (desarrollo de nuevos dispositivos de monitoreo), físicos y matemáticos (análisis estadístico de los datos generados) y fisiólogos y cardiólogos (interpretación clínica de los resultados estadísticos obtenidos) (véase p.ej. Task Force, 1996). Se han desarrollado decenas de índices de variabilidad diferentes, en los dominios del tiempo, de la frecuencia, de la complejidad y de la fractalidad (Ahmad et al., 2009). El índice más utilizado es la desviación estándar,

$$\sigma = \frac{1}{N} \sum_{i=1}^N \sqrt{\Delta t_i - \langle \Delta t \rangle} ,$$

aquí aplicada a una serie de N intervalos Δt_i entre latidos sucesivos del corazón, y que cuantifica cuánto en promedio los intervalos tienden a alejarse del intervalo promedio $\langle \Delta t \rangle$.

A pesar de su sencillez, la desviación estándar σ mide dos fenómenos muy diferentes: (i) el tipo de dinámica de la serie de tiempo, p.ej. fractal, aleatorio, caótico o periódico (compáranse los paneles (a), (b), (c) y (d) de la Fig. 6), y (ii) el rango dinámico en el cual fluctúa la serie de tiempo (compáranse los paneles (d) y (e) de la Fig. 6). Se ha reportado que cambios en la dinámica pueden indicar una variedad de enfermedades en adultos, mientras que con condiciones adversas en neonatos se observa más bien un cambio en el rango dinámico de fluctuación (Nelson et al., 1998).

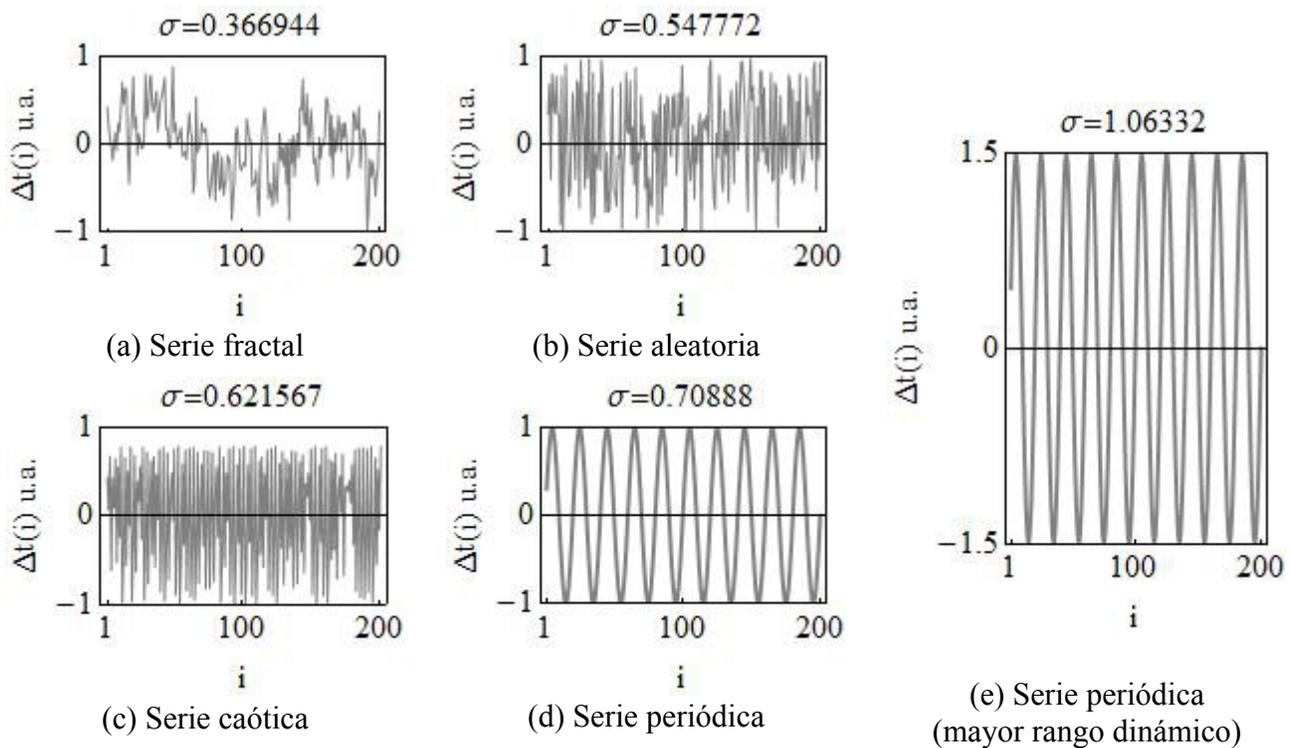


Figura 6. Series de intervalos artificiales que simulan dinámicas diferentes, (a) fractal, (b) aleatoria, (c) caótica, (d&e) periódica. Se muestran intervalos corregidos por su promedio, $\Delta t_i - \langle \Delta t \rangle$, que fluctúan alrededor de 0. Cuando todas las series fluctúan en el mismo rango dinámico, la variabilidad (cuantificada por la desviación estándar σ) se maximiza para la serie periódica, compárase el panel (d) con los paneles (a), (b) y (c). Del otro lado, independientemente del tipo de dinámica, un rango dinámico mayor corresponde con más variabilidad, compáranse los paneles (d) y (e).

Hay que distinguir entre la *variabilidad* y la *complejidad* de una serie de tiempo. Una serie periódica es la más variable pero su modo de fluctuar no es muy compleja; del otro lado una serie fractal con un patrón muy complejo de fluctuar es menos variable (compáranse los paneles (a) y (d) de la Fig. 6, y

véase la Ref. Lipsitz, 2002). En general, los índices HRV desarrollados en la literatura (Task force, 1996; Ahmad et al., 2009) pueden no ser los más adecuados para cuantificar la calidad del homeostasis o la capacidad de adaptabilidad del corazón o del cuerpo humano.

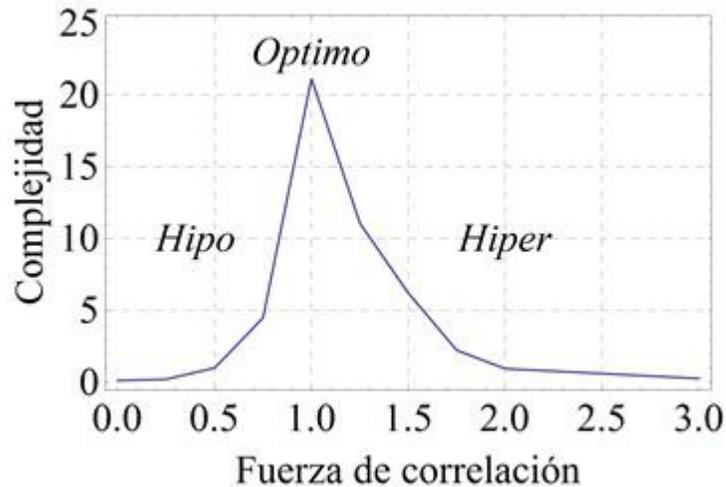


Figura 7. Relación entre la fuerza de correlación de una serie de tiempo y su complejidad. En este caso, se cuantifica la complejidad con el alcance de correlación (memoria) de la serie de tiempo. La complejidad se maximiza para una fuerza de correlación moderada, en el rango óptimo (justa medianía) entre los extremos de series aleatorias (no correlacionadas) y series rígidas (muy correlacionadas). Pequeñas variaciones en la fuerza de correlación causan pérdidas grandes de complejidad. Figura adaptada de Fossion (2011).

Series de tiempo, complejidad y “justa medianía”

Goldberger (1996) sugiere que cualquier serie de tiempo fisiológica procedente de un órgano sano (parecida a la serie del panel (a) de la Fig. 6) está en una justa medianía entre los extremos de una serie aleatoria o no-correlacionada (panel (b)) y una serie rígida/periódica que está infinitamente correlacionada (panel (d)). En un contexto cardiológico, una serie aleatoria correspondería con fibrilación, y una serie rígida con envejecimiento (Goldberger et al., 2002). Se ha propuesto que con las condiciones adversas de enfermedad y envejecimiento, el cuerpo humano pierde complejidad, tanto en sus estructuras anatómicas como en las series de tiempo de muchos observables fisiológicos (Lipsitz y Goldberger, 1992). Una simulación numérica confirma la hipótesis de Goldberger que la complejidad

de series de tiempo se maximiza en la justa medianía entre los dos extremos patológicos de series aleatorios y series rígidas, véase la Fig. 7. Recientemente, se hizo un estudio de la dinámica cardíaca en un grupo de adultos jóvenes. En la mitad del grupo, se encontraron pequeñas desviaciones hacia la rigidez, y en la otra mitad pequeñas desviaciones hacia lo aleatorio. Después de un entrenamiento físico de ejercicio de resistencia por 6 semanas, las series de ambos grupos convergieron hacia la justa medianía donde se maximiza la complejidad (véase la Fig. 8). El ejercicio físico se puede considerar como un tratamiento holístico porque beneficia a la totalidad del cuerpo humano, y porque se puede aplicar con el mismo efecto a dos estados patológicos opuestos.

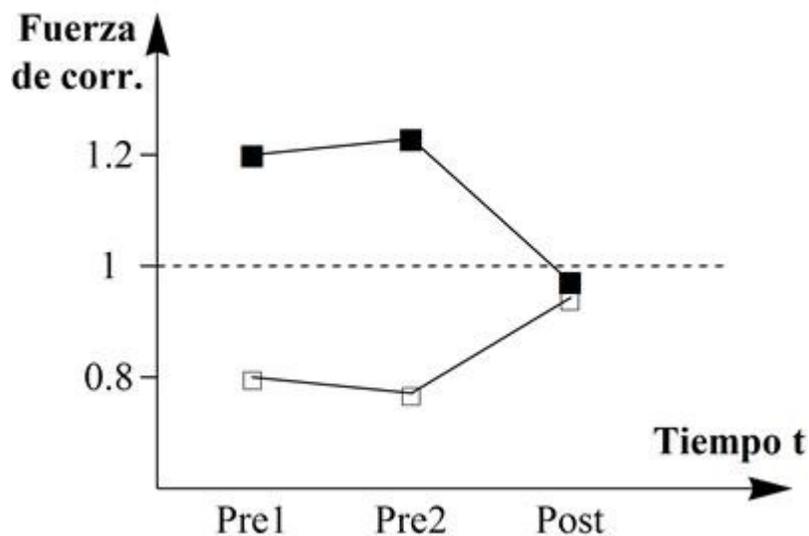


Figura 8. Evaluación de un tratamiento holístico con ejercicio físico de resistencia en adultos jóvenes sanos con la complejidad de series de tiempo cardíacas. En la medición basal (Pre1), las series de la mitad del grupo tienen ligeras desviaciones hacia la rigidez (símbolos negros), mientras que en la otra mitad del grupo hay ligeras desviaciones hacia lo aleatorio (símbolos blancos). No hubo cambios grandes en una segunda medición basal 4 semanas después (Pre2). Después de 6 semanas de ejercicio físico (Post), ambos grupos convergieron hacia la justa medianía con complejidad maximizada. Figura adaptada de Heffernan et al. (2008).

Señales de alerta temprana universales en sistemas complejos dinámicos

A primera vista, puede parecer extraño de interpretar procesos biológicos diferentes dentro de un mismo marco de pérdida de complejidad con condiciones adversas. Existen modelos matemáticos

especializados para funciones biológicas particulares, como p.ej. la marcha (Srinivasan y Ruina, 2006), el equilibrio (Loram y Lakie, 2002) y la dinámica cardíaca (Gois y Savi, 2009). Tales modelos específicos se necesitan para reproducir las particularidades de una función biológica, como p.ej. la forma exacta del complejo PQRST cardíaco, lo cual da información sobre el funcionamiento mecánico del corazón, es decir, el llenado y vaciado de las aurículas y ventriculas con sangre (véase la Fig. 3). Del otro lado, la secuencia de intervalos Δt_{RR} entre los complejos PQRST sucesivos y la manera en la cual esta serie de tiempo fluctúa (véase la Fig. 4), parece reflejar algo más general. Si se comparan los espectros de potencia de dos funciones biológicas tan diferentes como el ritmo cardíaco (Fig. 4b) y la pHmetría 24h del esófago (Fig. 5b), se puede apreciar que son muy parecidos. Todo parece indicar que complejidad de una serie de tiempo fisiológica cuantifica cómo contribuye esta función biológica al sistema de adaptabilidad del cuerpo humano.

La idea de que procesos muy diferentes bajo ciertas condiciones se pueden comportar de una manera similar, no es extraña en la física. Muchos sistemas dinámicos de diversos campos del conocimiento son sistemas compuestos, es decir, consisten de muchos componentes que interactúan entre si. Ejemplos son el clima global al cual contribuyen corrientes en la atmósfera y en el mar, la bolsa de valores donde compiten entre si muchas compañías nacionales e internacionales, la fisiología con muchos órganos o funciones biológicas interactuantes, etc. Muchas veces, el comportamiento de tales sistemas difícilmente se deja describir, modelar o predecir matemáticamente. Tales sistemas se llaman *sistemas complejos*. Sistemas complejos se pueden comportar de una manera muy diferente a los *sistemas lineales*, a los cuales estamos habituados, que responden a perturbaciones con una intensidad proporcional a la fuerza de perturbación. Recientemente se ha encontrado que muchos sistemas dinámicos complejos tienen un *comportamiento no-lineal*, que se caracteriza por *puntos críticos* donde el sistema cambia bruscamente de régimen. Para continuar con los ejemplos anteriormente mencionados, en el caso del clima global, hay preocupaciones de que podamos causar un cambio climático brusco si continuamos añadiendo gases de invernadero a la atmósfera; en el caso de la bolsa de valores, es de interés entender por qué y cuando suceden los cracs; y en la fisiología sería de gran utilidad saber cuáles son las condiciones que desencadenan ataques asmáticos o epilépticos.

Visto que los sistemas complejos difícilmente se dejan modelar matemáticamente, una aproximación alternativa es “escuchar” al sistema, estudiando estadísticamente series de tiempo de un observable asociado al sistema (Fossion et al., 2010b). En una serie de artículos recientes, se ha propuesto que cualquier sistema dinámico que se acerca a un punto crítico y está por cambiar de régimen, genera

señales de alerta temprana universales, sin importar el campo de conocimiento particular al cual pertenece el sistema, p.ej. la física, la fisiología, el clima, el mundo financiero, etc. (Scheffer et al., 2001; *ibid.*, 2009). Algunas de estas señales generales son las siguientes: (i) un incremento en la fuerza de la correlación y en su alcance (memoria), (ii) incremento en el sesgo y asimetría de la distribución o el histograma, (iii) incremento en la desviación estándar, y (iv) deceleración de regreso al estado de equilibrio después de una perturbación. Llama la atención que cuando Fried y colaboradores discuten las causas de la fragilidad asociada al envejecimiento, también mencionan al incremento en el tiempo necesitado por un sistema para recuperarse de perturbaciones (Varadhan et al., 2008).

Hay un tipo de sistemas complejos particulares, que se auto-organizan en un punto crítico entre regimenes extremos opuestos, lo cual se llama *criticalidad auto-organizada* (Bak, 1996). Hay cierta evidencia de que este concepto se puede aplicar a sistemas vivos. Como ejemplo particular, se puede mencionar a un virus RNA dentro de un huesped, que para él constituye un ambiente hostil. Para sobrevivir, el virus tiene que mutarse más rápidamente que el sistema inmunológico del huésped lo puede reconocer. Del otro lado, para tasas de mutación muy altas, el virus no puede transferir adecuadamente su información genética. Así que es muy probable que los virus RNA auto-ajustan su tasa de mutación en una justa medianía entre tasas de mutación demasiada bajas y demasiada altas, para optimizar su sobrevivencia. En esta justa medianía, las series de tiempo para observables asociados al virus RNA obedecen las estadísticas propuestas en el marco teórico de las señales de alerta temprana. Sin embargo, en este contexto de criticalidad auto-organizada, las señales de alerta temprana no avisan por un próximo colapso del sistema, pero al contrario parecen cuantificar la optimización del sistema dinámico (Fossion et al., 2012). Otro ejemplo de criticalidad auto-organizada, puede ser el sistema inmunológico, que auto-ajusta su fuerza de respuesta a patógenos entre los extremos opuestos de respuestas deficientes (tolerancia a antígenos propios y foráneos) y respuestas excesivas (autoinmunidad) (León et al., 2000; 2003).

Conclusión

En la Medicina se sabe muy bien que los valores promedios de ciertos observables fisiológicos no deben de rebasar ciertos límites. Por ejemplo, el ritmo cardíaco promedio en reposo debería de quedarse entre los límites de 50lpm (bradicardia) y 100lpm (taquicardia), y la pH del esófago no debería de rebasar el límite inferior de pH=4 porque de lo contrario el paciente padece de reflujo.

Nuevos desarrollos tecnológicos permiten monitorear la evolución en el tiempo de muchos observables fisiológicos importantes. En la física, la dinámica en el tiempo de un observable se llama una serie de tiempo, y se necesitan nuevas técnicas estadísticas para interpretar este nuevo tipo de datos. En la cardiología, se han realizado grandes avances en la interpretación de la variabilidad del ritmo cardíaco. Sin embargo, la variabilidad de una serie de tiempo parece no ser sinónimo a la complejidad. Parecería que es la complejidad cuantifica cómo contribuye una función biológica a la capacidad de adaptabilidad del cuerpo humano. Hay indicaciones de que la fisiología se puede interpretar como un sistema dinámico adaptable que se auto-organiza en una justa medianía (homeostasis) entre dos extremos patológicos opuestos de deficiencia y exceso. Si esta hipótesis de auto-organización resulta ser cierta, se puede aplicar al estudio del homeostasis y de la capacidad de adaptabilidad del cuerpo humano todo un marco teórico de señales de alerta temprana universales, que ya se está aplicando con éxito a muchos sistemas dinámicos de campos del conocimiento muy diversos.

Referencias

S. Ahmad, T. Ramsey, L. Huebsch et al., 2009. Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. *PloS ONE*, 4 (8), e6642.

P. Bak, 1996. *How Nature works: The science of self-organized criticality*. Springer-Verlag, New York.

T. G. Buchman, 2002. The community of the self. *Nature*, 420, pp. 246-251.

G. P. Chrousos, 2009. Stress and disorders of the stress system. *Nature Rev. Endocrinol.* 5, pp. 374-381.

M. Duarte, V. M. Zatsiorsky, 2001. Long-range correlations in human standing. *Phys. Lett. A* , 283, pp. 124–128.

T. C. Ferree, R. C. Hwa, 2003. Power-law scaling in human EEG: relation to Fourier power spectrum. *Neurocomputing* 52–54, pp. 755–761.

R. Fossion, 2010a. Una definición “compleja” de la fragilidad: Caos, fractales y complejidad en series de tiempo biológicas. Capítulo XVII en “Envejecimiento humano: Una visión transdisciplinaria”, eds. L. M. Gutiérrez Robledo, J. H. Gutiérrez Ávila y F. M. G. Ávila Fematt. **Instituto Nacional de Geriátría, México D.F., México** [<http://www.geriatria.salud.gob.mx/descargas/17.pdf>].

R. Fossion, E. Landa, P. Stránský, V. Velázquez, J.C. López Vieyra, I. Garduño, D. García y A. Frank, 2010b. Scale invariance as a symmetry in physical and biological systems: Listening to photons,

bubbles and heartbeats. *AIP Conf. Proc.*, 1323, pp. 74-90.

R. Fossion, 2011. Medidas de complejidad de series de tiempo fisiológicas. En “Memorias de la XIX Escuela de Verano en Física”, Instituto de Física (UNAM, México D.F.) y Instituto de Ciencias Físicas (UNAM, Cuernavaca), 25 de julio – 5 de agosto de 2011.

R. Fossion, D.A. Hartasánchez, O. Resendis-Antonio y A. Frank, 2012. Criticality, adaptability and early-warning signals in time series in a discrete quasispecies model. En preparación.

J. D. Gardner, W. Young, S. Sloan, M. Robinson y P. B. Miner Jr., 2005. The fractal nature of human gastro-oesophageal reflux. *Aliment. Pharmacol. Ther.*, 22, pp. 823-830.

S. R. F. S. M. Gois, M.A. Savi, 2009. An analysis of heart rhythm dynamics using a three-coupled oscillator model. *Chaos, Solitons and Fractals*, 41, pp. 2553–2565.

A. L. Goldberger, D. R. Rigney y B. J. West, 1990. Chaos and fractals in human physiology. *Sci. Am.* 262, pp. 34-41.

A. L. Goldberger, 1996. Non-linear dynamics for clinicians: Chaos theory, fractals and complexity at the bedside. *Lancet*, 347, pp. 1312-1314.

A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23), e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>].

A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. Ch. Ivanov, C.-K. Peng, H. E. Stanley, 2002. Fractal dynamics in physiology: Alterations with disease and aging. *Proc. Nat. Acad. Sci.*, 99 (suppl.1), pp. 2466-2472.

R. González González, 2012. Una forma de contextualizar Zheng Qi energía antipatógena. *Rev. Argentina Acup.*, 140, pp. 22-26.

J. M. Hausdorff, L. Zemaný, C.-K. Peng y A. L. Goldberger, 1999. Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children. *J. Appl. Physiol.*, 86(3), pp. 1040–1047.

K. S. Heffernan, J. J. Sosnoff, C. A. Fahs, K. K. Shinsako, S. Y. Jae, B. Fernhall, 2008. Fractal scaling properties of heart rate dynamics following resistance exercise training. *J. Appl. Physiol.*, 105, pp. 109-113.

K. León, R. Pérez, A. Lage y J. Carneiro, 2000. Modelling T-cell-mediated suppression dependent on interactions in multicellular conjugates. *J. Theo. Biol.*, 207, 231-254.

K. León Monzón, R. Pérez Rodríguez y A. Lage Dávila, 2003. Modelo matemático de la regulación del sistema inmune mediante la tolerancia dominante. *Biotechnol. Aplic.*, 20 (4), 253-257.

L. A. Lipsitz y A. L. Goldberger, 1992. Loss of “Complexity” and aging: Potential application of fractals and chaos theory to senescence. *J. Am. Med. Assoc.*, 267 (13), pp. 1806-1809.

L. A. Lipsitz, 2002. Dynamics of Stability: The Physiologic Basis of Functional Health and Frailty. *J. Gerontol.*, 57 A, pp. B115-B125.

D. Lloyd, M. A. Aon y S. Cortassa, 2001. Why homeodynamics, not homeostasis? *TheScientificWorld*, 1, pp. 133-145.

I. D. Loram y M. Lakie, 2002. Human balancing of an inverted pendulum: position control by small, ballistic-like, throw and catch movements. *J. Physiol.* (2002), 540.3, pp. 1111–1124.

M. C. Mackey y J. G. Milton, 1987. Dynamical disease. *Ann. N. Y. Acad. Sci.*, 504 (1), pp. 16-32.

M. Martínez-Lavín, A. G. Hermosillo, M. Rosas y M.-E. Soto, 1995. Circadian studies of autonomic nervous balance in patients with fibromyalgia: A heart rate variability analysis. *Arthr. Rheum.*, 41, pp. 1966-1971.

M. Martinez-Lavin, O. Infante, y C. Lerma, 2008. Hypothesis: The Chaos and Complexity Theory May Help our Understanding of Fibromyalgia and Similar Maladies. *Semin. Arthritis Rheum.*, 37 (4), pp. 260-264.

M. Martinez-Lavin, A. Vargas, 2009. Complex Adaptive Systems Allostasis in Fibromyalgia. *Rheum. Dis. Clin. N. Am.*, 35, pp. 285–298

J. C. Nelson, Rizwan-Uddin, M. P. Griffin, J. R. Moorman, 1998. Probing the Order within Neonatal Heart Rate Variability. *Ped. Res.*, 43 (6), pp. 823-831.

M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, B. Walker, 2001. Catastrophic shifts in ecosystems. *Nature*, 413, pp. 591-596.

M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, G. Sugihara, 2009. *Nature*, 461, pp. 53-59.

M. Srinivasan y A. Ruina, 2006. Computer optimization of a minimal biped model discovers walking and running. *Nature*, 439, 72-75.

Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996. Heart rate variability: Standards of measurement, physiological interpretation and clinical use. *Eur. Heart J.*, 17, pp. 354-381.

R. Varadhan, C. L. Seplaki, Q. L. Xue, K. Bandeen-Roch y L.P. Fried, 2008. Stimulus-response paradigm for characterizing the loss of resilience in homeostatic regulation associated with frailty. *Mech.Ageing Dev.*, 129, pp. 666-670.

F. E. Yates, 2008. Homeokinetics/Homeodynamics: A Physical Heuristic for Life and Complexity. *Ecol. Psychol.*, 20, pp. 148-179.

Sobre la expansión actual del universo

G. Germán*

*Instituto de Ciencias Físicas,
Universidad Nacional Autónoma de México,
Apdo. Postal 48-3, 62251 Cuernavaca, Morelos, México*

El universo acelerado es la observación de que éste parece estar expandiéndose a una velocidad cada vez mayor. El Premio Nobel de Física 2011 fue otorgado a Saul Perlmutter, Brian P. Schmidt y Adam G. Riess por el descubrimiento en 1998 de la expansión acelerada del Universo a través de observaciones de supernovas distantes con corrimiento al rojo $z \sim 0,5$. Este descubrimiento ha sido corroborado por fuentes independientes: la radiación de fondo de microondas cósmicas y estructura a grandes escalas, tamaño aparente de oscilaciones acústicas de bariones, la edad del universo así como por mediciones mejoradas de supernovas y de propiedades de rayos-X de cúmulos de galaxias. Usando supernovas muy distantes como candelas estándar es posible trazar la historia de la expansión cósmica del universo.

Una candela estándar (o fuente luminosa estándar), es un objeto astronómico de luminosidad conocida. Comparando la luminosidad conocida con la observada, la distancia al objeto puede ser calculada usando la ley inversa del cuadrado. Así, la distancia-luminosidad D_L de un objeto queda definida en términos de la relación entre su magnitud absoluta M y la magnitud aparente m según la ecuación $M = m - 5(\text{Log}_{10} D_L - 1)$, de donde se sigue que

$$D_L = 10^{\frac{m-M+5}{5}}. \quad (1)$$

Einstein dio instrumentos rigurosos para el estudio del universo al descubrir las ecuaciones de la Relatividad General

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu} + \Lambda g_{\mu\nu}. \quad (2)$$

Típicamente, las observaciones cosmológicas se realizan usando el espectro electromagnético completo. En orden creciente de energías se dispone de ondas de radio, microondas, infrarrojo, luz visible, ultravioleta, rayos-x y rayos gama. Sin embargo más recientemente se han realizado observaciones utilizando neutrinos (que son partículas que interactúan débilmente) y rayos cósmicos (partículas altamente relativistas). Las observaciones a muy grandes escalas (a nivel de supercúmulos de galaxias) apoyan el Principio Cosmológico, que es la

*Correo-e: gabriel@fis.unam.mx.

proposición de que el universo es homogéneo e isotrópico. Este principio encuentra su expresión matemática a través de la métrica de Friedmann-Robertson-Walker (FRW). Por la parte teórica la cosmología descansa básicamente en el Principio Cosmológico, las ecuaciones de la Relatividad General y la ecuación de estado del tipo de materia bajo estudio. El Principio Cosmológico es la hipótesis de que todos los puntos del universo son equivalentes. Este principio se implementa a través de la métrica FRW para un universo homogéneo e isotrópico. En este caso, el intervalo espacio-temporal está dado por

$$ds^2 = a(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right), \quad (3)$$

De aquí podemos leer directamente la métrica de FRW

$$g_{\mu\nu} = \text{diag} \left(-1, \frac{a^2(t)}{1 - kr^2}, a^2(t)r^2, a^2(t)r^2 \sin^2 \theta \right), \quad (4)$$

en tanto que el tensor de energía-momento para un fluido perfecto es

$$T_{\mu\nu} = \text{diag} (\rho, -p, -p, -p). \quad (5)$$

La ecuación de estado establece una relación entre la presión y la densidad de energía $p = p(\rho)$ y está dada por

$$p = \omega\rho. \quad (6)$$

De las ecuaciones de Einstein para $g_{\mu\nu}$ y $T_{\mu\nu}$ dadas arriba se sigue la ecuación de Friedmann

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho + \frac{k}{a^2}, \quad (7)$$

y la ecuación de aceleración

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3p). \quad (8)$$

La ecuación de continuidad o de conservación de la energía se deriva de las últimas dos

$$\dot{\rho} = -3H(\rho + p), \quad (9)$$

donde H es la “constante” de Hubble definida por

$$H = \frac{\dot{a}}{a}. \quad (10)$$

La constante k que aparece en la Ec. (3) determina la curvatura del universo. Si $k = 0$ el universo es plano, en tanto que si $k > 0$ o $k < 0$ el universo es esférico (cerrado) o hiperbólico (abierto), respectivamente. Observaciones recientes sugieren que el universo es plano o muy próximo a plano. Con el avance de la tecnología conceptos y predicciones de la Relatividad General pueden ser estudiados.

Mediante el corrimiento al rojo de candelas estándar (CE) es posible trazar la historia de la expansión cósmica del universo. La expansión cósmica estira distancias entre cúmulos de galaxias y los fotones sufren corrimiento al rojo dado por

$$z = \frac{\lambda_{\text{observada}} - \lambda_{\text{emitida}}}{\lambda_{\text{emitida}}}, \quad (11)$$

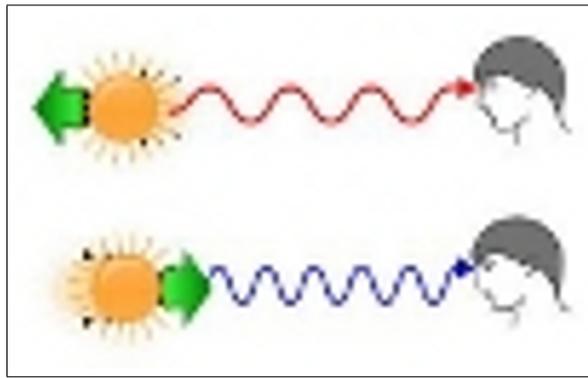


Figura 1: Corrimiento al rojo contra corrimiento al azul

El Corrimiento al rojo ocurre cuando la longitud de onda de la luz que viene desde un objeto en movimiento se incrementa proporcionalmente en su longitud de onda o se desplaza hacia el extremo rojo del espectro.

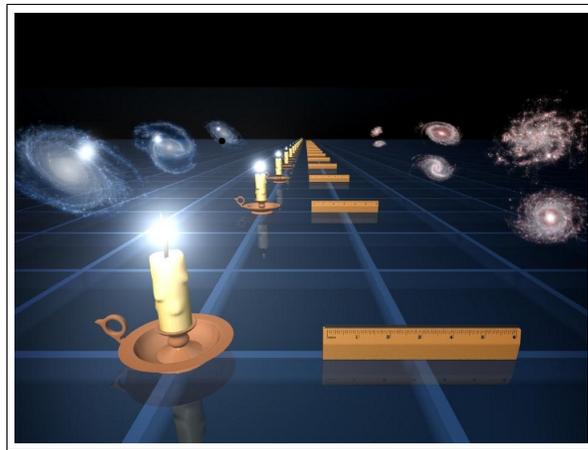


Figura 2: Candelas Estándar

Una candela estándar (o fuente luminosa estándar), es un objeto astronómico de luminosidad conocida. Comparando la luminosidad conocida con la observada, la distancia al objeto puede ser calculada usando la ley inversa del cuadrado. Una colección de mediciones de candelas estándar sobre un rango de distancias adecuado daría un historial de la expansión del universo.

z es un factor por el cual el universo se ha estirado desde que la luz dejó la fuente. Una colección de mediciones de candelas estándar sobre un rango de distancias adecuado daría un récord sobre la historia de expansión del universo.

E. Hubble descubrió la expansión del universo en 1929 usando galaxias como candelas estándar. Pero galaxias de tamaños y formas diversas son difíciles de usar como CE.

En los años 70's se sugirió usar como CE al miembro más brillante de una galaxia, sin embargo esto es susceptible de cambio evolucionario. Por ejemplo podemos usar supernovas como CE y así determinar la velocidad de recesión de la galaxia huésped.

En 1938, Walter Baade, trabajando con Fritz Zwicky, señaló que las supernovas eran candidatos prometedores para medir la expansión cósmica. Su pico de brillantes es bastante

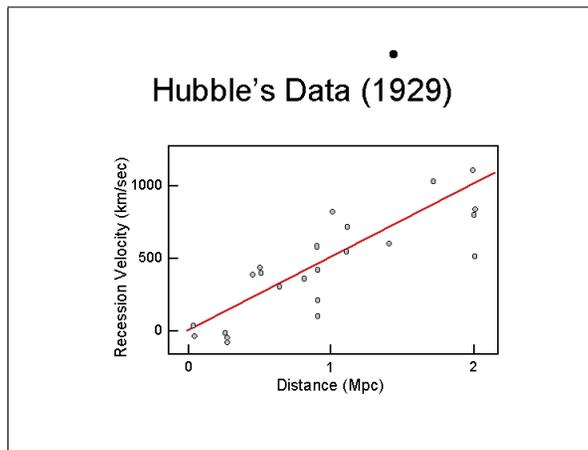


Figura 3: Diagrama de Hubble 1929

Diagrama de Hubble mostrando la velocidad de alejamiento de galaxias como función de la distancia. A partir de estos datos Hubble descubrió que la distancia a un galaxia es proporcional a su corrimiento al rojo.

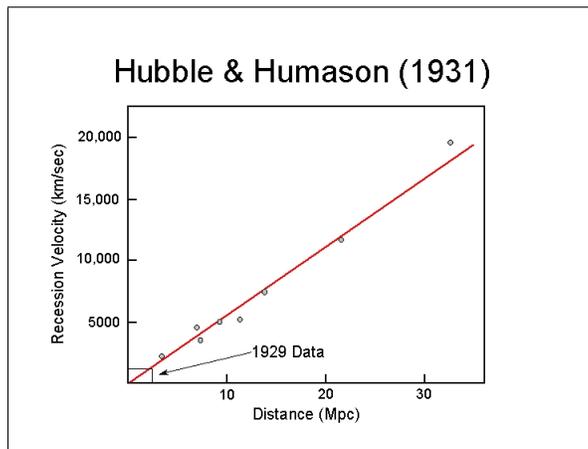


Figura 4: Diagrama de Hubble 1931

Diagrama de Hubble mostrando un número mayor de datos, compárese con la figura anterior.

uniforme, pero se encontró que formaban grupos bastante heterogéneos con amplio rango de picos de brillantes.

A principios de los 1980's surgió una nueva subclasificación de supernovas: Ia sin hidrógeno en su espectro pero con una característica de absorción de silicón a 6150A, también el tipo Ib sin absorción. Se descubrió una asombrosa consistencia en las supernovas tipo Ia. Supernovas más débiles lucen típicamente más rojas o en galaxia espirales inclinadas (o ambas) debido posiblemente a atenuación por polvo (gas) que absorbe más luz azul que roja.

Las supernovas tipo Ia son problemáticas porque son raras (una galaxia típica aporta un par de explosiones por milenio), aleatorias (no se sabe donde ni cuando explotarán), y de corta duración. La dificultad principal para su estudio se originaba además en un problema más pragmático: para observar las supernovas se requiere reservar un telescopio pero esto se

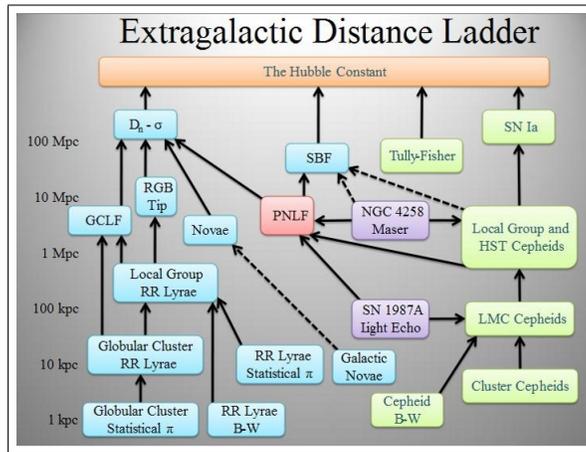


Figura 5: Escalera de distancias extragalácticas

La escalera de distancias extragalácticas (o escala de distancias extragalácticas) es el nombre que se le da a una sucesión de métodos que el astrónomo usa para determinar distancias a galaxias.

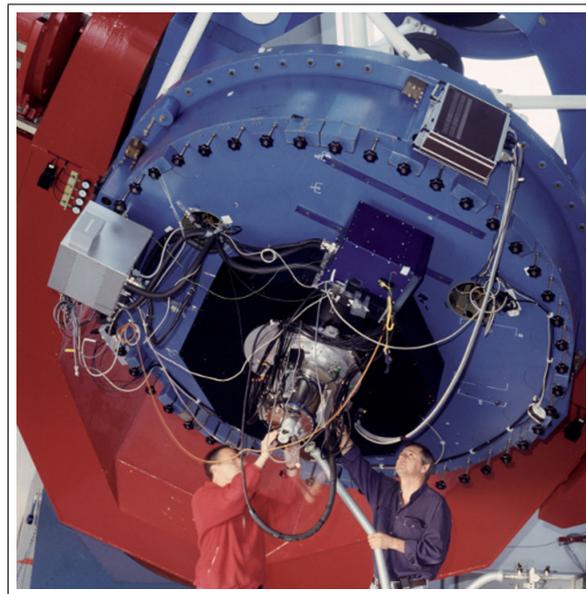


Figura 6: Cámara de campo grande

La cámara de campo grande WFI (Wide Field Imager) es una cámara focal reductora montada en el telescopio MPG/ESO de 2.2-m en La Silla, Chile, con un campo de visión de 34'x33'.

concedía sólo con conocimiento de a donde mirar, lo cual requiere a su vez disponer de un telescopio...

Los primeros intentos fueron poco alentadores: el equipo Danés de Hans Norgaard-Nielsen encontró sólo una supernova tipo Ia en dos años de observaciones y ya pasada de su pico. El problema finalmente se resolvió mediante el diseño y uso de una cámara de campo grande ya que una exposición captaría alrededor de mil galaxias y en una noche adecuada podrían

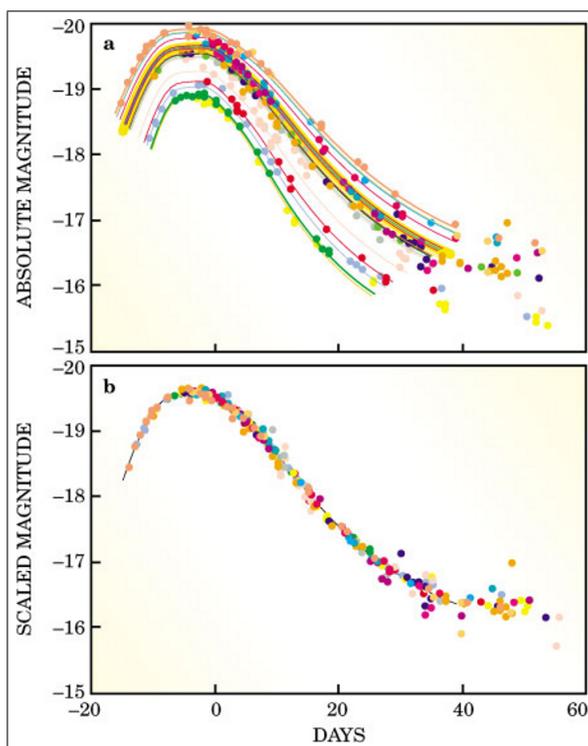


Figura 7: Curvas de luz

Curvas de luz de supernovas cercanas del tipo Ia medidas por Mario Hamuy y colaboradores. La figura superior muestra la magnitud absoluta contra el tiempo antes y después del pico de brillantés. La segunda figura muestra que todas las curvas de luz de supernovas tipo Ia pueden hacerse coincidir si estiramos las escalas de tiempo y después escalamos la brillantés por una determinada cantidad.

lograrse de 50 a 100 campos. Esto permitiría estudiar miles de galaxias distantes mejorando las expectativas de encontrar supernovas. Durante la fase oscura de la luna se usó el telescopio Hubble para dar seguimiento a las curvas de luz de las supernovas que se encontraban en un grado cuadrado del cielo.

Los primeros resultados fueron presentados a principios de 1997 por el grupo SCP (por sus siglas en inglés: Supernova Cosmology Project) en donde un análisis de siete supernovas de alto corrimiento al rojo parecían indicar un retardamiento en la expansión de acuerdo con los modelos usuales de expansión del universo. Sin embargo las barras de errores eran tan grandes no era posible establecer una conclusión definitiva. Conforme el número de supernovas estudiadas fue incrementándose quedó claro que los modelos más simples no podían explicar el comportamiento observado. Éstos se modificaron mediante la introducción de una constante cosmológica lo cual implicaba una expansión acelerada del universo en la época correspondiente a las observaciones, alrededor de unos seis o siete mil millones de años desde la época actual.

En las ecuaciones de Einstein, es la densidad de masa la que determina la velocidad de expansión del universo así como su curvatura, edad y destino. Por esto, el que un modelo con constante cosmológica sea requerido para explicar la luminosidad de las supernovas

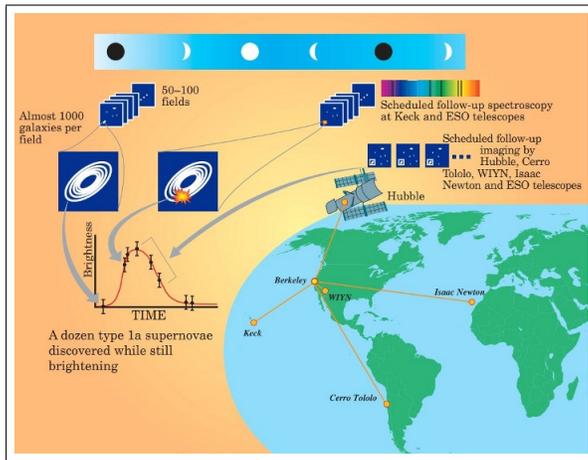


Figura 8: Estrategia de observación de supernovas

Se observa una pequeña región del cielo justo después de la luna nueva, cuando hay menos luz en el cielo, y posteriormente se observa la misma región justo antes de la siguiente luna nueva. Esto permite descubrir y dar seguimiento a por lo menos una docena de supernovas en la cercanía de su pico de brillantés. El seguimiento se hace con el telescopio Hubble y con telescopios terrestres.

estudiadas tiene a su vez una consecuencia notable: debe existir en el universo una cierta cantidad de energía-materia que haga la expansión. Dado que no parece tratarse de materia conocida (bariones) se le ha denominado *energía oscura* y su proporción con respecto al total de energía del universo debe corresponder al 72 por ciento del total!

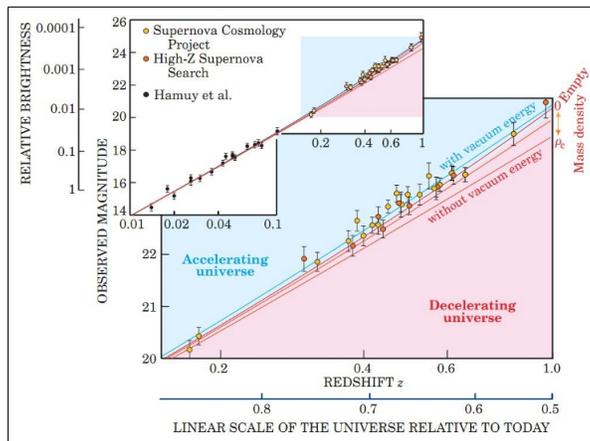


Figura 9: Magnitud observada contra corrimiento al rojo

Una gráfica de la magnitud observada contra corrimiento al rojo muestra que alrededor de $z \sim 0,5$ los datos empiezan a favorecer un universo con energía de vacío distinta de cero (constante cosmológica). Esto implica, a través de las ecuaciones de Einstein, una expansión acelerada del universo.

Descubrir la naturaleza de esta energía es uno de los principales problemas abiertos de la física actual.

Investigación realizada gracias al apoyo del “Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica” (PAPIIT) UNAM, IN103413-3, *Teorías de Kaluza-Klein, inflación y perturbaciones gravitacionales*.

Se agradece también al Dr. José Récamier por la gentileza de invitarme a participar en la XX Escuela de Verano en Física, mediante la impartición de una plática, resumida en estas notas.

Referencias

- [1] S. Perlmutter, *Supernovae, Dark Energy, and the Accelerating Universe*, Physics Today, April 2013.
- [2] The Nobel Prize in Physics 2011, Lectures: www.nobelprize.org.
- [3] E.W. Kolb and M.S. Turner, *The Early Universe*, Addison-Wesley, Pub. Co., 1990.
- [4] P.J.E. Peebles, *Principles of Physical Cosmology*, Princeton U. Press, 1993.
- [5] J.A. Peacock, *Cosmological Physics*, Cambridge U. Press, 1999.
- [6] A. Liddle, *An Introduction to Modern Cosmology*, Sec Ed., John Wiley & Sons, 2003.
- [7] S. Dodelson, *Modern Cosmology*, Academic Press, 2003.
- [8] S. Weinberg, *Cosmology*, Oxford University Press, 2008.

SIMULACIONES POR COMPUTADORA DE LA CRISTALIZACIÓN COLOIDAL

Agustín E. González

Centro de Ciencias Físicas, Universidad Nacional Autónoma de México,
Av. Universidad S/N, Col. Chamilpa, 62210 Cuernavaca, Morelos, México
Email: agus@fis.unam.mx

RESUMEN

Al efectuar simulaciones de Monte Carlo cinético en 2D de la cristalización coloidal encontramos que las fronteras de los núcleos cristalinos no son únicamente rugosas, como fue observado por los experimentales, sino también fractales, cuyo valor (d_f) se calculó. La frontera correspondiente para los cristales, arriba del tamaño crítico (N_c), también es fractal pero más lisa. Nuestras simulaciones nos permiten calcular además N_c , la tensión de línea (γ) así como la diferencia de potencial químico entre las dos fases ($\Delta\mu$). Sin embargo, a diferencia del procedimiento de los experimentales, encontramos que la fractalidad de las fronteras son necesarias para calcular γ y $\Delta\mu$.

La cristalización es un tema central en la física de la materia condensada. Reviste una gran importancia el tener un entendimiento cabal de la fabricación controlada de materiales cristalinos, para muchas aplicaciones. En particular, cristales formados por partículas coloidales son de gran interés para desarrollar materiales novedosos, tales como materiales con brechas de bandas fotónicas, dado el hecho de que el tamaño de dichas partículas así como la separación entre partículas vecinas cercanas son del orden de la longitud de onda de la luz visible. Más aún, los materiales coloidales ofrecen oportunidades únicas para estudiar la nucleación y el crecimiento de cristales en una escala microscópica, dado que los “átomos” (partículas) del sistema son de tamaño coloidal (mesoscópico) y, por lo tanto, son más fáciles de observar con simple microscopía óptica que átomos o moléculas simples. Además de esto, la dinámica de dichos sistemas coloidales se hace mucho más lenta, pudiendo entonces seguir en detalle, a nivel de partícula individual, todo el proceso de cristalización.

Las simulaciones por computadora también ofrecen una oportunidad única para estudiar procesos que experimentalmente son muy difíciles de visualizar, como el proceso de cristalización. El poder variar a nuestro antojo los parámetros relevantes de dichos procesos, como el potencial de interacción entre las partículas, la concentración, la temperatura, etc., les confiere a dichas simulaciones un carácter imprescindible, sin las cuales nuestro conocimiento de esos procesos se vería empobrecido enormemente. Además, las herramientas computacionales ofrecen la posibilidad de visualizar los resultados de las simulaciones, los cuales se pueden comparar con, por ejemplo, la videomicroscopía del proceso de cristalización coloidal.

Motivados por todo esto, estamos desarrollando una línea de investigación de la cristalización coloidal vía simulaciones por computadora. Ya hemos obtenido resultados nuevos del proceso de nucleación cristalina, que ocurre cuando se forman pequeñas regiones cristalinas (cristalitos) a partir de las fluctuaciones estructurales en un líquido. Estos cristalitos la mayor parte de las veces se empequeñecen y desaparecen aunque, de vez en cuando, si son suficientemente grandes, empiezan a crecer aún más para formar los cristales. También hemos obtenido resultados nuevos para el proceso de “maduración cristalina”, que consiste en que los cristales, al chocar en su crecimiento, forman las llamadas fronteras de grano,

las cuales eventualmente desaparecen en ciertos casos, formándose así un “grano cristalino” más grande a partir de los dos granos cristalinos que chocaron. El propósito de este artículo es describir nuestros resultados obtenidos en el proceso de la nucleación cristalina.

Para el caso de 2D, P. Pieranski [1] fue el primero en estudiar el ordenamiento de partículas coloidales de latex en la interfaz aire-agua, lo que se atribuyó a las fuerzas repulsivas de largo alcance entre las partículas *confinadas*. Sin embargo, para el estudio del ordenamiento en sistemas diluidos, se requieren fuerzas atractivas entre las partículas en lugar de fuerzas repulsivas. G. Y. Onoda [2] fue el primero que tuvo éxito en encontrar las partículas correctas, que consistían en partículas de poliestireno estabilizadas con dodecil sulfato de sodio injertado en sus superficies. Con este tensoactivo, él evitó que las partículas coloidales se agregaran mientras que mantenía aún la atracción de van der Waals entre ellas. En las fotografías mostradas por este autor, se puede ver claramente que las fronteras de los cristalitos tienen una estructura rugosa, posiblemente fractal, aunque no le prestó atención a este hecho.

La mayor parte de las teorías de cristalización y nucleación cristalina consideran a las fronteras entre los cristalitos o los granos cristalinos y el fluido que los rodea como *suave*. De hecho, en la teoría clásica de nucleación [3], se postula que los cristalitos tienen forma esférica (circular en 2D). El crecimiento de estas regiones a un grano cristalino depende de dos factores: una disminución en la energía del bulto, que favorece el crecimiento, y un aumento en la energía superficial, que lo inhibe. Estos dos factores se ven reflejados en el cambio de la energía libre para la formación de un cristalito circular (en 2D) de radio r :

$$\Delta G = 2\pi r \gamma - \pi r^2 \Delta\mu n. \quad (1)$$

En donde γ es la tensión de línea de la frontera de los cristalitos con el fluido, $\Delta\mu$ es la diferencia entre los potenciales químicos de las fases fluida y cristalina y n es el número de partículas en el cristal por unidad de área. Esta función tiene un pico en el radio crítico del cristalito $r_c = \gamma/\Delta\mu n$; abajo de r_c el cristalito tiende a empequeñecerse y desaparecer, mientras que arriba de r_c tiende a crecer a un grano cristalino. Como es algo difícil el cálculo de γ y $\Delta\mu$ de primeros principios, estas cantidades son a menudo tratadas como parámetros usados para ajustarse al experimento. No obstante esto, nosotros fuimos capaces de obtener dichas cantidades directamente de nuestras simulaciones [4]. Empezando con el trabajo de Onoda [2] y considerando que las unidades que cristalizan son coloides mesoscópicos suspendidos en un fluido, el procedimiento de simulación *ad hoc* debe ser un Monte Carlo cinético con pasos aleatorios cortos, dado que las partículas coloidales están efectuando este tipo de movimiento difusivo antes de unirse a otras partículas de un cristalito o de un grano cristalino.

Consideramos [4] entonces $N_{tot} = 53715$ discos distribuidos aleatoriamente en una caja con condiciones de frontera periódicas, teniendo cuidado de no traslapar los discos duros. Los discos duros ocupan una fracción de área de $\phi = 0.15$. Consideramos el potencial atractivo de interacción entre dos partículas, que se muestra en la Fig. 1, en donde r es la distancia entre los discos normalizada por el diámetro de la esfera dura. El algoritmo es como sigue: (i) se escoge una partícula al azar y se mueve un paso de longitud igual a un décimo del diámetro del disco duro en una dirección aleatoria. (ii) Se calcula la nueva energía de interacción U_f a la que se le sustrae la inicial: $\Delta U = U_f - U_i$. (iii) Si $\Delta U \leq 0$ se acepta el movimiento. (iv) Si $\Delta U > 0$ únicamente se acepta el movimiento si $exp(-\Delta U/k_B T) > R$, en donde R es un número aleatorio uniformemente distribuido en $[0,1)$. (v) En cualquiera de los dos últimos casos el tiempo de Monte Carlo t se aumenta por $1/N_{tot}$. Nos vamos entonces al punto (i). Cuando $\Delta t = 1$, que define una barrida de Monte Carlo, cada partícula ha

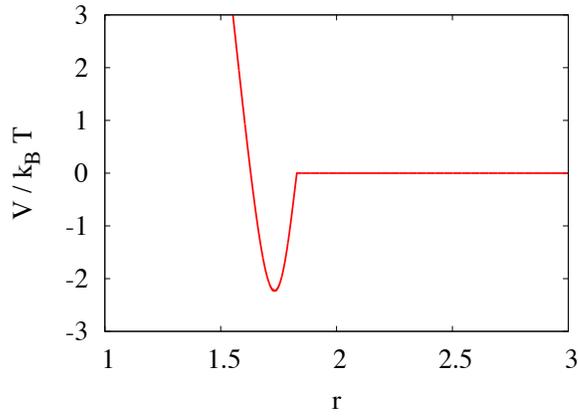


Figure 1: El potencial de interacción entre dos partículas. La distancia r se ha escalado con el diámetro de la esfera dura, mientras que la energía potencial V se ha escalado por $k_B T$.

intentado moverse una vez *en promedio*.

Visualizando los resultados de la simulación, fuimos capaces de ver la formación continua de cristalitas pequeños y de tamaño medio, con una simetría de red triangular; los pequeños desapareciendo casi inmediatamente después de la formación, mientras que algunos de los más grandes empezaron a crecer, indicando que nos encontramos en el régimen sobreenfriado y sobresaturado. Dado que observamos cristalitas altamente anisotrópicos, como los obtenidos experimentalmente por G. Y. Onoda [2], decidimos obtener el tamaño crítico en términos del número crítico de partículas, N_c , y no en términos de algún radio crítico. En la Fig. 2 mostramos una sección de la caja de simulación en donde, a la izquierda, se muestran todas las partículas en esa sección mientras que, a la derecha, se muestran únicamente las partículas que pertenecen a cristalitas con más de 10 partículas. Nótese el aspecto quebrado de las fronteras de los cristalitas, las que demostramos [4] que tienen una estructura fractal.

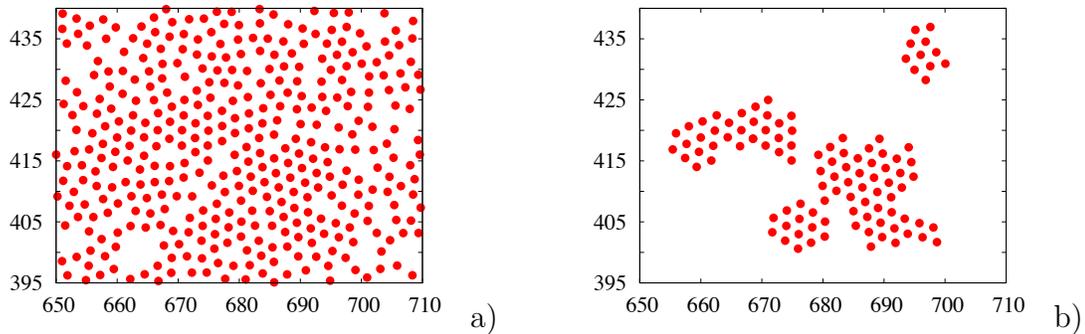


Figure 2: Una sección de la caja de la simulación en donde, a la izquierda, se muestran todas las partículas en esa sección mientras que, a la derecha, únicamente partículas que pertenecen a cristalitas con más de de 10 partículas son mostradas.

Para hacer el cálculo del tamaño crítico N_c , primero calculamos [4] la probabilidad de crecimiento $P_g(N)$ así como la probabilidad de decrecimiento $P_s(N)$, de los cristalitas de tamaño N . La función $P_g(N) - P_s(N)$ debe ser negativa abajo del tamaño crítico y positiva arriba de él. El punto de cruce de esta función con el eje horizontal N , marca la posición del tamaño crítico N_c . Esto se ve claramente en la Fig. 3, en donde se calculó dicha función en 7 intervalos de igual magnitud en la escala logarítmica N . De dicha gráfica se estimó el

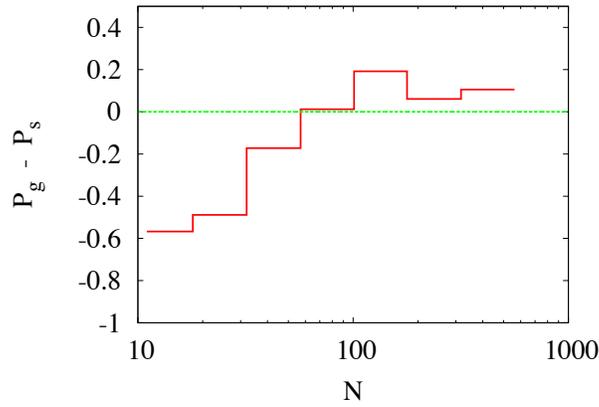


Figure 3: La probabilidad de crecimiento menos la probabilidad de decrecimiento de un cristalito, evaluada en 7 intervalos diferentes de tamaño constante en la escala logarítmica N . El tamaño crítico se estima de esta gráfica, quedando entre 70 y 80 partículas.

tamaño crítico en alrededor de 74 partículas.

Para hacer el cálculo de la dimensión fractal de la frontera de los cristalitos, graficamos en escala doble-logarítmica el número promedio de partículas en la frontera, L , contra el tamaño lineal promedio, ξ . La frontera debe escalar como $L \sim \xi^{d_f}$, en forma análoga a como lo hace la curva de Koch [5], en donde d_f es la dimensión fractal de la frontera. Haciendo este ejercicio [4] encontramos $d_f \approx 1.315$ para los cristalitos (abajo del tamaño crítico N_c), mientras que $d_f \approx 1.118$ para los cristales (arriba del tamaño crítico). Para la forma funcional de L en términos de N , obtuvimos [4]:

$$\ln L = 0.917 + 0.631 \ln N, \quad (2)$$

En lugar de la energía libre clásica de formación de un cristalito de radio r , la escribimos explícitamente en términos de la longitud de la frontera $l = L l_o$, en donde l_o es el espaciamento de la red cristalina (que corresponde a la posición del mínimo del pozo de potencial), así como del número de partículas N :

$$\Delta G = \gamma l - \Delta\mu N. \quad (3)$$

Como es bien sabido, el primer término γl domina este cambio para los cristalitos pequeños. La probabilidad de activación de un cristalito de tamaño N y perímetro l escala como $\exp(-\beta \Delta G)$. Por lo tanto, el número de pequeños cristalitos en el sistema escala como $NCr \sim \exp(-\beta \gamma l)$. De la pendiente de la línea recta en la Fig. 4, para los primeros puntos, podemos evaluar la tensión de línea en términos de la temperatura y el espaciamento de la red como

$$\gamma \simeq 0.399 k_B T / l_o. \quad (4)$$

Ahora sustituimos las ecuaciones 2 y 4 en la ecuación 3 para tener una forma funcional de ΔG en términos de N . Tomando la derivada con respecto a N e igualándola a cero en $N = N_c \simeq 74$, digamos, obtenemos

$$\Delta\mu \simeq 0.128 k_B T \quad (5)$$

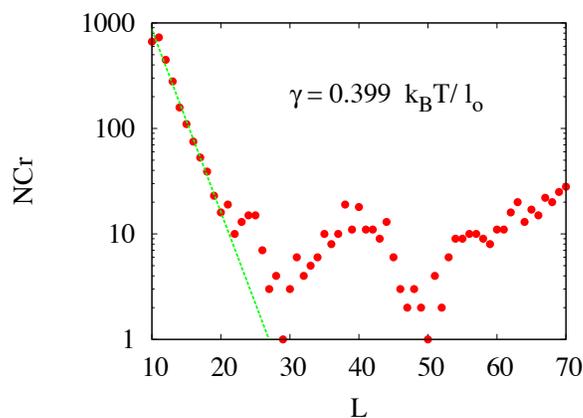


Figure 4: Gráfica semilogarítmica del número de cristalitos, NCr , contra L

Fue interesante notar que el tamaño crítico de los cristalitos, obtenido por el método de las probabilidades, coincidía aproximadamente con el punto de ruptura de las líneas rectas en el análisis fractal de las fronteras de los cristalitos y cristales. Para investigar este hecho producimos unas películas de los procesos de nucleación y crecimiento, obtenidas de nuestras simulaciones, que se pueden descargar del sitio <http://www.fis.unam.mx/~agus/>, para constatar cuáles son los mecanismos principales que actúan en esos dos procesos. En el archivo “2D_crystal_nucleation.mp4” podemos observar que los cristalitos son muy fugaces y efímeros, y solamente un atisbo de ellos puede ser obtenido. Pueden crecer adquiriendo más partículas del fluido circundante o, mucho más a menudo, pueden decrecer perdiendo partículas de sus fronteras que se van a la fase fluida, para luego desaparecer. No obstante, en la última parte del archivo “2D_crystal_nucleation.mp4” y en el otro archivo “2D_crystal_growth.mp4”, podemos ver que cuando los cristalitos alcanzan un tamaño apreciable (mayor que el tamaño crítico), se vuelven más o menos permanentes. No obstante esto, las fronteras no son fijas sino muy activas, adquiriendo más partículas del fluido que las rodea, que preferencialmente se van a las cavidades de estas fronteras en lugar de a las puntas, para poder establecer más enlaces con otras partículas de la estructura y “sentirse más a gusto” con una energía menor. Esta selección de zonas de agregación no puede ser obtenida con los cristalitos porque, durante el tiempo en que una partícula del fluido circundante trata de hacer esto, el cristalito desaparece. El efecto combinado de estos procesos es que, para los cristalitos, las fronteras se quedan aproximadamente como se formaron, mientras que, para los cristales, tenemos la formación de una frontera más lisa proveniente del llenado de las cavidades por las partículas del fluido circundante. Por lo tanto, la transición de cristalitos efímeros a cristales perdurables marca también la transición de una dimensión fractal de frontera relativamente alta a una dimensión fractal baja.

El autor agradece a DGAPA-UNAM (proyecto PAPIIT IN-106008) por el apoyo parcial. Agradece también al Comité de Supercómputo de la UNAM por los recursos computacionales asignados a través de DGSCA-UNAM.

References

- [1] P. Pieranski, Phys. Rev. Lett. **45**, 569 (1980).
- [2] G. Y. Onoda, Phys. Rev. Lett. **55**, 226 (1985).

- [3] K. F. Kelton, en: Solid State Physics, Vol. 45, editado por H. Ehrenreich and D. Turnbull, (Academic, New York, 1991) p. 75.
- [4] A. E. González y L. Ixtlilco-Cortés, Phys. Lett. A **376**, 1375 (2012).
- [5] B. B. Mandelbrot, The Fractal Geometry of Nature, (W. H. Freeman & Co., San Francisco, 1988).

Operadores de Escalera en 1,2 y 3 Dimensiones

Eugenio Ley Koo
Instituto de Física, UNAM
eleykoo@fisica.unam.mx

December 14, 2012

Abstract

Los operadores de creación y aniquilación del oscilador armónico en una dimensión cartesiana son familiares. En esta conferencia se ilustran sus extensiones en las versiones de dos dimensiones en coordenadas cartesianas y circulares, y en tres dimensiones en coordenadas cartesianas, circulares cilíndricas y esféricas. Se señalan también otras posibles extensiones.

1 1. Operadores de Ascenso y Descenso entre las Eigenfunciones del Oscilador Armónico en una Dimensión Cartesiana.

En esta sección se revisan las acciones de los operadores llamados de creación y aniquilación entre las eigenfunciones del oscilador armónico en una dimensión cartesiana. Aquí nos referimos a ellos con los nombres de operadores de escalera, de ascenso y de descenso, respectivamente.

El punto de partida es la ecuación de Schrödinger

$$\left[\frac{\hat{P}_x^2}{2m} + \frac{1}{2}m\omega^2 x^2 \right] \psi_n(x) = E_n \psi_n(x) \quad (1)$$

cuyo operador Hamiltoniano entre corchetes contiene a la coordenada de posición x , y el operador de la cantidad de movimiento es

$$\hat{P}_x = -i\hbar \frac{d}{dx} \quad (2)$$

en términos de la constante de Planck $\hbar = h/2\pi$. La unidad natural de energía es $\hbar\omega$, y las unidades asociadas de longitud $\sqrt{\hbar/m\omega}$ y de cantidad de movimiento $\hbar/\sqrt{\hbar/m\omega} = \sqrt{m\omega/\hbar}$.

Entonces es útil usar las cantidades adimensionales

$$\epsilon_n = \frac{E_n}{\hbar\omega}, \xi = \sqrt{\frac{m\omega}{\hbar}}x, \pi_x = \sqrt{\frac{\hbar}{m\omega}}P_x \quad (3)$$

y reescribir la ecuación (1) en la forma

$$\left[-\frac{1}{2} \frac{d^2}{d\xi^2} + \frac{1}{2} \xi^2 \right] \psi_n(\xi) = \epsilon_n \psi_n(\xi) \quad (4)$$

Los operadores de escalera

$$a_+ = \xi - \frac{d}{d\xi} \quad \text{y} \quad a_- = \xi + \frac{d}{d\xi} \quad (5)$$

se identifican como operadores de ascenso y descenso, respectivamente. Efectivamente si consideramos los productos de estos operadores

$$a_+ a_- = \xi^2 + \xi \frac{d}{d\xi} - \frac{d}{d\xi} \xi - \frac{d^2}{d\xi^2} = 2\hat{h} - 1 \quad (6)$$

$$a_- a_+ = \xi^2 - \xi \frac{d}{d\xi} + \frac{d}{d\xi} \xi - \frac{d^2}{d\xi^2} = 2\hat{h} + 1 \quad (7)$$

y se reconocen las dos formas alternativas de escribir el Hamiltoniano dentro del corchete de la ecuación (4):

$$h = \frac{1}{2} (a_+ a_- + 1) = \frac{1}{2} (a_- a_+ - 1). \quad (8)$$

Como se muestra en los libros de texto [1,2], esta factorización lleva a la identificación de $a_+ a_-$ con el operador de número n , y el espectro de niveles de energía igualmente espaciados $\epsilon_n = n + \frac{1}{2}$ para $n = 0, 1, 2, \dots$ a partir del nivel de energía más bajo o de estado base $\epsilon_0 = 1/2$.

Primero, se ilustra la construcción de la eigenfunción del estado base $\psi_0(\xi)$, a partir de que el operador a_- lo aniquila:

$$\left(\xi + \frac{d}{d\xi}\right)\psi_0 = 0 \quad (9)$$

La solución de esta ecuación es la exponencial gaussiana

$$\psi_0(\xi) = e^{-\xi^2/2}. \quad (10)$$

A continuación se aplica el operador a_+ a la eigenfunción del estado base para obtener el primer estado excitado con $n = 1$, $\epsilon_1 = 3/2$;

$$\left(\xi - \frac{d}{d\xi}\right)\psi_0 = \left(\xi - \frac{d}{d\xi}\right)e^{-\xi^2/2} = 2\xi e^{-\xi^2/2} = \psi_1(\xi) \quad (11)$$

en que $2\xi = H_1(\xi)$ se identifica como el polinomio de Hermite de orden 1.

Las aplicaciones sucesivas del operador a_+ lleva a los polinomios de Hermite $H_n(\xi)$ para los estados excitados mostrando la acción de ascenso sobre las eigenfunciones y los niveles de energía sucesivos con $n = 2, 3, \dots$

n	0	1	2	3	4
$H_n(x)$	1	2ξ	$4\xi^2 - 1$	$8\xi^3 - 12\xi$	$16\xi^4 - 4\xi^2 + 12$

También se invita al lector a aplicar el operador a_- a cualquier estado excitado y corroborar su efecto de descenso.

2 Osciladores Armónicos en Dos Dimensiones en Coordenadas Cartesianas y Circulares

El oscilador armónico isotrópico en dos dimensiones cartesianas tiene un Hamiltoniano que es la suma de los dos Hamiltonianos en las coordenadas x y y de la ecuación (1) o ξ y η de la ecuación (4). Aquí escribimos la última versión en la forma

$$\left[-\frac{1}{2}\left(\frac{\partial}{\partial\xi^2} + \frac{\partial}{\partial\eta^2}\right) + \frac{1}{2}(\xi^2 + \eta^2)\right]\psi_{n_\xi n_\eta}(\xi, \eta) = \epsilon_{n_\xi n_\eta}\psi_{n_\xi n_\eta}.$$

La aditividad de los Hamiltonianos hace posible la serparación de esta ecuación de Schrödinger con la eigenfunción factorizable

$$\psi_{n_\xi n_\eta}(\xi, \eta) = \psi_{n_\xi}(\xi)\psi_{n_\eta}(\eta) \quad (12)$$

y eigenenergías también aditivas

$$\epsilon_{n_\xi n_\eta} = n_\xi + \frac{1}{2} + n_\eta + \frac{1}{2}. \quad (13)$$

a partir de sus respectivas contribuciones en cada dimensión, discutidas en la sección anterior.

La identificación de los operadores escalera hace uso de los respectivos vectores de posición y de gradiente:

$$\bar{a}_{++} = \hat{i} \left(\xi - \frac{\partial}{\partial \xi} \right) + \hat{j} \left(\eta - \frac{\partial}{\partial \eta} \right) \quad (14)$$

$$\bar{a}_{--} = \hat{i} \left(\xi + \frac{\partial}{\partial \xi} \right) + \hat{j} \left(\eta + \frac{\partial}{\partial \eta} \right) \quad (15)$$

El lector puede aplicar sistemáticamente estos operadores y también los operadores identidad I_ξ y I_η tales que

$$I_\xi \psi_{n_\xi}(\xi) = \psi_{n_\xi}(\xi) \quad I_\eta \psi_{n_\eta}(\eta) = \psi_{n_\eta}(\eta) \quad (16)$$

para obtener algunos de los siguientes resultados y otros:

$$1) \quad a_-(\xi) a_-(\eta) \psi_0(\xi) \psi_0(\eta) = 0 \\ \psi_0(\xi) \psi_0(\eta) = e^{-\xi^2/2} e^{-\eta^2/2} \quad \epsilon_{00} = 1$$

para el estado base

$$2) \quad a_+(\xi) I_\eta \psi_0(\xi) \psi_0(\eta) = \psi_1(\xi) \psi_0(\eta) \\ I_\xi a_+(\eta) \psi_0(\xi) \psi_0(\eta) = \psi_0(\xi) \psi_1(\eta)$$

para los dos primeros estados excitados con la misma energía $\epsilon_{10} = \epsilon_{01} = 2$

$$3) \quad a_+(\xi) a_+(\eta) \psi_0(\xi) \psi_0(\eta) = \psi_1(\xi) \psi_1(\eta) \\ a_+^2(\xi) I_\eta \psi_0(\xi) \psi_0(\eta) = \psi_2(\xi) \psi_0(\eta) \\ I_\xi a_+^2(\eta) \psi_0(\xi) \psi_0(\eta) = \psi_0(\xi) \psi_2(\eta)$$

para los segundos estados excitados con la misma energía $\epsilon_{11} = \epsilon_{20} = \epsilon_{02} = 3$

4) y su generalización

$$a_+^{n_\xi}(\xi) a_+^{n_\eta}(\eta) \psi_0(\xi) \psi_0(\eta) = \psi_{n_\xi}(\xi) \psi_{n_\eta}(\eta)$$

para

$$n_\xi + n_\eta = N, \quad n_\xi = 0, 1, 2, \dots, N \\ n_\eta = N, N-1, N-2, \dots, 0$$

con un total de $N + 1$ estados con la misma energía $\epsilon_{n_\xi n_\eta} = n_\xi + n_\eta + 1$.

La extensión para coordenadas circulares es inmediata al reconocer que el vector de posición y los operadores de gradiente y Laplaciano son respectivamente:

$$\vec{\rho} = \rho \left(\hat{i} \cos \varphi + \hat{j} \sin \varphi \right) \quad (17)$$

$$\vec{\nabla} = (\hat{i} \cos \varphi + \hat{j} \sin \varphi) \frac{\partial}{\partial \rho} + (-\hat{i} \sin \varphi + \hat{j} \cos \varphi) \frac{1}{\rho} \frac{\partial}{\partial \varphi} \quad (18)$$

$$\nabla^2 = \frac{1}{\rho} \frac{\partial}{\partial \rho} \rho \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \varphi^2} \quad (19)$$

Entonces el Hamiltoniano en la ecuación de Schrödinger (12) toma la forma

$$\left[-\frac{1}{2} \left(\frac{1}{\rho} \frac{\partial}{\partial \rho} \rho \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \varphi^2} \right) + \frac{1}{2} \rho^2 \right] \psi_{n_\rho m}(\rho, \varphi) = \epsilon_{n_\rho m} \psi_{n_\rho m}(\rho, \varphi) \quad (20)$$

También se identifica el operador de momento angular

$$\begin{aligned} \vec{\ell} = \vec{\rho} \times \vec{p} &= \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \rho \cos \varphi & \rho \sin \varphi & 0 \\ -i\hbar \left(\cos \varphi \frac{\partial}{\partial \rho} - \frac{\sin \varphi}{\rho} \frac{\partial}{\partial \varphi} \right) & -i\hbar \left(\sin \varphi \frac{\partial}{\partial \rho} + \frac{\cos \varphi}{\rho} \frac{\partial}{\partial \varphi} \right) & 0 \end{vmatrix} \\ &= \hat{k} (-i\hbar) \frac{\partial}{\partial \varphi}. \end{aligned} \quad (21)$$

cuyo cuadrado aparece como factor en el segundo término de la ecuación (19). La ecuación de eigenvalores

$$-i\hbar \frac{\partial}{\partial \varphi} \Phi(\varphi) = m\hbar \Phi(\varphi) \quad (22)$$

es inmediatamente integrable

$$\Phi(\varphi) = \frac{e^{im\varphi}}{\sqrt{2\pi}} \quad (23)$$

y sus eigenvalores $m = 0, \pm 1, \pm 2, \dots$ están restringidos a números enteros garantizando su unicidad $\Phi(\varphi + 2\pi) = \Phi(\varphi)$ en el intervalo $\varphi [0, 2\pi]$.

La ecuación (20) admite la solución factorizable

$$\psi_{n_\rho m}(\rho, \varphi) = R_{n_\rho m}(\rho) e^{im\varphi} \quad (24)$$

transformándose en la ecuación radial

$$\left[-\frac{1}{2} \left(\frac{1}{\rho} \frac{d}{d\rho} \rho \frac{d}{d\rho} - \frac{m^2}{\rho^2} \right) + \frac{1}{2} \rho^2 \right] R_{n_\rho, m}(\rho) = \epsilon_{n_\rho, m}(\rho). \quad (25)$$

Esta ecuación tiene singularidades en $\rho = 0$ y $\rho = \infty$ asociadas a los términos de energía cinética rotacional y de la energía potencial cuadrática. Estas singularidades se remueven separando los factores ρ^m y $e^{-\rho^2/2}$

$$R_{n_\rho, m}(\rho) = \rho^m e^{-\rho^2/2} f(\rho) \quad (26)$$

El factor restante se identifica como un polinomio de Laguerre o función hipergeométrica confluyente de Kummer [3],

$$f(\rho) = M(-n_\rho, m+1, \rho^2). \quad (27)$$

La ecuación diferencial que satisface se obtiene de la sustitución de la ecuación (26) en la ecuación (25). Aquí en vez de seguir ese camino, ilustramos como se pueden construir usando los operadores de escalera de ascenso y de descenso para ambos números cuánticos radial y angular.

El eigenvalor de la energía en la ecuación (25) resulta ser

$$E_{n_\rho m} = 2n_\rho + |m| + 1 \quad (28)$$

que coincide con el de la ecuación (13) al incluir la relación

$$2n_\rho + |m| = (n_\xi + n_\eta) \quad (29)$$

para los estados degenerados.

Los operadores de ascenso y descenso en coordenadas cartesianas de las ecuaciones (14) y (15) se pueden traducir a sus respectivas representaciones en coordenadas circulares usando las ecuaciones (17) y (18):

$$\begin{aligned} \vec{a}_\pm = \hat{i} \left[\cos \varphi \left(\rho \mp \frac{\partial}{\partial \rho} \right) \pm \sin \varphi \frac{1}{\rho} \frac{\partial}{\partial \varphi} \right] \\ + \hat{j} \left[\sin \varphi \left(\rho \mp \frac{\partial}{\partial \rho} \right) \mp \cos \varphi \frac{1}{\rho} \frac{\partial}{\partial \varphi} \right]. \end{aligned} \quad (30)$$

A continuación usamos las relaciones de Euler

$$\cos \varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}, \quad \sin \varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i}$$

para reescribirlos en términos de los vectores complejos $\hat{i} \mp \hat{j}$ y las exponenciales de $\pm i\varphi$, con el resultado

$$\vec{a}_{\pm} = \frac{1}{2} (\hat{i} - i\hat{j}) e^{i\varphi} \left(\rho \mp \frac{\partial}{\partial \rho} \mp \frac{i}{\rho} \frac{\partial}{\partial \varphi} \right) + \frac{1}{2} (\hat{i} + i\hat{j}) e^{-i\varphi} \left(\rho \mp \frac{\partial}{\partial \rho} \pm \frac{i}{\rho} \frac{\partial}{\partial \varphi} \right)$$

En lo que sigue distinguimos a componentes de los operadores respectivos, agregando un segundo subíndice asociado al signo de la exponencial angular:

$$\begin{aligned} \hat{a}_{++} &= \frac{1}{2} e^{i\varphi} \left(\rho - \frac{\partial}{\partial \rho} - \frac{i}{\rho} \frac{\partial}{\partial \varphi} \right) \\ \hat{a}_{+-} &= \frac{1}{2} e^{-i\varphi} \left(\rho - \frac{\partial}{\partial \rho} + \frac{i}{\rho} \frac{\partial}{\partial \varphi} \right) \\ \hat{a}_{-+} &= \frac{1}{2} e^{i\varphi} \left(\rho + \frac{\partial}{\partial \rho} + \frac{i}{\rho} \frac{\partial}{\partial \varphi} \right) \\ \hat{a}_{--} &= \frac{1}{2} e^{-i\varphi} \left(\rho + \frac{\partial}{\partial \rho} - \frac{i}{\rho} \frac{\partial}{\partial \varphi} \right) \end{aligned}$$

En lo general, la aplicación de estos operadores al eigenestado $e^{im\varphi}$ lo transforman en los estados vecinos $e^{i(m\pm 1)\varphi}$ según sea el segundo subíndice, y con los respectivos coeficientes $m, -m, -m$ y m para el tercer término en $1/\rho$ que resultan después de aplicar la derivada angular. De esta manera queda establecido el carácter de operadores de ascenso y descenso para la parte angular.

A continuación se ilustra la acción complementaria sobre las funciones radiales $R_{n\rho m}$ de la ecuación (24)

$$\begin{aligned} \hat{a}_{++} \psi_{n\rho m} &= \frac{1}{2} e^{i(m+1)\varphi} \left(\rho - \frac{d}{d\rho} + \frac{m}{\rho} \right) R_{n\rho m} \\ \hat{a}_{+-} \psi_{n\rho m} &= \frac{1}{2} e^{i(m-1)\varphi} \left(\rho - \frac{d}{d\rho} - \frac{m}{\rho} \right) R_{n\rho m} \\ \hat{a}_{-+} \psi_{n\rho m} &= \frac{1}{2} e^{i(m+1)\varphi} \left(\rho + \frac{d}{d\rho} - \frac{m}{\rho} \right) R_{n\rho m} \\ \hat{a}_{--} \psi_{n\rho m} &= \frac{1}{2} e^{i(m-1)\varphi} \left(\rho + \frac{d}{d\rho} + \frac{m}{\rho} \right) R_{n\rho m} \end{aligned}$$

El lector puede anticipar los diferentes efectos de cada operador sobre la misma función con base en las diferentes combinaciones de signos en los términos sucesivos.

Empezamos con la acción sobre el estado base del operador de descenso, que lo debe aniquilar:

$$\hat{a}_- R_{00} = \left(\rho + \frac{d}{d\rho} \right) R_{00} = 0,$$

y que conduce a su forma explícita

$$R_{00}(\rho) = e^{-\rho^2/2}$$

en correspondencia con sus contrapartes cartesianas.

A continuación analizamos las acciones de los operadores de ascenso sobre el estado base:

$$\begin{aligned} \hat{a}_{+\pm} \psi_{00} &= \frac{1}{2} e^{\pm i\varphi} \left(\rho - \frac{d}{d\rho} \right) R_{00} \\ &= \frac{1}{2} e^{\pm i\psi} (\rho + \rho) e^{-\rho^2/2} = \psi_{0,\pm 1} \end{aligned}$$

Aquí se aprecia la potencia común $m = 1$ de los factores angular y radial, Ec. (24) y (26), así como la degeneración doble de los estados y su correspondencia con sus contrapartes cartesianas ($n_\xi = 1, n_\eta = 0$) y ($n_\xi = 0, n_\eta = 1$).

Las acciones de los operadores de ascenso sobre los primeros estados excitados se ilustran a continuación:

$$\begin{aligned} \hat{a}_{++} \psi_{01} &= \frac{1}{2} e^{i2\varphi} \left[\rho - \frac{d}{d\rho} + \frac{1}{\rho} \right] \rho e^{-\rho^2/2} = e^{i2\varphi} \rho^2 e^{-\rho^2/2} = \psi_{0,2} \\ \hat{a}_{+-} \psi_{0,-1} &= \frac{1}{2} e^{-i2\varphi} \left[\rho - \frac{d}{d\rho} + \frac{1}{\rho} \right] \rho e^{-\rho^2/2} = e^{-i2\varphi} \rho^2 e^{-\rho^2/2} = \psi_{0,-2} \\ \hat{a}_{+-} \psi_{0,1} &= \hat{a}_{++} \psi_{0,-1} = \frac{1}{2} \left[\rho - \frac{d}{d\rho} - \frac{1}{\rho} \right] \rho e^{-\rho^2/2} = -[1 - \rho^2] e^{-\rho^2/2} = \psi_{1,0} \end{aligned}$$

El binomio en la función radial corresponde a la función hipergeométrica confluyente de Kummer de la Ec. (27) con $n_\rho = 1$ y $m = 0$.

Los tres estados obtenidos comparten la misma energía con $N = 2$ en las ecuaciones (28-29), con degeneración de orden 3. La extensión para otros estados con mayor excitación se puede implementar de manera sistemática aplicando sucesivamente los operadores de ascenso. La aplicación de los operadores de descenso conduce también a cada estado con sus vecinos de menor excitación.

3 Osciladores Armónicos en Tres Dimensiones en Coordenadas Cartesianas, Cilíndricas, Circulares y Esféricas.

En esta sección consideramos el oscilador armónico isotrópico en las coordenadas adimensionales cartesianas (ξ, η, ζ) , cilíndricas circulares (ρ, φ, ζ) y esféricas (r, θ, φ) . Empezamos con sus conexiones familiares. El vector de posición:

$$\begin{aligned}\vec{r} &= \hat{i}\xi + \hat{j}\eta + \hat{k}\zeta \\ &= (\hat{i}\cos\varphi + \hat{j}\sin\varphi)\rho + \hat{k}\zeta \\ &= (\hat{i}\cos\varphi + \hat{j}\sin\varphi)r\sin\theta + \hat{k}r\cos\theta.\end{aligned}\quad (31)$$

El operador de gradiente

$$\begin{aligned}\vec{\nabla} &= \hat{i}\frac{\partial}{\partial\xi} + \hat{j}\frac{\partial}{\partial\eta} + \hat{k}\frac{\partial}{\partial\zeta} \\ &= (\hat{i}\cos\varphi + \hat{j}\sin\varphi)\frac{\partial}{\partial\rho} + (-\hat{i}\sin\varphi + \hat{j}\cos\varphi)\frac{1}{\rho}\frac{\partial}{\partial\varphi} + \hat{k}\frac{\partial}{\partial\zeta} \\ &= \left[(\hat{i}\sin\varphi + \hat{j}\cos\varphi)\sin\theta + \hat{k}\cos\theta \right] \frac{\partial}{\partial r} \\ &+ \left[(\hat{i}\cos\varphi + \hat{j}\sin\varphi)\cos\theta - \hat{k}\sin\theta \right] \frac{1}{r}\frac{\partial}{\partial\theta} \\ &+ (-\hat{i}\sin\varphi + \hat{j}\cos\varphi)\frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi}\end{aligned}\quad (32)$$

donde los vectores unitarios $\hat{\rho}, \hat{\varphi}$ y $\hat{r}, \hat{\theta}, \hat{\varphi}$, así como los factores de escala $h_\rho = 1, h_\varphi = \rho = r\sin\theta, h_r = 1, h_\theta = r$ son identificables directamente.

Los factores de escala permiten escribir los respectivos operadores Laplacianos, y aquí simplemente los incorporamos en las ecuaciones de Schrödinger siguientes:

$$\left\{ -\frac{1}{2} \left(\frac{\partial^2}{\partial\xi^2} + \frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\zeta^2} \right) + \frac{1}{2} (\xi^2 + \eta^2 + \zeta^2) \right\} \psi(\xi, \eta, \zeta) = \epsilon_N \psi(\xi, \eta, \zeta) \quad (33)$$

$$\left\{ -\frac{1}{2} \left(\frac{1}{\rho} \frac{\partial}{\partial\rho} \rho \frac{\partial}{\partial\rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial\varphi^2} + \frac{\partial^2}{\partial\zeta^2} \right) + \frac{1}{2} (\rho^2 + \zeta^2) \right\} \psi(\rho, \varphi, \zeta) = \epsilon_N \psi(\rho, \varphi, \zeta) \quad (34)$$

$$\left\{ -\frac{1}{2} \left[\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2} \left(\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \sin\theta \frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\varphi^2} \right) \right] + \frac{1}{2} r^2 \right\} \psi(r, \theta, \varphi) = \epsilon_N \psi(r, \theta, \varphi) \quad (35)$$

La separabilidad de las ecuaciones en coordenadas cartesianas y circulares cilíndricas se extiende para incluir el tercer grado de libertad ζ con respecto

a la sección anterior. Efectivamente tenemos eigenfunciones factorizales,

$$\psi_{n_\xi n_\eta n_\zeta}(\xi, \eta, \zeta) = \psi_{n_\xi}(\xi) \psi_{n_\eta}(\eta) \psi_{n_\zeta}(\zeta) \quad (36)$$

$$\psi_{n_\rho m_\varphi n_\xi}(\rho, \varphi, \xi) = R_{n_\rho m}(\rho) \Phi_m(\varphi) \psi_{n_\xi}(\xi) \quad (37)$$

y eigenvalores de energía

$$\epsilon_{n_\xi n_\eta n_\zeta} = n_\xi + n_\eta + n_\zeta + \frac{3}{2} \quad (38)$$

$$\epsilon_{n_\rho m n_\zeta} = 2n_\rho + |m| + n_\zeta + \frac{3}{2} \quad (39)$$

con energías comunes para $n_\xi + n_\eta + n_\zeta = 2n_\rho + |m| + n_\zeta = N$.

Los operadores de escalera de ascenso y descenso

$$\vec{a}_\pm = (\vec{r} \mp \nabla) \quad (40)$$

incluyen en la dirección \hat{k} las componentes cartesianas comunes $\left(\zeta \mp \frac{\partial}{\partial \zeta}\right)$. Los efectos de aplicar los operadores de la ecuación (40) a las eigenfunciones de las ecuaciones (36,37) son del mismo tipo que los efectos ya estudiados en las secciones 1 y 2. El lector puede analizar de manera sistemática la generación del estado base y de los estados excitados usando \vec{a}_- para el estado con $N = 0$, y \vec{a}_+ para los otros estados.

En la ecuación (35) se reconoce la presencia del operador $\partial^2 / \partial \varphi^2$ en común con el término correspondiente en la ecuación (34), incluyendo su coeficiente común donde $\rho^2 = r^2 \sin^2 \theta$. En la sección 2, ya se reconoció que $-i\partial / \partial \varphi$ es el operador de la componente z del momento angular. Correspondientemente, en la ecuación (35) el operador angular dentro del paréntesis se identifica como el cuadrado de momento angular total incluyendo el signo negativo común afuera del corchete:

$$\hat{\ell}^2 = - \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right). \quad (41)$$

Los operadores $\hat{\ell}^2$ y $\hat{\ell}_z$ conmutan entre sí y tienen eigenfunciones comunes que son los armónicos esféricos

$$\hat{\ell}^2 Y_{\ell m}(\theta, \varphi) = \ell(\ell + 1) Y_{\ell m}(\theta, \varphi) \quad (42)$$

$$\hat{\ell}_z Y_{\ell m}(\theta, \varphi) = m Y_{\ell m}(\theta, \varphi) \quad (43)$$

donde

$$Y_{\ell m}(\theta, \varphi) = N_{\ell m} \sin^m \theta P_{\ell}^m(\cos \theta) e^{im\varphi} \quad (44)$$

es el producto de los polinomios asociados de Legendre y la eigenfunción de $\hat{\ell}_z$; además $\ell = 0, 1, 2, \dots$ y $m = -\ell, -\ell + 1, \dots, -1, 0, 1, 2, \dots, \ell$. El operador vectorial de momento angular

$$\begin{aligned} \vec{\ell} &= \vec{r} \times \vec{p} = \hat{i}\hat{\ell}_x + \hat{j}\hat{\ell}_y + \hat{k}\hat{\ell}_z \\ &= \frac{1}{2}(\hat{i} - i\hat{j})(\ell_x + i\ell_y) + \frac{1}{2}(\hat{i} + i\hat{j})(\ell_x - i\ell_y) + \hat{k}\hat{\ell}_z \end{aligned} \quad (45)$$

se puede escribir en términos de sus componentes cartesianas o de sus combinaciones complejas

$$\hat{\ell}_{\pm} = \hat{\ell}_x \pm i\hat{\ell}_y$$

Estos juegan el papel de operadores de ascenso y descenso para el número cuántico magnético m , conectando los $(2\ell + 1)$ armónicos esféricos con un mismo valor de ℓ_z [1,2], específicamente

$$\hat{L}_{\pm} Y_{\ell m} = \sqrt{(L \mp m)(\ell \pm m + 1)} Y_{\ell, m \pm 1}(\theta, \varphi) \quad (46)$$

Las soluciones de la ecuación (35) son separables

$$\psi_{n_r \ell m}(r, \theta, \varphi) = R_{n_r \ell}(r) Y_{\ell m}(\theta, \varphi) \quad (47)$$

y su factor radial satisface la ecuación

$$\left\{ -\frac{1}{2} \left[\frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} - \frac{\ell(\ell+1)}{r^2} \right] + \frac{1}{2} r^2 \right\} R_{n_r \ell}(r) = \epsilon_N R_{n_r \ell}(r) \quad (48)$$

Nótese la estructura semejante con su contraparte circular, ecuación (20). Los puntos $r = 0$ y $r = \infty$ son también singulares y conducen a soluciones del tipo

$$R_{n_r \ell}(r) = r^{\ell} e^{-r^2/2} f(r)$$

donde $f(r)$ se identifica como la función hipergeométrica confluyente de Kummer $M(-n_r, \ell + \frac{3}{2}, r^2)$. Los eigenvalores de la energía son

$$\epsilon_{n_r \ell} = \left(2n_r + \ell + \frac{3}{2} \right) = \left(N + \frac{3}{2} \right)$$

como la contraparte de las ecuaciones(38-39).

Los operadores de ascenso y descenso en coordenadas esféricas se pueden escribir en base a las Ecs. (40), (31), (32) usando directamente las expresiones para el vector de posición y el vector de gradiente. Aquí seguimos un camino diferente, aprovechando la conexión entre el momento lineal y el momento angular. Efectivamente, calculando el producto vectorial

$$\hat{r} \times \vec{L} = \hat{r} \times (\vec{r} \times \vec{P}) = \vec{r} (\hat{r} \cdot \vec{P}) - (\hat{r} \cdot \vec{r}) \vec{P} \quad (49)$$

se identifican las componentes radial y transversales del vector de cantidad de movimiento y de gradiente

$$\vec{P} = \hat{r} P_r - \frac{\hat{r} \times \vec{L}}{r} = -i\hbar \nabla. \quad (50)$$

La ventaja de este camino es que podemos aprovechar las acciones de los operadores de momento angular sobre los armónicos esféricos, Ecs. (43) y (46).

Las expresiones para los operadores escalera en la base de Euler son las siguientes:

$$\begin{aligned} \vec{a}_{\pm} = & \frac{1}{2} (\hat{i} - \hat{j}) \left\{ \text{sen } \theta e^{ir} \left[\left(r \mp \frac{\partial}{\partial r} \right) \pm \frac{1}{r} L_z \right] \mp \frac{\cos \theta}{r} L_+ \right\} \\ & + \frac{1}{2} (\hat{i} + \hat{j}) \left\{ \text{sen } \theta e^{-ir} \left[\left(r \mp \frac{\partial}{\partial r} \right) \mp \frac{1}{r} L_z \right] \pm \frac{\cos \theta}{r} L_- \right\} \\ & + \hat{k} \left\{ \cos \theta \left(r \mp \frac{\partial}{\partial r} \right) \mp \frac{\text{sen } \theta}{2r} [-e^{i\varphi} L_- + e^{-i\varphi} L_+] \right\} \end{aligned} \quad (51)$$

Nótese que para $\theta = \frac{\pi}{2}$; r se convierte en ρ , y los primeros términos dentro de las llaves en los dos primeros renglones se convierten en los operadores respectivos en coordenadas circulares; mientras los segundos términos se anulan. Por otra parte, para $\theta = 0$ o π : r se convierte en z y el primer término dentro de las llaves en el tercer renglón se convierte en los operadores respectivos a lo largo del eje z , mientras que el segundo término se anula. En general, los términos en L_+ y L_- son los novedosos y a continuación identificamos sus efectos en los renglones sucesivos: En el primer renglón L_+ aplicado a un armónico esférico ℓm , lo convierte en $\ell, m + 1$, que es el

mismo efecto de sus términos compañeros con el factor común $e^{i\varphi}$; en el segundo renglón también se aprecia el efecto de L_- sobre ℓm , convirtiéndolo en $\ell, m - 1$, también compartido con sus compañeros con el factor $e^{-i\varphi}$; en el tercer renglón los efectos de $e^{-i\varphi}L_+$ y $e^{-i\varphi}L_-$ son comunes dejando ℓm en ℓm . En resumen y de acuerdo con las Ecs. (46) y (43), en los dos primeros renglones se tienen los efectos de subir y bajar m en una unidad, $\Delta m = \pm 1$, respectivamente, y en el tercero $\Delta m = 0$.

Adicionalmente, se reconoce la presencia de las componentes del vector unitario \hat{r} en la base de Euler $\sin\theta e^{\pm i\varphi}$ y $\cos\theta$ que corresponden a los armónicos esféricos con $\ell = 1$ y $m' = \pm 1, 0$. El acoplamiento vectorial de los momentos angulares $\vec{\ell} + \vec{1} = \vec{L}$ conduce a los estados con $L = \ell + 1$ y $\ell - 1$, y $M = m + m'$

$$Y_{LM}(\theta, \varphi) = \sum_{m, m'} Y_{\ell m}(\theta, \varphi) Y_{1 m'}(\theta, \varphi) c_{LM}^{\ell m / m'}$$

mostrando los efectos de ascenso y descenso para la magnitud del momento angular total, siendo $c_{LM}^{\ell m / m'}$ los coeficientes de Clebsch-Gordan. Usando la notación $a_{\pm\Delta\ell, \Delta m'}$ para los operadores de ascenso y descenso en los diferentes grados de libertad, identificamos sus formas explícitas y sus acciones

$$\begin{aligned} a_{++\Delta m'} \psi_{n_r \ell m}(r, \theta, \varphi) &= Y_{\ell+1, M}(\theta, \varphi) \left[r - \frac{d}{dr} + \frac{\ell}{r} \right] R_{n_r \ell}(r) \\ a_{+-\Delta m'} \psi_{n_r \ell m}(r, \theta, \varphi) &= Y_{\ell-1, M}(\theta, \varphi) \left[r - \frac{d}{dr} - \frac{\ell+1}{r} \right] R_{n_r \ell}(r) \\ a_{-+\Delta m'} \psi_{n_r \ell m}(r, \theta, \varphi) &= Y_{\ell+1, M}(\theta, \varphi) \left[r + \frac{d}{dr} - \frac{\ell}{r} \right] R_{n_r \ell}(r) \\ a_{--\Delta m'} \psi_{n_r \ell m}(r, \theta, \varphi) &= Y_{\ell-1, M}(\theta, \varphi) \left[r + \frac{d}{dr} + \frac{\ell+1}{r} \right] R_{n_r \ell}(r) \end{aligned}$$

Nótese que los coeficientes de los términos en $1/r$ dentro de los operadores de ascenso y descenso radiales solo dependen del número cuántico ℓ y no dependen de las proyecciones m, m', M , debido al carácter central del potencial de oscilador armónico isotrópico. Nótese también la misma estructura de estos operadores con los de sus contrapartes bidimensionales, con las diferencias en los coeficientes asociados a los respectivos momentos angulares. Se invita al lector a construir el estado base $\psi_{000} = Y_{00}R_{00}$ a partir de su

aniquilación con el operador del tercer renglón, y los demás por aplicaciones sucesivas de los operadores del primero y segundo renglones.

$$\left[r + \frac{d}{dr} \right] R_{00} = 0 \quad \therefore \quad R_{00} = e^{-r^2/2}$$

$$a_{--\Delta m'} \psi_{000} = Y_{1M}(\theta, \varphi) \left[r - \frac{d}{dr} \right] e^{-r^2/2} = Y_{1M}(r+r) e^{-r^2/2} = \psi_{01M}(r)$$

que corresponden a los tres estados degenerados con $M = 1, 0, -1$, y $N = 1$.

$$\begin{aligned} a_{++\Delta m'} \psi_{01M} &= Y_{2M} \left[r - \frac{d}{dr} + \frac{1}{r} \right] r e^{-r^2/2} \\ &= Y_{2M} [r^2 - 1 + r^2 + 1] e^{-r^2/2} = \psi_{02M} \\ a_{+-\Delta m'} \psi_{01M} &= Y_{00} \left[r - \frac{d}{dr} - \frac{2}{r} \right] r e^{-r^2/2} \\ &= [r^2 - 1 + r^2 - 2] e^{-r^2/2} = -3 \left[1 - \frac{r^2}{3} \right] e^{-r^2/2} = \psi_{100} \end{aligned}$$

que corresponden a los $5 + 1 = 6$ estados degenerados con $N = 2$, y a identificar $M(-1, 3/2, r^2)$ como el binomio en el último renglón.

4 Discusión

En las secciones 1, 2 y 3 se han presentado los operadores de ascenso y descenso para el oscilador armónico en las respectivas dimensiones y sistemas de coordenadas, cartesianas, circulares y esféricas. Se han identificado las conexiones entre los operadores en las diferentes coordenadas, y se han ilustrado sus acciones sobre las eigenfunciones respectivas. En las secciones 2 y 3 se destacan los papeles de los operadores de momento angular como partes de los operadores completos. En este punto llamamos la atención de los lectores a las referencias adicionales [3] y [5].

Por otra parte, las referencias [6-7], presentan la construcción de los operadores correspondientes para la construcción de los operadores correspondientes para la partícula libre confinada en cajas lineales, circulares y esféricas. Además en [5] se muestra que tales operadores son válidos para la partícula libre en general. El uso del oscilador armónico se reduce al de la partícula

libre en el límite de $\omega \rightarrow 0$ en que la acción de la fuerza desaparece, y los operadores respectivos se reducen a las componentes del operador de gradiente o de cantidad de movimiento. En la referencia [4] se encuentran operadores escalera para osciladores armónicos y el átomo de hidrógeno confinados en ángulos diedros, y en la referencia [10] se construyeron tres conjuntos de operadores escalera para los armónicos de moléculas asimétricas.

References

1. S. Gasiorowicz, Quantum Physics, Third Edition 2003, Wiley.
2. S.H. Dong, Factorization Method in Quantum Mechanics, 2007, Springer.
3. M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, 1972, Dover.
4. Q.H. Liu, D.M. Xun and L. Shan, Raising and Lowering Operators for Orbital Momentum Quantum Numbers, Int. J. Theor. Phys. **49** (2010), 2164.
5. R. Méndez-Fragoso and E. Ley-Koo, Rotations of the Hydrogen Atom and Asymmetric Molecules in Free and Confined Configurations, Advances in Quantum Chemistry, Vol. **62** (2011), 137-212.
6. S.H. Dong and Z.Q. Ma, The Hidden Symmetry for a Quantum System with an Infinitely Deep Square-Well Potential, Am. J. Phys. **70** (2002), 520.
7. G.H. Sun and S.H. Dong, New Type Shift Operators for Circular Well Potential in Two Dimensions, Phys. Lett. A **374** (2010), 4112.
8. G.H. Sun and S.H. Dong, New Type Shift Operator for Three-Dimensional Well Potential, Modern Phys. Lett. A **26** (2011), 351.
9. E. Ley Koo and G.H. Sun, Ladder Operators for Quantum Systems Confined by Dihedral Angles, SIGMA **8** (2012), 060.
10. R. Méndez-Fragoso and E. Ley-Koo, Ladder Operators for Lamé Sphericoonal Harmonic Polynomials, SIGMA **8** (2012) 074.

Migración Planetaria

Frédéric S. Masset

December 4, 2012

1 Abstract

La migración planetaria es el proceso mediante el cual un protoplaneta, todavía rodeado por el disco a partir del cual se formó, ve variar su semi-eje mayor con el tiempo, a raíz de una interacción de marea con el disco. Este proceso es sumamente eficaz y rápido para planetas de hasta 10 masas terrestres, y se hace en general hacia dentro, trayendo los planetas a órbitas de radio y periodo muy pequeños. La cantidad clave para determinar la velocidad y dirección de migración de un protoplaneta es la torca de marea entre el planeta y el disco. Esa última tiene dos componentes: la torca de Lindblad, y la torca de corotación. Describimos en este trabajo las características de ambas, haciendo énfasis en los desarrollos de estos últimos años.

2 Introduction

The extraordinary diversity of extrasolar planetary systems has challenged our understanding of how planets form and how their orbits evolve as they form. Among the many processes contemplated thus far to account for the observed properties of extrasolar planets, the gravitational interaction between planets and their parent protoplanetary disc plays a prominent role.

Even if there is compelling evidence that other processes are capable of altering the orbits, the disk's tidal torque is in principle capable of varying the planets semi-major axis by orders of magnitude over the disk lifetime.

The main difficulty in planetary migration theories lies in predicting the torque value. In addition to being a difference between several large ampli-

tude torques, it is very sensitive to the disc properties near the planet’s orbit (e.g., density, temperature profiles).

The torque acting on a low-mass planet in circular orbit can be decomposed into two components: (i) the differential Lindblad torque, arising from material passing by the planet at supersonic velocities, which is deflected by the latter and therefore exchanges angular momentum and energy with the planet, and (ii) the corotation torque, arising from material slowly drifting with respect to the planet, in the vicinity of its orbit. The differential Lindblad torque has been extensively studied from the early times of planetary migration theories. The corotation region, which has been under intense scrutiny over the last five years, has proved to have a much more complex dynamics than previously thought. In particular, the value of the corotation torque depends sensitively on the radiative properties of the gas disc, and may exhibit large values when the gas is radiatively inefficient (as is generally expected in regions of planet formation).

3 Physical model and notations

In most of the following we shall consider two-dimensional discs, considering vertically averaged or vertically integrated quantities where appropriate. At the present time, most of the recent investigation on the migration of low-mass planets has been undertaken in two dimensions (with a few exceptions) and much insight can be gained into the mechanisms of the different components of the torque exerted by the disc on the planet through a two-dimensional analysis.

We consider a planet of mass M_p orbiting a star of mass M_\star with orbital frequency Ω_p . We denote by $q = M_p/M_\star$ the planet-to-star mass ratio. The planet is assumed to be on a prograde circular orbit of semi-major axis a , coplanar with the disc, so that we do not consider in this work eccentric, inclined, or retrograde planets. The protoplanetary disc in which the planet is embedded is modelled as a two-dimensional viscous disc in radial equilibrium, with the centrifugal acceleration and the radial acceleration related to the pressure gradient balancing the gravitational acceleration due to the central star. We assume in most of what follows that the surface density Σ and temperature T profiles are power laws of radius, with indices α and β , respectively:

$$\Sigma \propto r^{-\alpha} \tag{1}$$

and

$$T \propto r^{-\beta}. \quad (2)$$

The disc pressure scale length is $H = c_s/\Omega$, with Ω the gas orbital frequency and c_s the sound speed. We define the disc aspect ratio by $h = H/r$. When T is a power law of radius, so is h , with an index f dubbed the flaring index:

$$h \propto r^f, \quad (3)$$

which satisfies $\beta = 1 - 2f$. In almost all studies of planet–disc interactions, the disc is modelled with a stationary kinematic viscosity ν , aimed at modelling the disc’s turbulent properties.

The governing equations of the flow are the equation of continuity, the Navier-Stokes equations and the energy equation (except when dealing with isothermal discs), together with the closure relationship provided by the equation of state, which is that of an ideal gas. These read respectively:

The equation of continuity reads, in the frame corotating with the planet:

$$\partial_t \Sigma + \frac{1}{r} \partial_r (\Sigma r v_r) + \frac{1}{r} \partial_\phi (\Sigma v_\phi) = 0, \quad (4)$$

where v_r and v_ϕ are respectively the radial and azimuthal velocities in the frame corotating with the planet. The Euler equations read, respectively in r and ϕ :

$$\partial_t v_r + v_r \partial_r v_r + \frac{v_\phi}{r} \partial_\phi v_r - r \Omega_p^2 - 2 \Omega_p v_\phi - \frac{v_\phi^2}{r} = -\frac{\partial_r P}{\Sigma} - \partial_r \Phi \quad (5)$$

and

$$D_t j = -\frac{\partial_\phi p}{\Sigma} - \partial_\phi \Phi, \quad (6)$$

where $D_t \equiv \partial_t + v_r \partial_r + \frac{v_\phi}{r} \partial_\phi$, $j = r^2 \Omega$ is the specific angular momentum, and Φ is the gravitational potential, that includes the potential of the star, the potential of the planet and, in the usual cases in which the frame is not centered on the center of mass of the system but rather on the star, a so-called indirect term to account for the fictitious force originating from the frame acceleration. The equation of state that we adopt is that of ideal gases, which reads $P = \Sigma T$. In that case, the internal energy density is $e = p/(\gamma - 1)$, and the energy equation reads:

$$\Sigma D_t \left(\frac{e}{\Sigma} \right) = -p \vec{\nabla} \cdot \vec{v}. \quad (7)$$

In case one wants to add some degree of realism, this latter equation may include source and sink terms, such as the cooling of the disk by radiation through its photospheres or a viscous heating term, and heat diffusion.

4 Migration of low-mass embedded planets: Type I migration

4.1 Differential Lindblad torque

The differential Lindblad torque accounts for the exchange of angular momentum between the planet and the trailing density waves (spiral wakes) that it generates in the disc. The density waves propagating inside the planet's orbit carry away negative angular momentum, and thus exert a positive torque on the planet, named the inner Lindblad torque. Similarly, the spiral density waves propagating outside the planet's orbit carry away positive angular momentum, which corresponds to a negative torque on the planet (the outer Lindblad torque). The residual torque, called differential Lindblad torque, results from a balance between the inner and outer torques.

It may be shown that the Lindblad torque scales with the following quantity:

$$\Gamma_{\text{ref}} = \Sigma q^2 \Omega^2 a^4 / h^2. \quad (8)$$

An accurate determination of the asymmetry between the inner and outer torques yields a dimensionless factor to be put in front of Γ_{ref} to give the expression for the differential Lindblad torque. This issue has triggered a lot of theoretical efforts in the last three decades, and it is not completely solved yet. Presently, the two main results are:

- An expression obtained by solving numerically the linearised equations of the flow in two dimensions with a softened planet potential, in discs with arbitrary gradients of surface density and temperature [1, 2]. It takes the form:

$$\frac{\Gamma_L}{\Gamma_{\text{ref}}} = -(2.5 + 1.7\beta - 0.1\alpha) \left(\frac{0.4}{\epsilon/H} \right)^{0.71}, \quad (9)$$

with α and β defined in Eqs. (1) and (2). The expression in Eq. (9) is most accurate for softening lengths $\epsilon \sim 0.4H$.

- An expression obtained by a semi-analytic method for globally isothermal, three-dimensional discs with arbitrary gradients of surface density [3]:

$$\frac{\Gamma_L}{\Gamma_{\text{ref}}} = -(2.34 - 0.10\alpha). \quad (10)$$

We sum up the following important results

- The differential Lindblad torque corresponds to the net rate of angular momentum carried away by density waves (wakes) the planet generates in the disc at Lindblad resonances.
- The sign and magnitude of the differential Lindblad torque arise from a slight asymmetry in the perturbed density distribution associated to each wake.
- The differential Lindblad torque is a stationary quantity, largely independent of the disc's turbulent viscosity. Alone, it would drive the migration of Earth mass embedded planets in as short a time as a few $\times 10^5$ yrs in typical protoplanetary discs.

4.2 Corotation torque

The other component of the tidal torque, the corotation torque, has long been neglected in studies of planetary migration. It is usually linked to the so-called horseshoe drag, which corresponds to the exchange of angular momentum between the planet and its horseshoe region. The planet's horseshoe region encompasses the disc region where fluid elements are on horseshoe streamlines with respect to the planet orbit. It has been established in the past few years that this torque component features three terms: one which scales with the radial gradient of specific vorticity at the orbit (i.e. it scales with $3/2 - \alpha$), one which scales with the radial gradient of entropy at the orbit, and another one which scales with the radial gradient of temperature at the orbit.

The former has the following expression:

$$\Gamma_{\text{HS}}^{\text{vort}} = 0.93 \left(\frac{3}{2} - \alpha \right) \Sigma \Omega_p^2 q^2 a^4 h^{-2}, \quad (11)$$

whereas the second component reads:

$$\Gamma_{\text{HS}}^{\text{entr}} = -\frac{1.3\mathcal{S}}{\epsilon/H}\Sigma\Omega_p^2q^2a^4h^{-2}, \quad (12)$$

where the above expression has been derived in the framework of a flat temperature profile, and assuming a ratio of specific heats $\gamma = 1.4$, and where ϵ is the softening length of the potential. In Eq. (12), all disc quantities are to be evaluated at the planet’s orbital radius, and the quantity \mathcal{S} is a dimensionless expression of the radial entropy gradient. Finally, the dependence of the corotation torque on the temperature gradient is still to be determined, and is the subject of current work.

5 Migration of gap-opening planets: Type II and III migration

The disc response to a low-mass planet has been studied in details in § 4, where we have focused on the two components of the type I migration torque. The aim of this section is to examine the range of planet masses that is relevant to type I migration in § 5.1, and to give a concise description of planet–disc interactions for planets that are massive enough to significantly perturb the disc’s mass distribution.

5.1 Shock formation and gap-opening criterion

The wakes generated by a planet in a disc carry angular momentum as they propagate away from the planet. This angular momentum is eventually deposited in the disc through some wave damping processes, which leads to redistributing the disc mass. An efficient wave damping mechanism relies on the non-linear wave evolution of the wakes into shocks [4]. The (negative) angular momentum deposited by the inner wake decreases the semi-major axis of the fluid elements in the disc region inside the planet’s orbit (the inner disc). Similarly, the (positive) angular momentum deposited by the outer wake increases the semi-major axis of the fluid elements in the outer disc.

The distance d_s from the planet where the planet-generated wakes become

shocks is given by [4, 5]:

$$d_s \approx 0.8 \left(\frac{\gamma + 1}{12/5} \frac{q}{h^3} \right)^{-2/5} H(a), \quad (13)$$

where γ is the gas adiabatic index, and a denotes the planet's semi-major axis. As the magnitude of the one-sided Lindblad torque peaks at $\sim 4H(a)/3$ from the planet's orbit, a linear description of the differential Lindblad torque thus fails when $|d_s| \sim 4H(a)/3$. This condition can be recast as $q \leq 0.3h^3$ for $\gamma = 5/3$. When $|d_s| \leq 2H(a)/3$, wakes turn into shocks within their excitation region. Fluid elements just outside the planet's horseshoe region are pushed away from the planet orbit after crossing the wakes, which directly affects the planet's coorbital region by inducing asymmetric U-turns [6]. Horseshoe fluid elements therefore get progressively repelled from the planet orbit after each U-turn, and the planet slowly depletes its coorbital region. The equilibrium structure (width, depth) of the annular gap the planet forms around its orbit is determined by a balance between gravity, viscous and pressure torques [7].

Shock formation and its damping efficiency are very sensitive to the disc's viscosity, and gap-opening results from a balance between (i) a planet mass large enough to induce shocks where the wake excitation takes place, and (ii) a viscosity small enough to maximise the amount of angular momentum deposited by the shocks in the planet's immediate vicinity:

1. The first condition reads $|d_s| \leq 2H(a)/3$, which corresponds to $q \geq 1.5h^3$ for $\gamma = 5/3$. It means that the planet's Bondi radius $r_B = GM_p/c_s^2$, where the pressure distribution is most strongly perturbed by the planet, as well the planet's Hill radius become comparable to the local pressure scaleheight. This is known as the thermal criterion for gap opening [8].
2. The second condition, known as the viscous criterion, can be expressed as $q \geq 40/\mathcal{R}$, where $\mathcal{R} = a_p^2\Omega_p/\nu$ is the Reynolds number¹ [9, 10].

¹Although traditionally dubbed Reynolds number essentially for dimensional considerations, this ratio has little to do with the dimensionless ratio that must be considered to assess whether a flow is laminar or turbulent. If one regards the planet as an obstacle in the sheared Keplerian flow, it would be more appropriate to consider as a characteristic scale the size of its Roche lobe or $\sim x_s$, and as a characteristic velocity $2|A|x_s$.

The above two conditions for gap-opening have been revisited by [7], who provide a unified criterion that takes the form

$$1.1 \left(\frac{q}{h^3} \right)^{-1/3} + \frac{50\alpha_v h^2}{q} \leq 1, \quad (14)$$

where we have written the disc's kinematic viscosity $\nu = \alpha_v h^2 a^2 \Omega$ [11], and where in Eq. (14) h and α_v are to be evaluated at the planet's semi-major axis.

5.2 Partial gap-opening: type III migration in massive discs

So far, we have addressed the properties of planet migration through a direct analysis of the tidal torque, the latter being directly proportional to the migration rate. This approach is valid for low-mass planets that do not open a gap, for which migration has a negligible feedback on the tidal torque (note that a weak, negative feedback slightly decreases the magnitude of the entropy-related horseshoe drag [12]). Nevertheless, migrating planets that open a partial gap around their orbit experience an additional corotation torque due to fluid elements flowing across the horseshoe region [13]. If for instance the planet migrates inwards, fluid elements circulating near the inner separatrix of the horseshoe region enter the horseshoe region, and execute an outward U-turn when they reach the vicinity of the planet. Upon completion of the U-turn, these fluid elements leave the horseshoe region as the planet keeps migrating, and end up circulating in the outer disc. Consequently, the mass distribution within the horseshoe region may become asymmetric, as the horseshoe region adopts approximately a trapezoidal shape in the azimuth-radius plane [13]. As a consequence, in the case of an inward migrating planet, there is more mass behind the planet than ahead of it, owing to the partial depletion of the asymmetric horseshoe region. Similarly, if the planet migrates outwards, fluid elements circulating near the outer separatrix may embark on single inward U-turns across the horseshoe region.

Assuming steady migration at a moderate rate (this point will be clarified below), the additional corotation torque experienced by the planet due to the orbit-crossing flow is, to lowest order in x_s/a ,

$$\Gamma_{\text{cross}} = 2\pi a \dot{\Sigma}_s \times \Omega a x_s, \quad (15)$$

where Σ_s is the surface density at the inner (outer) horseshoe separatrix for a planet migrating inwards (outwards). The term $2\pi a\dot{a}\Sigma_s$ in the right-hand side of Eq. (15) is the mass flux across the horseshoe region. The second term (Ωax_s) is the amount of specific angular momentum that a fluid element near a horseshoe separatrix exchanges with the planet when performing a horseshoe U-turn. Note that the above expression for Γ_{cross} assumes that all circulating fluid elements entering the coorbital region embark on horseshoe U-turns, whereas a fraction of them may actually become trapped inside the planet's circumplanetary disc. Since Γ_{cross} is proportional to, and has same sign as \dot{a} , migration may become a runaway process. We now discuss under which circumstances a runaway may happen.

The planet and its coorbital material (which encompasses the horseshoe region, with mass M_{hs} , and the circumplanetary disc, with mass M_{cpd}) migrate at the same drift rate, \dot{a} , which we assume to be constant. The rate of angular momentum change of the planet and its coorbital region includes (i) the above contribution Γ_{cross} to the corotation torque, and (ii) the tidal torque that, for planets opening a partial gap, essentially reduces to the differential Lindblad torque Γ_{LR} :

$$\frac{\Omega}{2}a\dot{a}(M_p + M_{\text{cpd}} + M_{\text{hs}}) = 2\pi a\dot{a}\Sigma_s \times \Omega ax_s + \Gamma_{\text{LR}}. \quad (16)$$

Eq. (16) can be written as

$$\frac{\Omega}{2}a\dot{a}\tilde{M}_p = \frac{\Omega}{2}a\dot{a}\delta m + \Gamma_{\text{LR}}, \quad (17)$$

where $\tilde{M}_p = M_p + M_{\text{cpd}}$ corresponds to an *effective* planet mass, and where $\delta m = 4\pi ax_s\Sigma_s - M_{\text{hs}}$ is called the coorbital mass deficit [13]. It represents the difference between (i) the mass the horseshoe region would have if it had a uniform surface density equal to that of the separatrix-crossing flow, and (ii) the actual horseshoe region mass.

Runaway migration is based on the planet's ability to build up a coorbital mass deficit by opening a gap. It does not apply to low-mass planets, for which $\delta m \ll \tilde{M}_p$. It does not apply to high-mass planets neither, which open a wide, deep gap, so that the surface density of the separatrix-crossing flow is too small to produce a significant mass deficit. It rather concerns intermediate-mass planets, marginally satisfying the gap-opening criterion in Eq. (14), in massive discs (the larger the disc mass, the larger the density

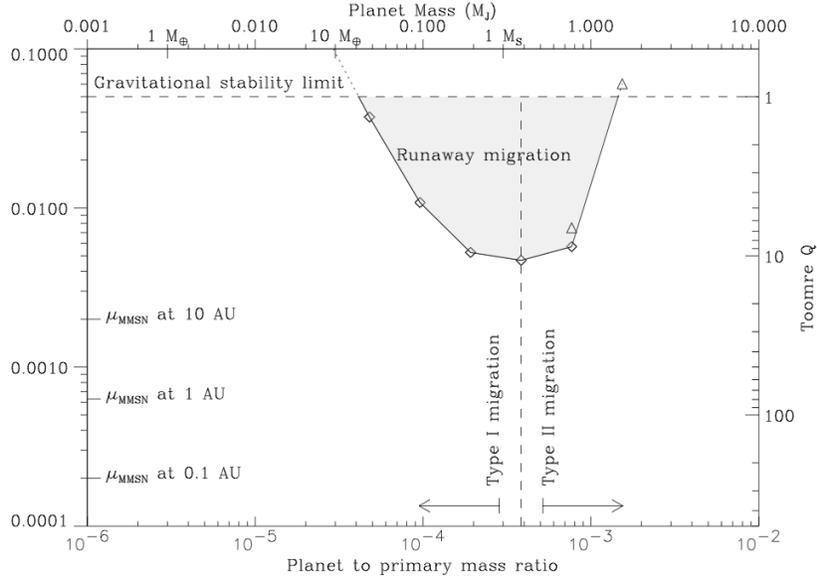


Figure 1: Occurrence for type I, II and III (runaway) migrations with varying the planet-to-star mass ratio (bottom x -axis) and the disc-to-primary mass ratio at the planet location (left y -axis). The disc’s aspect ratio is $h = 0.05$ and its alpha viscosity is $\alpha_v = 4 \times 10^{-3}$. The right y -axis shows the Toomre Q -parameter at the planet location. The upper part of the plot is limited by the gravitational instability limit (dashed line). From [13].

of the orbit-crossing flow). Its occurrence is illustrated in Fig. 1 for a disc with aspect ratio $h = 5\%$ and alpha viscosity $\alpha_v = 4 \times 10^{-3}$, where we see that runaway migration may be particularly relevant to Saturn-mass planets in massive discs (with a Toomre- Q parameter at the planet’s orbital radius typically less than about 10).

The simple model described above helps understand the condition for migration to enter a runaway regime. However, since it assumes steady migration (that is, constant \dot{a}), this model is no longer valid when migration actually enters the runaway regime, where the migration rate increases exponentially over a time comparable to the horseshoe libration period. A more general approach can be found in [13, 6, 14]. Numerical simulations find that, depending on the resolution of the gas flow surrounding the planet, the timescale for inward runaway type III migration can be as short as a few 10^2

orbits [13, 15].

The sign of Γ_{cross} is dictated by the initial drift of the planet. Runaway migration can therefore be directed inwards or outwards, depending on the sign of \dot{a} before the runaway takes place. In particular, migration may be directed outwards.

6 Conclusions

We have reviewed the recent progress made in understanding planet–disc interactions, and the resulting planets’ orbital migration. Being for a while the second-place actor of planet migration theories, the corotation torque has been shown to play a prominent role in realistic protoplanetary discs, where it can slow down, stall, or reverse type I migration. The type II and type III migration regimes for gap-opening planets are also examined. Being aimed at migration of planets on circular orbits, this presentation has set aside interesting recent developments on the tidal interactions of eccentric or inclined planets with their discs. We have also focused essentially on the mechanisms that drive the migration of a single planet in a disc, and we have therefore excluded most results about the migration of several planets. For a recent review covering these topics, the reader is referred to [16]. This list of restrictions of the present review stresses that the research on planet–disc interactions is a very active branch of planet formation, with a growing body of avenues. We finally make a plea in favor of planetary migration: planetary migration is not overrated. The tremendous value of each of the tidal torque components exerted on a given planet, associated to the great sensitivity of these torques to the underlying disc structure, appeals for a detailed knowledge of the properties of protoplanetary discs, and significant efforts toward an accurate determination of each torque component. This also reasserts tidal interactions as a prominent process in shaping forming planetary systems.

References

- [1] S.J. Paardekooper, J.C.B. Papaloizou, *A&A***485**, 877 (2008). DOI 10.1051/0004-6361:20078702

- [2] S. Paardekooper, C. Baruteau, A. Crida, W. Kley, *MNRAS***401**, 1950 (2010). DOI 10.1111/j.1365-2966.2009.15782.x
- [3] H. Tanaka, T. Takeuchi, W.R. Ward, *ApJ***565**, 1257 (2002)
- [4] J. Goodman, R.R. Rafikov, *ApJ***552**, 793 (2001)
- [5] R. Dong, R.R. Rafikov, J.M. Stone, *ApJ***741**, 57 (2011). DOI 10.1088/0004-637X/741/1/57
- [6] F.S. Masset, in *EAS Publications Series, EAS Publications Series*, vol. 29 (2008), *EAS Publications Series*, vol. 29, pp. 165–244. DOI 10.1051/eas:0829006
- [7] A. Crida, A. Morbidelli, F. Masset, *Icarus* **181**, 587 (2006). DOI 10.1016/j.icarus.2005.10.007
- [8] D.N.C. Lin, J.C.B. Papaloizou, in *Protostars and Planets III*, ed. by E.H. Levy, J.I. Lunine (1993), pp. 749–835
- [9] D.N.C. Lin, J. Papaloizou, *MNRAS***186**, 799 (1979)
- [10] G. Bryden, X. Chen, D.N.C. Lin, R.P. Nelson, J.C.B. Papaloizou, *ApJ***514**, 344 (1999). DOI 10.1086/306917
- [11] N.I. Shakura, R.A. Sunyaev, *A&A***24**, 337 (1973)
- [12] F.S. Masset, J. Casoli, *ApJ***703**, 857 (2009). DOI 10.1088/0004-637X/703/1/857
- [13] F.S. Masset, J.C.B. Papaloizou, *ApJ***588**, 494 (2003)
- [14] A. Pepliński, P. Artymowicz, G. Mellema, *MNRAS***386**, 179 (2008). DOI 10.1111/j.1365-2966.2008.13046.x
- [15] A. Crida, C. Baruteau, W. Kley, F. Masset, *A&A***502**, 679 (2009). DOI 10.1051/0004-6361/200811608
- [16] W. Kley, R.P. Nelson, *ARA&A***50**, 211 (2012). DOI 10.1146/annurev-astro-081811-125523

Módulos estructurales y precursores de fibras: anticuerpos y amiloidosis

Nina Pastor
Facultad de Ciencias, UAEM
nina@uaem.mx

Los anticuerpos son proteínas que circulan en la sangre y se encargan de unirse a moléculas que nos son ajenas, llamadas antígenos. Como se podrán imaginar, el número de antígenos es inmenso, tan grande que es imposible tener prefabricados los anticuerpos para lidiar con todos. La solución biológica a este problema es armar anticuerpos de forma modular, haciendo combinaciones al azar y luego mejorando las adivinanzas buenas. Los módulos son de dos tipos: variables y constantes; los constantes indican la clase de anticuerpo (los más comunes son tipo gamma, de ahí que también se les diga gammaglobulinas a los anticuerpos). Los variables son los que se unen a los antígenos. Cada anticuerpo está hecho de cuatro proteínas, dos cadenas pesadas iguales (llamadas así porque tienen cuatro o cinco módulos, uno de los cuales es variable) y dos ligeras iguales (que tienen dos módulos, uno variable y otro constante). Los anticuerpos son moléculas en forma de Y; en la punta de los brazos de la Y tienen dos sitios idénticos de unión con los antígenos, formados por la juxtaposición del dominio variable de una cadena ligera con el de una cadena pesada. Existen cinco clases (con sus subclases) de cadenas pesadas, dos clases (con ~70 subclases) de cadenas ligeras, combinados de forma aleatoria. Además existe un proceso llamado maduración, en el que un anticuerpo funcional es mutado de forma aleatoria, con la intención de seleccionar variantes más afines por el antígeno. Cada uno de nosotros hace su propio repertorio de anticuerpos, incluso ante el reto con el mismo antígeno.

Uno pensaría que las células B, que son las encargadas de sintetizar a los anticuerpos, hacen cantidades equimolares de cadenas ligeras y pesadas. Se ha visto que siempre hacemos un pequeño exceso de cadenas ligeras, que son secretadas y circulan por la sangre en forma de dímeros. ¿Qué hacemos con estos dímeros? En condiciones normales, llegan al riñón donde son filtrados por el glomérulo y pasan al túbulo proximal. Ahí son capturados por células que los degradan hasta aminoácidos; los que escapan son excretados en la orina. En ciertos tipos de disfunción de células B se producen muchas cadenas ligeras. Existen al menos cuatro enfermedades derivadas del manejo inadecuado por parte del riñón de los dímeros de cadenas ligeras, dependiendo de la región del riñón en la que se atoren. Afortunadamente son enfermedades poco comunes. Si se quedan en el glomérulo pueden generar depósitos amorfos (desencadenando la enfermedad por deposición de cadenas ligeras, o LCDD) o depósitos de fibras amiloides (generando amiloidosis de cadena ligera, o AL). Si se acumulan en el túbulo proximal, forman cristales característicos de la enfermedad de Fanconi; si llegan al túbulo distal se agregan con otras proteínas para formar estructuras en forma de moldes cilíndricos que bloquean el túbulo. Este tipo de bloqueos es común en mieloma múltiple y se conoce como nefropatía de moldes (cast nephropathy, o CN). Una vez que el riñón falla, el exceso de cadenas ligeras permanece en circulación, y puede depositarse por todo el cuerpo; lo único que perdonan es el sistema nervioso central. Los depósitos de fibras amiloides en corazón, hígado,

articulaciones, tracto gastrointestinal y articulaciones son la marca distintiva de la AL. También se ha propuesto que estas cadenas ligeras pueden formar poros que se insertan en las membranas de las células y las revientan. Ya sea por la deposición de material o por muerte celular, estos desórdenes afectan la función de muchos órganos, causando irremisiblemente la muerte del paciente en pocos años. Cada paciente presenta una proteína diferente, y puede tener más de un tipo de depósitos simultáneamente.

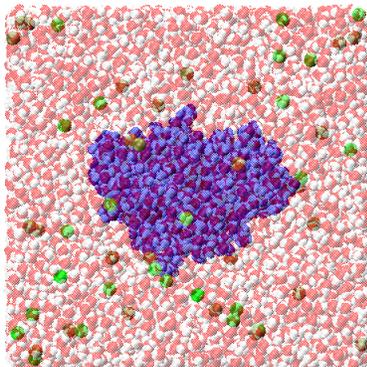
¿Qué hace que una cadena ligera capaz de formar un anticuerpo funcional acabe haciendo fibras, poros o agregados amorfos? No se sabe. Nosotros estudiamos módulos variables de las cadenas ligeras de un par de subclases ($\lambda 3$ y $\lambda 6$) que son los más comunes en los casos clínicos de AL. En particular, en este caso comparamos dos cadenas $\lambda 6$: JTO proviene de un paciente con CN, y WIL de un paciente con AL. No se conoce la estructura de las proteínas en los depósitos de CN. Para AL, sabemos que las fibras amiloides son largas, sin ramas, y extraordinariamente polimórficas en cuanto a su grosor, curvatura y periodo de giro sobre el eje de la fibra. Estas fibras presentan un patrón de difracción conocido como β -cruzado, compatible con una estructura en forma de sándwich, en el que cada pan está hecho de hebras β paralelas o antiparalelas entre sí.

Sabemos que las proteínas en su estado nativo no son capaces de formar fibras. Un modelo sugiere que todas las proteínas tienen la capacidad de hacer fibras amiloides, pero esconden las regiones pro-fibra en el *core* de la proteína de forma que no pueda asociarse con estas regiones pro-fibra en una proteína vecina. Se ha propuesto que coexisten los embudos de plegamiento y asociación, y que estos embudos se comunican a través de los intermediarios de plegamiento de las proteínas monoméricas. No sabemos qué estructura tienen estos intermediarios, y no es por falta de empeño. Son especies moleculares de vidas cortas (son metaestables) y de baja concentración (por ser menos estables que el estado nativo), lo cual complica severamente su caracterización experimental. Volviendo a JTO y a WIL, dado que el primero prefiere agregados amorfos y el segundo fibras amiloides, se ha postulado que pueblan distintos intermediarios. Hacer fibras requiere nucleación homogénea, y esto lo hace un proceso lento, ya que hay que concentrar especies poco frecuentes en el ensamble de conformaciones poblado en condiciones fisiológicas. Por otro lado, hacer agregados amorfos no requiere que todos los monómeros sean iguales, y consecuentemente, es un proceso más rápido. Una gran sorpresa fue encontrar que JTO sí puede hacer fibras *in vitro*, un orden de magnitud más lentamente que WIL.

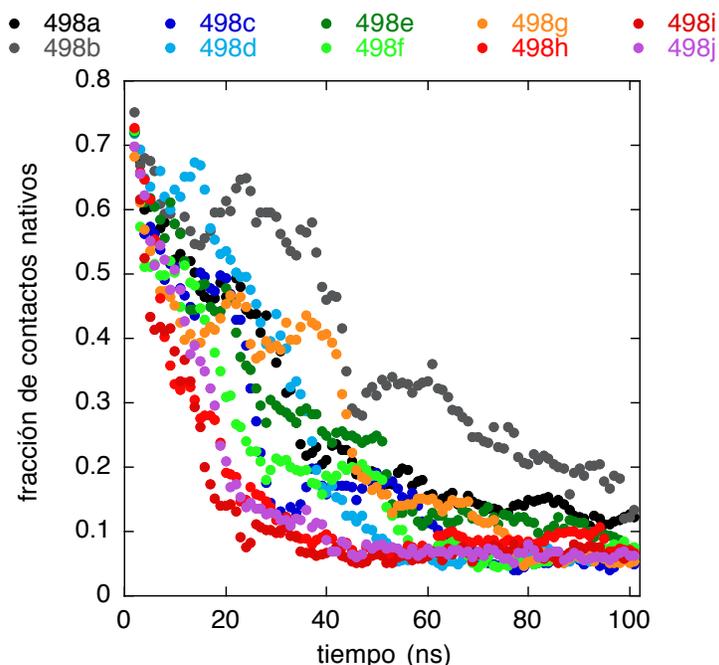
¿Cómo nos imaginamos el proceso de formación de fibras amiloides? Ante una perturbación ambiental (digamos, pH bajo, aumento de sal, presencia de urea, aumento de temperatura) la proteína pierde estructura nativa y puebla de manera más o menos eficiente una colección de intermediarios de plegamiento. Algunos de estos intermediarios se asocian para formar oligómeros, en el proceso de nucleación. Llegando a un tamaño crítico (desconocido), estos núcleos se vuelven competentes para reclutar un monómero a la vez, estableciéndose el proceso de elongación de las fibras. Se desconoce cuánto debe desnaturalizarse la proteína para poder formar núcleos, la estructura de los monómeros que son reclutados durante la elongación, y cuánto de la proteína original forma la estructura de β -cruzada en la fibra. Nosotros estudiamos computacionalmente el paisaje conformacional de los monómeros, para cazar a los intermediarios. Ya que los tenemos, hacemos modelos de las fibras que se podrían armar con estos intermediarios.

LOS PAISAJES DE PLEGAMIENTO

Partimos de la estructura nativa de JTO y WIL, depositada en el Protein Data Bank en los registros 1CD0 y 2CD0 respectivamente. Rodeamos a cada proteína (figura oscura) con una caja cúbica de moléculas de agua (gris y blanco) y iones de potasio y cloro (esferas oscuras), simulando una solución 0.1M de KCl. Describimos las interacciones interatómicas con el potencial CHARMM22 + CMAP, y corremos dinámicas moleculares con NAMD por 50 o 100 ns, en al menos cinco simulaciones independientes. Buena parte de este esfuerzo de simulación se ha hecho en colaboración con César Millán, en KanBalam y en el Centro Nacional de Supercómputo en San Luis Potosí.



Para inducir desplegamiento, aumentamos la temperatura a 398K y a 498K. Sólo la segunda condición despliega eficientemente. Realizamos diez ensayos de desnaturalización para cada proteína, y seguimos el proceso con la pérdida de contactos intramoleculares.



Como puede apreciarse, cada ensayo de desnaturalización sigue un curso temporal distinto; un par pierde abruptamente la estructura a tiempos cortos, mientras que otras se detienen por intervalos más o menos largos en pasos intermedios de desnaturalización; a estas zonas

las llamamos mesetas. Estos son indicios de la existencia de intermediarios de desplegamiento. Si nos amparamos en el postulado de reversibilidad microscópica, estos mismos intermediarios deberían poblarse cuando la proteína se pliega. Tanto JTO como WIL tienen intermediarios en estos ensayos. Ante nosotros tenemos al menos dos escenarios posibles: los intermediarios son los mismos, y lo que es diferente es la población de los mismos; o los intermediarios son distintos, y por eso llevan a destinos diferentes. Las siguientes dos estructuras son tomadas de mesetas de larga vida en JTO y WIL respectivamente, ambas con ~40% de contactos nativos.

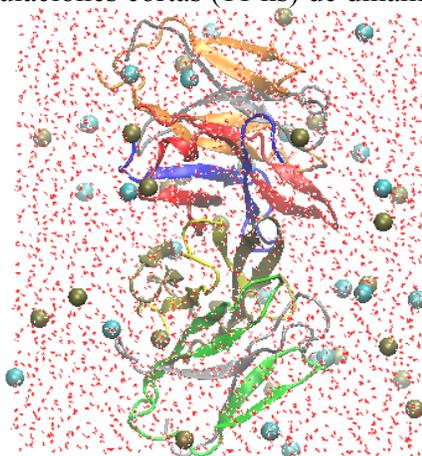


Hay mucha estructura residual, pero es cualitativamente distinta, lo cual podría sugerir diferentes rutas de desplegamiento. Actualmente estamos trabajando en el desarrollo de herramientas computacionales que nos permitan caracterizar a nivel atómico las rutas de desnaturalización, para poder resolver esta parte del mecanismo de formación de monómeros competentes para formar fibras. Como las temperaturas extremas se antojan muy artificiales, Gilberto Valdés está iniciando una serie de simulaciones a temperaturas más bajas pero con urea, de forma que podamos comparar más eficientemente con datos de resonancia magnética nuclear del laboratorio de Carlos Amero en la UAEM, y de desnaturalización con urea del laboratorio de Alejandro Fernández en la UNAM.

MODELOS DE FIBRAS Y LOS INTERMEDIARIOS DE DESPLEGAMIENTO

Otra manera de pensar sobre la ruta de desnaturalización y sus posibles intermediarios es proponer un modelo jerárquico de desensamble de la proteína, como si fuera una alcachofa. Resulta que la estructura de los módulos de los anticuerpos es rica en hebras β , organizadas como un sándwich de hebras antiparalelas, al igual que las fibras amiloides. Estos módulos son inventos muy viejos entre las proteínas, y traen naturalmente mecanismos antiagregación: los bordes del sándwich están ocluidos por hebras torcidas. Fácilmente nos podemos imaginar que podríamos ir quitando estos mecanismos protectores, y luego una hebra a la vez, para generar una familia de conformaciones capaces de polimerizar para formar una fibra. Con estos sub-módulos hemos hecho dos cosas. Por una parte, hemos buscado su población en las simulaciones de desnaturalización. En paralelo, Jessica Araujo y Diana Valenzo han construido tetrámeros con estos sub-módulos, como modelos mínimos de lo que pudiera ser la estructura básica de una fibra amiloide de estas proteínas. Al construir estos modelos nos encontramos con una explosión de isómeros, dado que las hebras pueden asociarse de manera paralela o antiparalela, y usando ambos bordes del sándwich. Esto pudiera ser la raíz del polimorfismo de las fibras amiloides, observada al microscopio electrónico. Al comparar la cantidad de maneras viables de ensamblar

tetrámeros entre variantes con mayor o menor eficiencia para formar fibras experimentalmente, encontramos que a mayor eficiencia, más maneras de hacer fibras. Otra propiedad que puede estar relacionada con la facilidad para formar fibras es la estabilidad de los núcleos. Para ensayar esta propiedad, los modelos fueron rodeados de agua y sal (0.1 M KCl), y refinados con simulaciones cortas (11 ns) de dinámica molecular.



No todos los modelos son estables, y la correlación entre energía interna y eficiencia para hacer fibras no es perfecta, pero sí hay sub-módulos en los cuales esta correlación se encuentra. Esto último pudiera sugerir que estos son los sub-módulos relevantes para describir el proceso de formación de fibras. Todavía falta mucho modelado por hacer, porque los tetrámeros no logran girar 360° para constituir una unidad periódicamente replicable, pero vamos por buen camino.

Cuando preguntamos en las dinámicas moleculares cuánto tiempo persisten estos sub-módulos y las interacciones entre las hebras protectoras, encontramos que JTO es más estable que WIL a todas las temperaturas (298K, 398K y 498K). Esto concuerda con los datos experimentales reportados para la estabilidad térmica de las proteínas. Al buscar la población de estructuras de sub-módulos desprotegidos en los bordes, encontramos que WIL puebla preferencialmente a los sub-módulos más cercanos al módulo completo original, mientras que JTO prefiere aquellos que están más desnaturalizados. Esto apoyaría la idea de distintas rutas de desplegamiento, con la población de intermediarios adecuados para hacer fibras preferencialmente en el caso de WIL, y agregados amorfos en el caso de JTO. En este análisis también encontramos que hay pares de hebras β particularmente estables. Estas horquillas son los elementos estructurales de poros que atraviesan las membranas celulares. En una estancia de verano, Diego Granados construyó un modelo de poro usando como monómeros horquillas largas correspondientes a monómeros 50% desnaturalizados. Su modelo de poro tiene las dimensiones medidas con microscopía de fuerza atómica para poros hechos con proteínas homólogas, y tiene el perfil de carga adecuado para ser débilmente selectivo a cationes, como se ha reportado. Es un modelo preliminar, pero prometedor.

Para los que quieran leer mucho sobre amiloides en todas sus facetas, recomiendo el número especial publicado este año (2012) del Journal of Molecular Biology, volumen 421, dedicado a los mecanismos moleculares y celulares de las amiloidosis.

Este trabajo está financiado por CONACyT (133294).

ATMOSPHERIC TURBULENCE AND
WIND ENERGY UTILIZATION

EDUARDO RAMOS

RENEWABLE ENERGY INSTITUTE

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

1 Introduction

It is of great importance to understand and predict the dynamic and transport properties of turbulent flows since they are commonly found in nature and in engineering. Although a formal definition of turbulence is elusive, it can be said that in all turbulent flows the fluid is in a time dependent, irregular motion. In turbulent flows, kinetic energy flows from large unstable vortical structures to smaller vortices which in turn subsequently generate even smaller vortices up to a scale where the kinetic energy is dissipated by viscosity and converts it into heat. Since the flow is highly irregular in both time and space, most of the descriptions of turbulent flows must be made in terms of statistical theories. The reader interested in the general theory of turbulent flows should consult standard specialized texts like [3] and [10]. In these notes, we shall concentrate the discussion of turbulence only in the context of wind energy utilization.

Wind is regarded as one of the most promising renewable energy sources for the future [6]. Already some countries rely on this form of energy to generate a substantial fraction of their national needs. The most widespread method to convert the kinetic energy in the atmosphere in versatile energy is to use mechanical artifacts that rotate as a result of their interaction with the wind and generate electrical energy. The optimization of the energy conversion procedures require a deep knowledge of the dynamic properties of the wind and of the device itself. The time dependent amplitude of the wind fluctuations is one of its most notable characteristics and it is this feature that may make the source unreliable. The detailed knowledge of the wind motion can turn this renewable energy source into a source that can be trusted.

In the particular application of wind energy utilization, turbulent flows appear in many instances. To describe a few examples, we can say that the atmospheric boundary layer where the wind turbines are located is turbulent due to the interaction of the air in the atmosphere and the ground or sea surfaces. Also, the turbulence is generated at the surface of the turbine blades as vorticity is peeled off from their surface, modifying the lift and drag forces on the blades. Also, in wind energy farms, the turbulent wakes shed by individual wind generators interact with each other and with other wind turbines, reducing the overall performance of the farms.

2 Wind intermittency in the atmosphere

The air motion in the atmosphere is a consequence by the concurrent action of the gravity force and the density differences due to uneven heating of the sun at different latitudes. As it is obvious from our personal experience, the wind changes with the season of the year, time of the day and even at smaller time scales and also changes from one place to another. The geographical non uniformity of the wind is easier to deal with and if enough information is available, favorable sites sites for installing wind generators can be identified. As

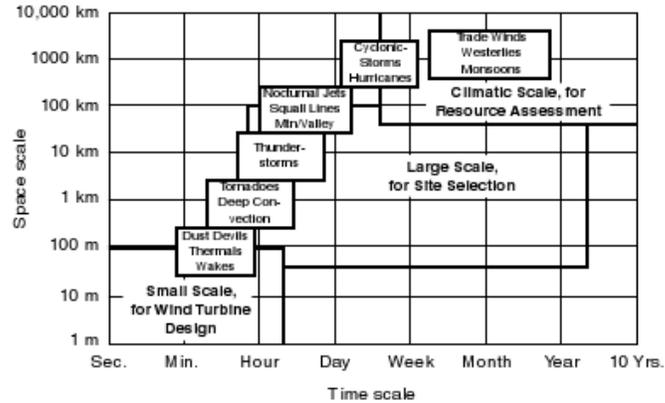


Figure 1: Scales of air motion in the atmosphere [13]. Note the logarithmic scale in the ordinates and the nonlinear scale in the abscissas.

will be explained below, this is the case of coastlines and offshore sites. A broad definition of geographical variations include the presence or absence of water bodies where the surface roughness is much smaller than that occurring in land where mountains, hills, woods and vegetation in general change the properties of the atmospheric boundary layer. Generally speaking, the upper Troposphere wind velocity is reached at lower altitudes over the sea than over the land. This effect will be briefly described in Section 3. Even though at atmospheric scale the motion of air is generated by natural convection, due to the small vertical magnitude as compared to the horizontal magnitudes in the atmosphere, the wind blows mostly horizontally influenced by horizontal pressure gradients and Coriolis forces caused by the rotation of the earth. The resultant of these effects is called the geostrophic wind.

The motion of the atmosphere at planetary scale fuels large and small scale circulations [11]. In Figure 1, a chart of the different spatial and temporal scales of the air motion in the atmosphere are shown. At a climatic scale we find the trade winds, monsoons and hurricanes with magnitudes 10^3 to 10^4 km. Large scale circulations have spatial scales of the order of 10 to 100 km and include tornadoes and mountain/valley winds.

The planetary-scale or climatic-scale air motion and the large scale circulations interact with non uniformities of land and sea distribution and on orographic features like mountains, hills or valleys generating the small scale circulations or local winds. The contrast between the physical properties like thermal capacity between land and water bodies generate land and sea breezes. This phenomenon is particularly notable in the coast lines where diurnal convective cells with air ascending in the land regions and descending in the water zones are generated by the quicker heating of the land due to its relative lower

thermal capacity as compared with water. At night time, the flow is reversed due to the faster cooling of the land. This phenomenon can be used with relative certainty to predict the performance and yield of the of wind generators.

Temporal wind variations are present in several scales, starting from geological or multi-year variations down to fractions of a second due to turbulence fluctuations i.e. dynamic instabilities at small spatial scales where viscosity turns kinetic energy into heat and are described by Kolmogoroff theory (See [4]). The time scales most relevant for wind energy utilization are from seasonal variations relevant for site selection to tenths of a second which corresponds to the response time of small wind turbines.

Detailed description of air motion would require a three-component vector field as a function of three spatial and one temporal coordinates. For a relative well documented spatial and time resolution, this would constitute enormous amounts of information and although desirable, it might be extremely difficult to handle. However, a reasonable alternative description is the use of probability distributions.

3 Atmospheric boundary layer

The atmospheric boundary layer is the lower part of the atmosphere and is the region where the geostrophic wind adjusts itself to the ground velocity. This is the region where the wind turbines are located and its description and prediction of its dynamical features is of great importance to wind energy utilization. The extension of the boundary layer varies with the roughness of the terrain and with the time of the day and the season of the year and it is turbulent. The mean vertical velocity of the atmospheric wind in the boundary layer has a well established profile that can be calculated as follows (See reference [16]). At the lower region of the atmospheric boundary layer, the momentum balance can be assumed to be average steady and is approximately given by

$$\frac{\partial p}{\partial x} = \frac{\partial \tau_{xz}}{\partial z} \quad (1)$$

where p is the pressure and x and z are the horizontal and vertical coordinates respectively. The shear stress in direction x with normal in direction z is denoted by τ_{xz} . At short enough distances from the surface the pressure can be assumed to be independent of the vertical coordinate¹, and the previous equation can be integrated to become

$$\tau_{xz} = \tau_o + z \frac{\partial p}{\partial x} \quad (2)$$

where τ_o is the shear stress at the surface. At small altitudes, the second term is small and can be neglected to conclude that the shear at low altitudes may

¹In reality, the pressure decreases with height. The international standard atmosphere valid to an elevation of 5000 m is $p = 101.29 - (0.011837)z + (4.793 \times 10^{-7})z^2$ where z must be in given in meters and the pressure is in kPa .

be considered constant. An expression for the shear stress can be obtained by invoking Prandtl mixing length theory [12],

$$\tau_{xz} = \rho l^2 \left(\frac{\partial \bar{u}_h}{\partial z} \right)^2 \quad (3)$$

where ρ is the density, \bar{u}_h is the time average² horizontal component of the velocity and l is the mixing length. From the two previous equations, we have

$$\frac{\partial \bar{u}_h}{\partial z} = \frac{1}{l} \sqrt{\frac{\tau_o}{\rho}} = \frac{\bar{u}_h^*}{l}. \quad (4)$$

The *friction velocity* is defined by $\bar{u}_h^* = \sqrt{\tau_o/\rho}$. Using experimental information, the mixing length l can be approximated as a linear function of z as $l = kz$ where $k = 0.4$ is known as the von Kármán's constant. With this expression, equation (4) can be integrated to give the *logarithmic wind profile*:

$$\bar{u}_h(z) = \frac{\bar{u}_h^*}{k} \ln \frac{z}{z_o} \quad (5)$$

here z_o is a reference length that reflects the roughness of the surface. This quantity must be estimated with experimental data; a few values of this variable are given in Table 1.

Type of surface	z_o (mm)
Sea	0.2-0.5
Grass	8.0
Few trees	100
Forest	250
City	3000

4 Statistical description of a turbulent velocity

Given that the flow in the atmosphere is turbulent, we choose to describe its velocity $\mathbf{u}(\mathbf{x}, t)$ in the following terms:

$$\mathbf{u}(\mathbf{x}, t) = \bar{\mathbf{u}}(\mathbf{x}) + \mathbf{u}'(\mathbf{x}, t) \quad (6)$$

where $\bar{\mathbf{u}}(\mathbf{x})$ is the time average of the velocity defined by

$$\bar{\mathbf{u}}(\mathbf{x}) = \frac{1}{T} \int_o^T \mathbf{u}(\mathbf{x}, t) dt = \frac{\delta t}{N} \sum_{n=0}^N \mathbf{u}_n(\mathbf{x}). \quad (7)$$

²Time average is defined in equation (7)

³The friction velocity is estimated from the profile of boundary layers in simplified flows. For a discussion on the estimation of the friction velocity, see reference [5]

and the departure from the average $\mathbf{u}'(\mathbf{x}, t)$ is defined by expressions (6) and (7). In equation (7), we have used the continuous representation for the velocity and the more realistic discrete version where the n reading of the velocity is denoted by $\mathbf{u}_n(\mathbf{x})$. The sampling rate is a constant equal to δt and the following relation holds $n\delta t = T$.

When dealing with random variables, the usual procedure is as follows. A variable of the flow is observed, for instance, one component of the velocity u , a then some kind of statistical processes are apply with the hope that the properties of the flow that we are interested in are captured. The observations can be ordered according to their magnitude:

$$u_1 < u_2 < u_3 < \dots < u_N \quad (8)$$

if the magnitude of u_i is smaller than the magnitude of u_{i+1} . In this expression, the monitored variables can be been relabeled if required.

The following statistical definitions are introduced: The *Cumulative distribution function* (CDF) is:

$$F(u) = \frac{\text{Number of samples with a magnitude smaller than } u}{\text{Total number of elements}} \quad (9)$$

The CDF has the following properties:

$$\begin{aligned} F(-\infty) &= 0 \\ F(\infty) &= 1 \\ F(u_1) &< F(u_2) \quad \text{if} \quad u_1 < u_2 \end{aligned} \quad (10)$$

Observe that equations (10) imply that CDF is a nondecreasing function of variable u .

We define now the *Probability density function* PDF as

$$f(u) = \frac{dF(u)}{du} \quad (11)$$

From properties (10) of the CDF, we conclude that

$$f(u) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f(u^*) du^* = 1. \quad (12)$$

A *Gaussian* or *Normal distribution* is described by the following cumulative distribution function:

$$F(u) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{u - \bar{u}}{\sqrt{2\sigma^2}} \right) \right) \quad (13)$$

The Probability density function of a normal distribution is defined by

$$f(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left(-\frac{1}{2\sigma^2} (u - \bar{u})^2 \right) \quad (14)$$

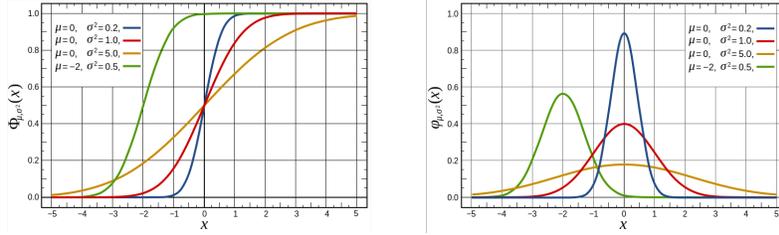


Figure 2: Cummulative distributions (left) and probability density functions (right) of Normal or Gaussian distributions.

The several plots of $f(u)$ for different \bar{u} and σ are given in Figure 2.

This distribution is illustrated in Figure 2. Observe that CDF and PDF depend on two parameters \bar{u} and σ and that the properties listed in equation (10) are satisfied.

An important property of the probability density function is the *standard deviation* defined as:

$$\frac{1}{T} \int_0^T u^2 f(u) du. \quad (15)$$

Observe that in the case of a normal distribution, the standard deviation is equal to σ .

Another useful distribution for the description of wind properties is the *Log-normal distribution* of a random variable u which is such that $\ln u$ is normally distributed.

The probability density function for wind speeds can be obtained by taking readings of the air velocity and processing the data to obtain their statistical properties. However, experience indicates that this is not a straightforward procedure. It has been found that the specific probability density distribution depends on the frequency and averaging of the data, as well as the specific geographical place where the data are taken. For instance, in many geographical locations, it has has been found that the hourly average of wind variation over a year can be represented by the following cumulative density function:

$$F_W(u) = e^{-(u/c)^k} \quad (16)$$

Formally, $F(u)$ is the number of readings that exceed u . The parameters c and k control the scale and shape of the curve respectively. The corresponding probability density function is the *Weibull* probability density function defined by

$$f_W(u) = -\frac{dF}{du} = k \frac{u^{k-1}}{c^k} e^{-(u/c)^k}, \quad (17)$$

and its corresponding annual mean speed is

$$\langle u \rangle = \int_0^{\infty} u^* f(u^*) du^* = c\Gamma(1 + 1/k). \quad (18)$$

In the previous equation, Γ is the generalization of the factorial function and is known as the Gamma function. For a mathematical description of the Gamma function, see reference [15]. Note that the case $k = 2$ of the Weibull distribution is also known as the Rayleigh distribution.

Even though hourly averages of wind recordings at many geographical places can be described using the Weibull distribution. This is not always the case and some other distributions must be used. In some cases, the bi-Weibull distribution reflects more faithfully the statistical properties of the wind.

5 Effect of the wind intermittency on the available wind energy

The transformation of the kinetic energy from the wind into an energy that can be manipulated to make it useful, requires a relatively long chain of energy conversions. The first of these conversions is accomplished by using of a device that turns the turbulent kinetic energy of the air into another form of energy. The most common device, although by no means the only possibility, is of course a wind turbine that rotates as result of its interaction with the wind. In this way, the (on average) one dimensional motion of the wind becomes a rotatory motion that generates a torque. The second link in the chain is the conversion of rotatory power into electricity. Here we will describe some features of these transformations. Before we discuss how a wind machine power can be calculated, it is convenient to give some pertinent definitions. See standard texts on wind engineering like references [1] or [7] for more complete descriptions. The power contained in the wind flowing in a stream tube of cross section A is

$$P_w = \frac{1}{2}\rho A \bar{u}^3 \quad (19)$$

where ρ is the density of air and \bar{u} is the average velocity of the wind. For a wind turbine with horizontal axis, the normalized characteristic velocity is the *tip speed ratio* defined as

$$\lambda = \frac{\Omega R}{\bar{u}} \quad (20)$$

where Ω is the angular velocity of the turbine and R is the length of the blade or equivalently, the radius of the wind turbine. This parameter depends on the aerodynamics of the blades and mechanical components of the turbine. The *rotor power coefficient* is defined by

$$C_p = \frac{\text{Rotor power}}{\text{Wind power}} = \frac{P_{rotor}}{\frac{1}{2}\rho A \bar{u}^3} \quad (21)$$

The rotor coefficient is a function of the incoming wind velocity and the aerodynamic properties of the turbine. It can be demonstrated that the maximum attainable value, regardless of the turbine design is the Betz value and is equal to $C_p^M = 16/27$, or approximately 60%. See references [1] or [7]. The rotor power coefficient can be calculated theoretically using the Blade-Element Momentum (BEM) theory that incorporates information on the wind velocity, number of blades, drag coefficient and other geometrical and dynamic properties [1]. However, in practice this information is often obtained from experimental data. In Figure 3 the rotor power coefficient is shown as a function of the normalized tip speed for several turbine designs.

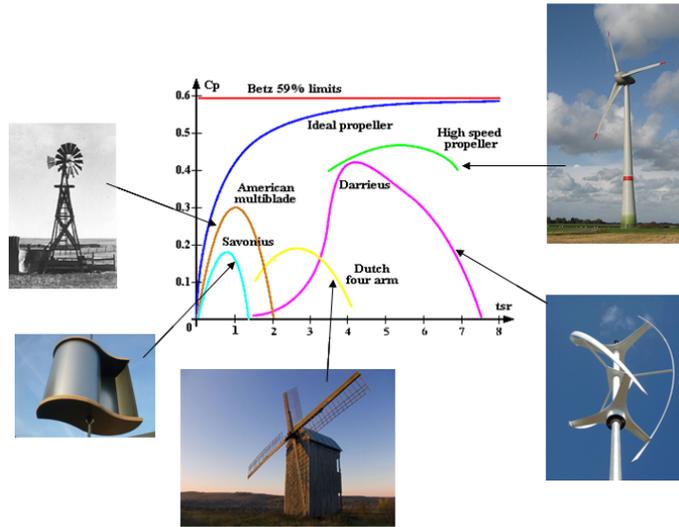


Figure 3: Rotor power coefficient (C_p) as a function of the normalized blade tip speed for different wind turbine designs. Note the Betz limit at $C_p \sim 0.6$.

Equation (21) can be written as

$$P_{rotor} = \frac{1}{2} \rho A C_p \bar{u}^3, \quad (22)$$

and finally, in order to describe the relation of the mechanical power of the rotor with the mechanical power delivered by the entire gear train, we define the *machine power* as

$$P_m = \eta P_{rotor}, \quad (23)$$

where η is the drive train efficiency.

Now, for a given wind probability distribution $p(u)$, the average wind ma-

chine power \bar{P} is:

$$\bar{P} = \int_0^{\infty} P_m(u^*) f(u^*) du^* \quad (24)$$

or using equation (23), the average wind machine is

$$\bar{P} = \frac{1}{2} \rho A \eta \int_0^{\infty} C_p u^{*3} f(u^*) du^* \quad (25)$$

With the previous equations, it is possible to calculate the theoretical maximum power delivered by a wind turbine [2]. Assume an idealized wind turbine with a constant power coefficient equal to the maximum possible to attain by a continuous running turbine, namely, the Betz coefficient $C_p = 16/27$. See Figure 3. Also, take $\eta = 1$ and assume that the hourly mean wind speed over a year is described by a Rayleigh probability distribution and probability density functions as defined in equations (16) and (17) with $k = 2$. The maximum average wind machine power becomes:

$$\bar{P}_M = \frac{1}{2} \rho A \frac{16}{27} u_c^3 \int_0^{\infty} x^3 (2x \exp(-x^2)) dx = \frac{1}{2} \rho A \frac{16}{27} u_c^3 3\sqrt{\pi}/4 \quad (26)$$

where $u_c = 2\bar{u}/\sqrt{\pi}$. Finally, expressing the area of the rotor in terms of the diameter of the rotor as $A = \pi D^2/4$, the maximum average wind machine power simplifies to

$$\bar{P}_M = \rho \left(\frac{2D}{3} \right)^2 \bar{u}^3 \quad (27)$$

As has been shown above, the yearly estimation of the power generated by a wind turbine depends critically on the probability density function of the wind. Unfortunately, data bases of real observations indicate that the information has large variations depending on the geographical location. In the case of Mexico, this problem is even more acute since there are very few reliable data bases and even for those locations, the time span of the data is only few years. With this limited information is impossible to establish the probability density function that describes the wind in the Mexican territory.

The largest data sets available are those of the offshore buoys in the American coastline [8]. It is found that the two parameter Weibull distribution is in many cases the best fit. See Figure 4 for representative fits of offshore data. However, it must be emphasized that the optimum fit depends on the specific site and that in some cases, the best fit would also be dependent on the particular parameter of the turbine that one is interested in estimating [8].

5.1 Turbulence effects on wind turbines

The yearly calculation of the power generated by a wind turbine is a long term estimate of the performance of the turbine, but in order to refine the estimate it is also important to consider the wind variation at other much smaller time and spatial scales. The wind generators have resonant frequencies of 1 Hz. or larger

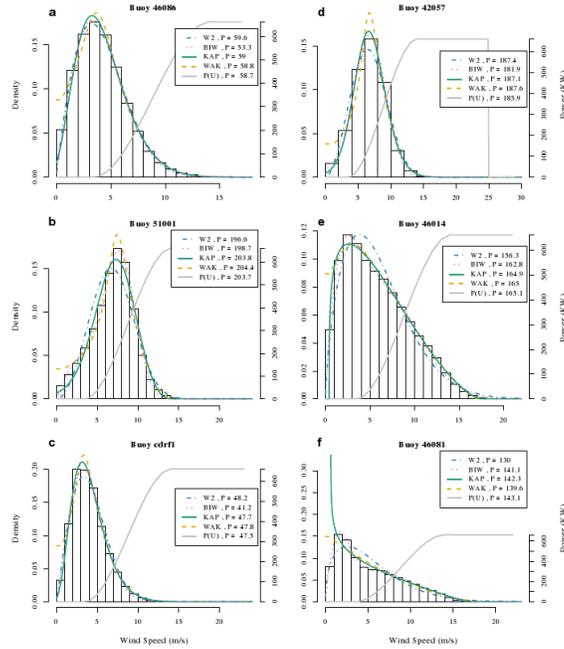


Figure 4: Probability density distribution for several examples of observed data from offshore wind speed measurements [8]. W2: Two parameter Weibull distribution, BIW: Bimodal Weibull model distribution, KAP: Kappa distribution, WAK: Wakeby distribution. The continuous lines indicate the power output. Observe that higher wind speeds have a large contribution for the total power.

and therefore one hour or 10 min. averages are not adequate to estimate the effect of the turbulent wind variation on the turbine. Also, in order to assess the dynamical response of the turbines to turbulent excitation, it is convenient to use the *velocity variation* instead of the velocity itself. The velocity variation is

$$\delta u_{\tau}(t) = u(t + \tau) - u(t) \quad (28)$$

The probability density function of the velocity variation of the wind and its dependence on the delay variable τ are strongly non Gaussian with larger values at the tails, indicating that strong gusts are present more often than those predicted by a Gaussian distribution. An illustration of this wind property is given in figure 5. A specific example from reference [14] indicates that for a 7σ event for velocity increment, the Gaussian probability density function underestimates the probability by a factor of 10^7 , see Figure 5. This is equivalent of saying the Gaussian distribution would predict that an extreme event would happen every 1250 years while in reality it happens once per hour. Observe that according to Taylor hypothesis that indicates that time averages are equivalent

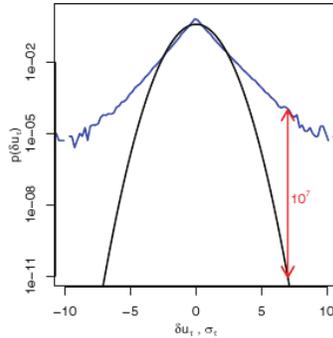


Figure 5: Non Gaussian probability density function for a wind data base. For reference, the Gaussian distribution with the same standard deviation is also shown. The red arrow indicates the position 7σ away from the line of symmetry [14].

to ensemble averages, for an average velocity of 6.6 m/s, and a time lag of $\tau = 3s$, the characteristic separation is 20 m which is comparable to size of a large wind turbine.

In order to illustrate the importance of considering non-Gaussian probability density functions, Figure 6, shows observed data of the variations of the wind velocity for $\tau = 1s$ and the torque variation on a wind turbine due to the wind fluctuations calculated with a stochastic intermittent wind field model. and with a Kaimal wind model. These two distribution functions are compared with the Gaussian distribution to emphasize the large difference of prediction of rare events in the tails. See Figure 5. It is important to remark that large torque variations, imply corresponding stress on the turbine structure and potential failures.

The importance of the tools described in this section has only recently been recognized by the leading groups at international level. Unfortunately, in Mexico there is not enough groups are working on this aspect of the wind energy analysis and the benefits of this theory have not been applied to engineering applications.

6 Concluding remarks

In these notes we review some properties of the wind turbulence that are relevant for wind energy utilization. It is recognized that although in recent years the technology of wind turbines has seen an enormous improvement and the aerodynamics behind the technology is far better understood than it used to be 20 or 30 years ago, the advancement in the knowledge of the physics of the turbulence in the wind is far from satisfactory. A deeper knowledge will have a large impact in improving the performance of the wind turbines and also in

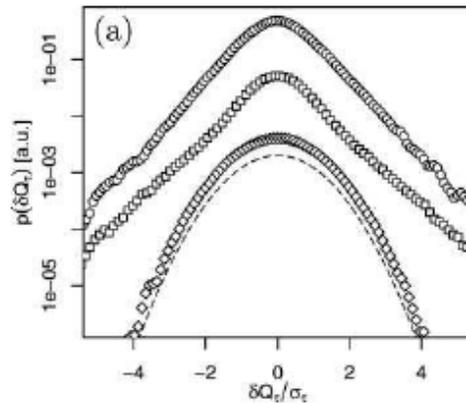


Figure 6: Upper set of points (circles): Probability density function of a wind variation with $\tau = 1s$ obtained at the GROWIAN German wind field measurement station. Points in the middle (squares): Corresponding torque variation on a wind turbine calculated with a stochastic intermittent wind field model. Lower points (diamonds): simulation of the torque variations considering a Kaimal model. A Gaussian distribution is given with the broken line for comparison [14].

their durability and reliability. At a national level, the knowledge of the science and technology of wind energy utilization lags far behind that of other countries and although some groups have started research programs in wind energy, the effort should be multiplied before a progress in the developing the technology can be visible.

References

- [1] Burton T., Sharpe D., Jenkins N., Bossanyi E. (2001) Wind energy handbook. Wiley, New York.
- [2] Carlin P.W. (1997) Analytic expressions for maximum wind turbine average power in a Rayleigh wind regime. Proc. 1997 ASME/AIAA Wind Symposium, 255-263.
- [3] Davidson P. A.(2004) Turbulence. Cambridge University Press.
- [4] Frisch U.(1995) Turbulence. Cambridge University Press.
- [5] Hsu, S.A. (2003) Estimating overwater friction velocity and exponent of power-law wind profile from gust factor during storms. J. Water, Port, Coastal and Ocean Engineering, **129**(4), 174 - 177.

- [6] Lu X., Mc Elroy M.B. and Kiviluoma J. (2009) Global Potential for wind-generated electricity. *Proc. Nat. Acad. Sci.* **106**, 10933-10938.
- [7] Manwell J.F., Mc Gowan J.G. and Rogers A.L. (2008) *Wind energy explained. Theory, design and application.* J. Wiley and Sons.
- [8] Morgan E.C., Lackner M., Vogel R.M. and Baise L.G. (2011) Probability distributions for offshore wind speeds. *Energy Conv. Manag.* **52**, 15-26.
- [9] Mucke T., Kleinhans D., and Peinke J. (2011) Atmospheric turbulence and its influence on the alternating loads on wind turbines. *Wind Energy* **14**, 301316.
- [10] Pope S.B. (2000) *Turbulence.* Cambridge University Press.
- [11] Rohatgi, J. S. and Nelson, V. (1994) *Wind Characteristics: An Analysis for the Generation of Wind Power.* Alternative Energy Institute, Canyon, TX.
- [12] Schlichting H. and Gersten K. (2000) *Boundary layer theory.* Springer.
- [13] Spera, D. A. (Ed.) (1994) *Wind Turbine Technology: Fundamental Concepts of Wind Turbine Engineering.* ASME Press, New York.
- [14] Waechter M., Heisselmann H., Holling M., Morales A., Milan P., Mucke T., Peinke J., Reinke N. and Rinn P. (2012) The turbulent nature of the atmospheric boundary layer and its impact on the wind energy conversion process *J. Turbulence* **13** No 26, 1-21.
- [15] Weber H.J. and Arfken G.B. (2003) *Essential mathematical methods for physicists.* Academic Press.
- [16] Wortman, A. J. (1982) *Introduction to Wind Turbine Engineering.* Butterworth, Boston, MA.

Estados Coherentes
José Récamier
Instituto de Ciencias Físicas
Universidad Nacional Autónoma de México

Resumen

En este curso se presenta la construcción de los estados coherentes de oscilador armónico a partir de sus tres definiciones alternativas [1] y su generalización a álgebras distintas a las del oscilador armónico utilizando el formalismo de operadores deformados [2]. En particular se considera el caso de un potencial tipo Pöschl-Teller trigonométrico con un número infinito de estados ligados y uno modificado con un número finito de estados ligados[3].

1 Introducción

La propuesta de crear un estado coherente para un sistema cuántico se debe a Schrödinger quien en 1926 los propone como los estados cuánticos más parecidos a los estados clásicos del oscilador armónico, esto es, los estados cuánticos que tengan la menor dispersión posible acorde con el principio de incertidumbre de Heisenberg. No fue sino hasta 1963 cuando Glauber los utilizó para el estudio de las funciones de correlación del campo electromagnético [1]. Casi al mismo tiempo, Klauder [4] desarrolló un conjunto de estados continuos que contienen las bases para la generación de estados coherentes para grupos de Lie arbitrarios. Posteriormente, Gilmore [5] y Perelomov [6] lograron la construcción completa de estados coherentes para grupos de Lie. Un resultado importante de esta construcción es que los estados coherentes no tienen que estar restringidos a el oscilador armónico sino que pueden ser generalizados a otros sistemas físicos. Una excelente revisión de este tema puede encontrarse en Zhang [7].

2 Estados coherentes del campo

Glauber mostró que para el álgebra del oscilador armónico, es posible construir los estados coherentes a partir de cualquiera de las siguientes definiciones: *i)* como estados propios del operador de aniquilación $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$, *ii)* como estados de mínima incertidumbre $(\Delta X)(\Delta P) = \hbar/2$ con $(\Delta X) = (\Delta P)$ y *iii)* como los estados que se obtienen al desplazar el vacío $D(\alpha)|0\rangle = |\alpha\rangle$ siendo $D(\alpha) = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a})$. Veremos a continuación que estas tres definiciones son equivalentes.

Para el caso del oscilador armónico es útil escribir a los operadores de posición \hat{x} y momento \hat{p} en términos de operadores \hat{a} , \hat{a}^\dagger definidos mediante:

$$\hat{x} = \sqrt{\frac{\hbar}{2m\Omega}}(\hat{a} + \hat{a}^\dagger), \quad \hat{p} = i\sqrt{\frac{\hbar m\Omega}{2}}(\hat{a}^\dagger - \hat{a}) \quad (1)$$

en donde Ω es la frecuencia del oscilador, m su masa y $\hbar = h/2\pi$ siendo h la constante de Plank. El Hamiltoniano del oscilador armónico

$$H_{oa}(\hat{x}, \hat{p}) = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\Omega^2\hat{x}^2$$

escrito en términos de los operadores \hat{a} , \hat{a}^\dagger toma la forma:

$$H_{oa}(\hat{a}, \hat{a}^\dagger) = \hbar\Omega(\hat{a}^\dagger\hat{a} + \frac{1}{2}) \equiv \hbar\Omega(\hat{n} + \frac{1}{2})$$

cuando se han usado las definiciones dadas en la ecuación 1. Los estados de número (o estados de Fock) $|n\rangle$ con $(n=0, 1, \dots)$ son eigenestados de H_{oa} [8], esto es:

$$H_{oa}|n\rangle = \hbar\Omega(\hat{n} + \frac{1}{2})|n\rangle = \hbar\Omega(n + \frac{1}{2})|n\rangle. \quad (2)$$

Debido a que el conmutador de los operadores \hat{a} , \hat{a}^\dagger es un escalar $[\hat{a}, \hat{a}^\dagger] = 1$, es posible escribir el operador de desplazamiento en forma de un producto de exponenciales [9]:

$$D(\alpha) = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a}) = \exp(-\frac{|\alpha|^2}{2}) \exp(\alpha\hat{a}^\dagger) \exp(-\alpha\hat{a}). \quad (3)$$

Utilizando ahora el desarrollo de la exponencial

$$\exp(\alpha\hat{O}) = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} \hat{O}^n$$

y el hecho de que los operadores de aniquilación \hat{a} , de creación \hat{a}^\dagger y de número \hat{n} actuando sobre los estados propios del oscilador armónico (o estados de Fock) $|n\rangle$ dan: $\hat{a}|n\rangle = \sqrt{n}|n-1\rangle$, $\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$, $\hat{n}|n\rangle = n|n\rangle$ obtenemos el estado coherente $|\alpha\rangle$ al aplicar el operador de desplazamiento dado por la ecuación (3) sobre el estado de vacío $|0\rangle$:

$$|\alpha\rangle = D(\alpha)|0\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (4)$$

Cabe destacar que los estados dados en la ecuación 4 están normalizados. Mostraremos ahora que los estados $|\alpha\rangle$ dados arriba son eigenestados del operador de aniquilación \hat{a} . Tomando el desarrollo dado en la ecuación 4 tenemos:

$$\begin{aligned} \hat{a}|\alpha\rangle &= e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} \hat{a}|n\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} \sqrt{n}|n-1\rangle \\ &= \alpha e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^{n-1}}{\sqrt{(n-1)!}} |n-1\rangle = \alpha e^{-\frac{|\alpha|^2}{2}} \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} |m\rangle = \alpha|\alpha\rangle. \end{aligned} \quad (5)$$

el último paso es posible debido a que el número de estados que intervienen en la suma es infinito. Una forma alternativa de obtener el estado coherente que será útil más adelante es la siguiente. Escribimos el estado coherente como un desarrollo en términos de estados de número con coeficientes c_n y pedimos que se cumpla $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$

$$\hat{a} \sum_{n=0}^{\infty} c_n |n\rangle = \sum_{n=0}^{\infty} c_n \sqrt{n} |n-1\rangle = \sum_{m=0}^{\infty} c_{m+1} \sqrt{m+1} |m\rangle = \alpha \sum_{n=0}^{\infty} c_n |n\rangle \quad (6)$$

de donde se obtiene $c_{n+1} = \alpha c_n / \sqrt{n+1}$. Aplicando nuevamente el operador \hat{a} al estado coherente se obtiene $c_{n+2} = \alpha^2 c_n / \sqrt{(n+1)(n+2)}$, de forma que el desarrollo puede escribirse en términos de c_0 como:

$$|\alpha\rangle = c_0 (|0\rangle + \frac{\alpha}{\sqrt{1}}|1\rangle + \frac{\alpha^2}{\sqrt{2}}|2\rangle + \dots + \frac{\alpha^n}{\sqrt{n!}}|n\rangle + \dots) = c_0 \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \quad (7)$$

siendo ahora c_0 una constante de normalización.

Finalmente mostraremos que estos estados saturan la relación de incertidumbre de Heisenberg. Para ello notamos que:

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle \quad \text{y} \quad \langle\alpha|\hat{a}^\dagger = \alpha^*\langle\alpha|$$

así que, el valor esperado de los operadores \hat{x} y \hat{p} entre estados coherentes es:

$$\langle \alpha | \hat{x} | \alpha \rangle = \sqrt{\frac{\hbar}{2m\Omega}} (\alpha + \alpha^*), \quad \langle \alpha | \hat{p} | \alpha \rangle = i\sqrt{\frac{\hbar m\Omega}{2}} (\alpha^* - \alpha). \quad (8)$$

Los operadores \hat{x}^2 y \hat{p}^2 dados en términos de los operadores \hat{a} y \hat{a}^\dagger son:

$$\hat{x}^2 = \left(\frac{\hbar}{2m\Omega} \right) (\hat{a}^2 + 2\hat{n} + 1 + \hat{a}^{\dagger 2}), \quad \hat{p}^2 = - \left(\frac{\hbar m\Omega}{2} \right) (\hat{a}^2 - 2\hat{n} - 1 + \hat{a}^{\dagger 2}) \quad (9)$$

de donde, los valores esperados entre estados coherentes quedan:

$$\langle \alpha | \hat{x}^2 | \alpha \rangle = \left(\frac{\hbar}{2m\Omega} \right) (\alpha^2 + 2|\alpha|^2 + 1 + \alpha^{*2}) \quad (10)$$

$$\langle \alpha | \hat{p}^2 | \alpha \rangle = - \left(\frac{\hbar m\Omega}{2} \right) (\alpha^2 - 2|\alpha|^2 - 1 + \alpha^{*2}) \quad (11)$$

y las dispersiones son:

$$\Delta \hat{x} = \sqrt{\langle \alpha | \hat{x}^2 | \alpha \rangle - \langle \alpha | \hat{x} | \alpha \rangle^2} = \left(\frac{\hbar}{2m\Omega} \right)^{\frac{1}{2}} \quad (12)$$

$$\Delta \hat{p} = \sqrt{\langle \alpha | \hat{p}^2 | \alpha \rangle - \langle \alpha | \hat{p} | \alpha \rangle^2} = \left(\frac{\hbar m\Omega}{2} \right)^{\frac{1}{2}} \quad (13)$$

así que el producto de las dispersiones

$$(\Delta \hat{x})(\Delta \hat{p}) = \frac{\hbar}{2} \quad (14)$$

es independiente del valor del parámetro α que es, en general, un número complejo arbitrario.

Si calculamos ahora el valor esperado del Hamiltoniano entre estos estados coherentes obtenemos

$$\langle \alpha | H_{oa} | \alpha \rangle = \hbar\Omega \langle \alpha | \hat{n} + \frac{1}{2} | \alpha \rangle = \hbar\Omega \left(|\alpha|^2 + \frac{1}{2} \right) \quad (15)$$

así que el número medio de excitaciones en el oscilador está dado por el cuadrado del *tamaño* del estado coherente $|\alpha|^2$.

La probabilidad de encontrar ocupado el estado n en un estado coherente $|\alpha\rangle$ está dada por

$$P_n(\alpha) = |\langle n|\alpha\rangle|^2 = \left| e^{-\frac{|\alpha|^2}{2}} \frac{\alpha^n}{\sqrt{n!}} \right|^2 = e^{-|\alpha|^2} \frac{|\alpha|^{2n}}{n!} \quad (16)$$

que es una distribución de Poisson. Como puede verse en la figura 1, para $|\alpha| = 1$ el máximo en la probabilidad de ocupación se encuentra cerca de $n = 1$, cuando $|\alpha| = \sqrt{5}$ el máximo en la distribución está localizado cerca de $n = 5$. La dispersión en el número de ocupación $\Delta n = \sqrt{\langle \hat{n}^2 \rangle - \langle \hat{n} \rangle^2}$

$$\Delta n = \sqrt{\langle \alpha | \hat{n}^2 | \alpha \rangle - \langle \alpha | \hat{n} | \alpha \rangle^2} = |\alpha| \quad (17)$$

crece con el *tamaño* del estado coherente como puede observarse también en la figura 1.

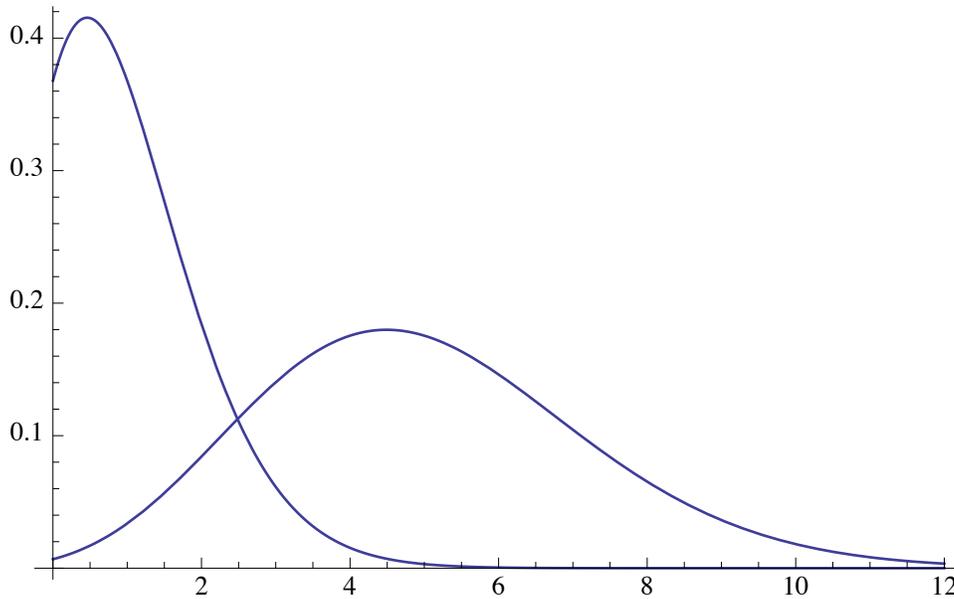


Figure 1: Probabilidad de ocupacion del estado n para un estado coherente con $\alpha = 1$ y $\alpha = 1 + 2i$

3 Oscilador deformado

Un oscilador deformado es uno en el que los operadores de creación y aniquilación son modificados por medio de una función del operador de número. Esto es, se definen los operadores deformados \hat{A} , \hat{A}^\dagger como [2]:

$$\hat{A} = \hat{a}f(\hat{n}), \quad \hat{A}^\dagger = f(\hat{n})\hat{a}^\dagger \quad (18)$$

siendo la función $f(\hat{n})$ una función del operador de número que especifica el tipo de deformación del Hamiltoniano. Usando las definiciones dadas arriba se obtienen los conmutadores:

$$[\hat{A}, \hat{n}] = \hat{A}, \quad [\hat{A}^\dagger, \hat{n}] = -\hat{A}^\dagger, \quad [\hat{A}, \hat{A}^\dagger] = (\hat{n} + 1)f^2(\hat{n} + 1) - \hat{n}f^2(\hat{n}) \quad (19)$$

vemos que el límite armónico se obtiene haciendo $f(\hat{n}) = 1$. Como mencionamos anteriormente, la función $f(\hat{n})$ especifica la forma del oscilador deformado, a continuación presentaremos un par de ejemplos que ilustran este punto.

3.1 Potencial de Pöschl-Teller modificado

Este potencial ha sido muy utilizado en problemas de física atómica y molecular y tiene la forma:

$$V_{PT} = U_0 \tanh^2(ax) \quad (20)$$

en donde U_0 es la profundidad del pozo, a es el rango del potencial y x es la distancia con respecto a la posición de equilibrio.

Las soluciones a la ecuación de Schrödinger correspondiente y sus eigenvalores son conocidos [10] y están dados por:

$$\Psi_n^\epsilon(\zeta) = N_n^\epsilon (1 - \zeta^2)^{\epsilon/2} F(-n, \epsilon + s + 1; \epsilon + 1, (1 - \zeta)/2) \quad (21)$$

y

$$\begin{aligned} E_n &= U_0 - \frac{\hbar^2 a^2}{8\mu} \left(-(2n + 1) + \sqrt{1 + \frac{8\mu U_0}{\hbar^2 a^2}} \right)^2 = U_0 - \frac{\hbar^2 a^2}{2\mu} (s - n)^2 \\ &= \frac{\hbar^2 a^2}{2\mu} (s + 2sn - n^2) = \frac{\hbar^2 a^2 s}{\mu} \left(n + \frac{1}{2} - \frac{n^2}{2s} \right) \end{aligned} \quad (22)$$

en donde N_n^ϵ es un factor de normalización, $\zeta = \tanh(ax)$, μ es la masa reducida de la molécula, el parámetro s está relacionado con la intensidad

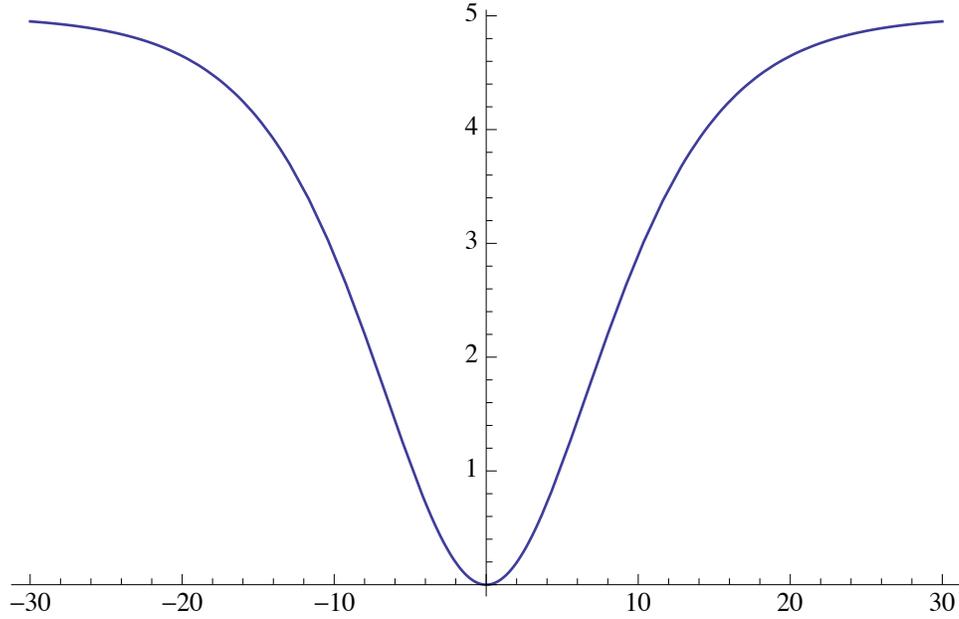


Figure 2: Potencial de Pöschl-Teller modificado, aquí tomamos $U_0 = 5$ y $a = 0.1$ corresponde a un sistema que soporta 20 estados ligados

del potencial y su alcance mediante la expresión $s(s+1) = 2\mu U_0/\hbar^2 a^2$, $\epsilon = \sqrt{-2\mu(E - U_0)}/\hbar a$ y $F(a, b; c, z)$ es el símbolo que representa a la función hipergeométrica [11]. De la expresión para los eigenvalores vemos que el número de estados ligados para este potencial esta determinado por la condición $\epsilon = s - n = 0$. Tomando en cuenta que el parámetro s es un número entero, la función propia para el estado con energía nula ($n = s$) no es normalizable. Por tanto, el último estado ligado corresponde a $n_{max} = s - 1$ [12]. Nótese que el espectro corresponde al de un oscilador armónico modificado por una contribución no lineal cuyo signo es responsable de que el potencial soporte solamente un número finito de estados ligados.

Consideremos ahora un Hamiltoniano de la forma del oscilador armónico escrito en términos de operadores deformados:

$$H_D = \frac{\hbar\Omega}{2}(\hat{A}\hat{A}^\dagger + \hat{A}^\dagger\hat{A}). \quad (23)$$

Si elegimos la función de deformación $f(\hat{n})$ como:

$$f^2(\hat{n}) = \frac{\hbar a^2}{2\mu\Omega}(2s + 1 - \hat{n}) \quad (24)$$

entonces, el Hamiltoniano deformado dado por la Ec.23 queda:

$$H_D = \frac{\hbar^2 a^2}{2\mu}(s + 2s\hat{n} - \hat{n}^2) = \frac{\hbar^2 a^2 s}{\mu}(\hat{n} + \frac{1}{2} - \frac{\hat{n}^2}{2s}) \quad (25)$$

cuyos eigenvalores calculados entre estados de número son idénticos a los dados en la ecuación 22.

3.2 Potencial de Pöschl-Teller trigonométrico

Este potencial tiene la forma

$$V(x) = U_0 \tan^2(ax) \quad (26)$$

en donde U_0 es una constante real positiva, a es el rango del potencial y x es la distancia a la posición de equilibrio. Para este potencial, la solución a la ecuación de Schrödinger es [13]:

$$\psi_n(x) = \sqrt{\frac{a(\lambda + n)\Gamma(2\lambda + n)}{\Gamma(n + 1)}}(\cos(ax))^{1/2} P_{n+\lambda-1/2}^{1/2-\lambda}(\sin(ax)) \quad (27)$$

y los valores propios son:

$$E_n = \frac{\hbar^2 a^2}{2\mu}(n^2 + 2n\lambda + \lambda) = \frac{\hbar^2 a^2 \lambda}{\mu}(n + \frac{1}{2} + \frac{n^2}{2\lambda}) \quad (28)$$

siendo μ la masa reducida de la molécula, el parámetro λ esta relacionado con la profundidad y alcance del potencial mediante $\lambda(\lambda - 1) = 2\mu U_0 / \hbar^2 a^2$ y $P_{n-\lambda-1/2}^{1/2-\lambda}$ son los polinomios de Legendre. De la ecuación para λ vemos que ésta tiene que ser mayor que 1 y de la expresión para los eigenvalores vemos que en este caso el número de estados ligados que soporta el potencial es infinito.

Procediendo en forma similar a la usada en el caso anterior, proponemos una función de deformación

$$f^2(\hat{n}) = \frac{\hbar a^2}{2\mu\Omega}(\hat{n} + 2\lambda - 1) \quad (29)$$

y el Hamiltoniano deformado toma ahora la forma

$$H_D = \frac{\hbar^2 a^2}{2\mu} (\hat{n}^2 + 2\lambda\hat{n} + \lambda) \quad (30)$$

cuyos eigenvalores coinciden con los dados en la ecuación 28.

Con estos dos ejemplos se ha mostrado cómo es posible modelar a un potencial no armónico dado al escoger una forma explícita para la función de deformación. En la referencia [14] se muestran algunos otros ejemplos y sus estados coherentes.

4 Estados coherentes para potenciales generales

Construiremos los estados coherentes generalizados como estados propios del operador de aniquilación deformado

$$\hat{A}|\alpha, f\rangle = \alpha|\alpha, f\rangle \quad (31)$$

Procediendo en forma similar a la utilizada para el caso del oscilador armónico, escribimos el estado $|\alpha, f\rangle$ como un desarrollo en términos de estados de número y se encuentra que los coeficientes del desarrollo toman la forma:

$$|\alpha, f\rangle = N_f \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!f(n)!}} |n\rangle \quad (32)$$

en donde N_f es una constante de normalización, $|n\rangle$ es un estado propio del oscilador no deformado y $f(n)! = f(0)f(1)\cdots f(n)$. Sustituyendo el desarrollo dado en la ecuación 40 en la ecuación 31 podemos comprobar que los estados $|\alpha, f\rangle$ son estados propios del operador de aniquilación deformado.

Esta expresión para construir los estados coherentes generales es aplicable para cualquier función de deformación y proporciona estados coherentes aproximados cuando se trata de un potencial que soporta un número finito de estados ligados y estados coherentes exactos cuando el número de estados ligados que soporta el potencial es infinito. Considerando primeramente el caso del potencial de Pöschl-Teller modificado tenemos que el número de términos en la suma es finito ($N_{max} = s - 1$) y por lo tanto los estados

$|\alpha, f\rangle$ son aproximados. Al reemplazar la forma explícita de la función de deformación obtenemos:

$$|\alpha, f\rangle = N_f \sum_{n=0}^{s-1} \sqrt{\frac{\Gamma(2s+1-n)}{n!\Gamma(2s+1)}} \left(\frac{2s}{\chi}\right)^{n/2} \alpha^n |n\rangle \quad (33)$$

en donde $\chi = \hbar a^2 s / \mu \Omega$.

Para el caso en que los operadores deformados y el operador de número cierran un álgebra de Lie, es posible obtener una expresión analítica para el operador de desplazamiento deformado $D_D(\alpha) = \exp(\alpha \hat{A}^\dagger - \alpha^* \hat{A})$ [15].

Consideremos primeramente el caso del potencial de Pöschl-Teller modificado. El conjunto de operadores relevantes cumple con las reglas de conmutación:

$$[\hat{A}, \hat{n}] = \hat{A}, \quad [\hat{A}^\dagger, \hat{n}] = -\hat{A}^\dagger, \quad [\hat{A}, \hat{A}^\dagger] = \frac{\hbar a^2}{\mu \Omega} (s - \hat{n}) \quad (34)$$

nótese que el conmutador entre los operadores deformados es un escalar más una función *lineal* del operador de número. Este hecho nos permite generalizar el método del operador de desplazamiento ya que si escribimos $[\hat{A}, \hat{A}^\dagger] = g(\chi \hat{n})$, obtenemos el conjunto de conmutadores:

$$[\hat{A}, g(\chi \hat{n})] = -\frac{\chi}{s} \hat{A}, \quad [\hat{A}^\dagger, g(\chi \hat{n})] = \frac{\chi}{s} \hat{A}^\dagger, \quad [\hat{A}, \hat{A}^\dagger] = g(\chi \hat{n}). \quad (35)$$

Claramente los operadores \hat{A} , \hat{A}^\dagger y $g(\chi \hat{n})$ son elementos de un álgebra de Lie, por lo tanto es posible escribir la exponencial de la suma en forma de un producto de exponenciales [15]. De esta manera es posible escribir el operador de desplazamiento generalizado como:

$$\begin{aligned} D_D(\alpha) &= \exp[\alpha \hat{A}^\dagger - \alpha^* \hat{A}] \\ &= \exp \left[\alpha \frac{\tan(|\alpha| \sqrt{\chi/2s})}{|\alpha| \sqrt{\chi/2s}} \hat{A}^\dagger \right] \exp \left[\frac{\ln(\cos(|\alpha| \sqrt{\chi/2s}))}{\chi/2s} g(\chi \hat{n}) \right] \times \\ &\times \exp \left[-\alpha^* \frac{\tan(|\alpha| \sqrt{\chi/2s})}{|\alpha| \sqrt{\chi/2s}} \hat{A} \right]. \end{aligned} \quad (36)$$

Si escribimos $\alpha = |\alpha| e^{i\phi}$ y definimos $\zeta = e^{i\phi} \tan(|\alpha| \sqrt{\chi/2s})$ la ecuación Ec. 36 queda:

$$D_D(\zeta) = \exp \left(\zeta \frac{\hat{A}^\dagger}{\sqrt{\chi/2s}} \right) \left(\frac{1}{1 + |\zeta|^2} \right)^{sg(\chi \hat{n})/\chi} \exp \left(-\zeta^* \frac{\hat{A}}{\sqrt{\chi/2s}} \right) \quad (37)$$

y es posible obtener los estados coherentes aplicando el operador $D_D(\zeta)$ al estado de vacío. Al hacerlo obtenemos:

$$|\zeta\rangle = \frac{1}{(1 + |\zeta|^2)^s} \sum_n \frac{\zeta^n}{\sqrt{(\chi/2s)^n}} \frac{f(n)!}{\sqrt{n!}} |n\rangle. \quad (38)$$

Tomando el caso particular del potencial de Pöschl-Teller modificado en donde la suma debe restringirse a los primeros $s - 1$ términos (debido a que se cuenta con un número finito de estados ligados) se obtienen los estados coherentes aproximados

$$|\zeta\rangle \simeq \frac{1}{(1 + |\zeta|^2)^s} \sum_{n=0}^{s-1} \sqrt{\frac{\Gamma(2s + 1)}{n! \Gamma(2s + 1 - n)}} \zeta^n |n\rangle. \quad (39)$$

Para el caso del potencial de Pöschl-Teller trigonométrico, el número de estados ligados es infinito y las expresiones que se obtienen para los estados coherentes son exactas. Para los construidos como estados propios del operador de aniquilación deformado obtenemos:

$$|\alpha, f\rangle = N_{f,\alpha} \sum_{n=0}^{\infty} \sqrt{\frac{\Gamma(2\lambda)}{n! \Gamma(2\lambda + n)}} \left(\frac{2\mu\Omega}{\hbar a^2} \right)^{n/2} \alpha^n |n\rangle. \quad (40)$$

mientras que los obtenidos aplicando el operador de desplazamiento generalizado quedan:

$$|\zeta\rangle = (1 - |\zeta|^2)^\lambda \sum_{n=0}^{\infty} \sqrt{\frac{\Gamma(2\lambda + n)}{n! \Gamma(2\lambda)}} \zeta^n |n\rangle. \quad (41)$$

Nótese que aunque los estados dados en las ecuaciones 40 y 41 son exactos no son iguales entre sí. Cabe destacar que sus propiedades estadísticas son diferentes, mientras que unos cumplen con una estadística super-poissoniana los otros cumplen con una sub-poissoniana sin embargo los valores medios de diversas observables calculados con estos estados coherentes son prácticamente idénticos (al menos para el caso de estados coherentes para el potencial de Morse) [16].

References

- [1] R J Glauber, Phys. Rev. Lett. **10**, 84 (1963).
- [2] V I Man'ko, G Marmo, F Zaccaria and ECG Sudarshan, Phys. Scr. **55**, 528 (1997).
- [3] O. de los Santos and J. Récamier, J. Phys. A: Math. Theor. **44**, (2011) 145307
- [4] J R Klauder, J. Math. Phys. **4**, 1055 (part 1); 1058 (part 2) (1963).
- [5] R Gilmore, Ann. Phys. (NY) **74**, 391 (1972); Rev. Mex. Fís. **23**, 142 (1974).
- [6] A M Perelomov, Commun. Math. Phys. **26**, 222 (1972).
- [7] W M Zhang W, D H Feng and R Gilmore, Reviews of Modern Physics **62**, 868 (1990).
- [8] Messiah Albert 1999 *Quantum Mechanics*, Ch. XII, (Dover, New York).
- [9] Merzbacher E 1970 *Quantum mechanics*, (Second edition, Wiley, New York)
- [10] Landau L, Lifshitz E, 1967 *Mechanique Quantique: Théorie non Relativiste* (Moscou: Deuxième édition MIR) p. 94.
- [11] M. Abramowitz and I A Stegun, 1972 *Handbook of Mathematical Functions* (New York, Dover) p 555
- [12] R Lemus and R Bernal, Chem. Phys. **283**, 401 (2002).
- [13] M M Nieto, Phys. Rev. A **17**, 1273 (1978)
- [14] R. Román-Ancheyta, O. de los Santos-Sánchez and J. Récamier, J. Phys A: Math. Theor. **44**, (2011) 435304
- [15] O. de los Santos Sánchez, *Estados coherentes no lineales para potenciales generales*, tesis doctoral, Universidad Autónoma de Puebla, Instituto de Física Luis Rivera Terrazas, junio 2011.
- [16] J. Récamier, W. L. Mochán, M. Gorayeb, J. L. Paz, R. Jáuregui, Int. J. Modern Physics B **20**, (11-12) 1851-1859 (2006).

Las simulaciones numéricas de sistemas moleculares

Humberto Saint Martin

*Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México,
Apartado Postal 48-3, Cuernavaca, Morelos 62251, México*

6 de Marzo de 2013

Resumen: En este curso de verano se plantea una jerarquía de simulaciones de sistemas reales con diferentes requisitos de precisión. La elección del método de simulación adecuado depende de la escala propia del fenómeno de interés y del nivel de detalle que se desee incluir. La explicación de varias propiedades macroscópicas de las fases condensadas de la materia puede obtenerse a partir de las interacciones de las moléculas que la componen. Para ello se utilizan los métodos de la *Química Cuántica* y de las simulaciones por *Monte Carlo* y por *dinámica molecular*, que se describen en estas notas.

Palabras clave (*Keywords*): simulaciones numéricas, modelos moleculares, fases condensadas.

PACS: 61.20.Ja, 61.46.Fg, 64.70.dg, 82.39.Wj, 87.15.ak, 87.16.dp, 87.16.Vy

Sobre el Rigor en la Ciencia

En aquel Imperio, el Arte de la Cartografía
logró tal Perfección que el Mapa de una sola
Provincia ocupaba toda una Ciudad,
y el Mapa del Imperio, toda una Provincia.
Con el tiempo, estos Mapas Desmesurados
no satisficieron y los Colegios de Cartógrafos
levantaron un Mapa del Imperio,
que tenía el Tamaño del Imperio y
coincidía puntualmente con él.

Menos Adictas al Estudio de la Cartografía,
las Generaciones Sigüientes entendieron
que ese dilatado Mapa era Inútil y no sin Impiedad
lo entregaron a las Inclemencias del Sol y los Inviernos.
En los Desiertos del Oeste perduran despedazadas
Ruinas del Mapa, habitadas por Animales y por Mendigos;
en todo el País no hay otra reliquia de las Disciplinas Geográficas.

Suárez Miranda: Viajes de varones prudentes
Libro Cuarto, Cap. XLV, Lérica 1658

De: *El hacedor* (1960), por **Jorge Luis Borges**

Introducción

En la actualidad el modelado molecular ha alcanzado tal sofisticación que se ha hecho común el uso de las simulaciones numéricas para representar, describir e incluso predecir el comportamiento de diversos sistemas moleculares con creciente precisión. Las técnicas actuales, así como las capacidades de cómputo masivo, permiten abordar problemas con diferentes grados de complejidad; dependiendo del fenómeno de interés, se puede elegir un nivel de descripción que va desde representaciones de tipo de dinámica de Langevin, a dinámica Browniana; a dinámica de grano grueso con detalle molecular; a dinámica molecular con detalle atómico, a efectos de las nubes electrónicas, a cálculos cuánticos de las funciones de onda electrónicas.

Para elegir el nivel de descripción a utilizar en un caso específico, convendría saber de antemano la relevancia de los detalles que un modelo dado soslaye en el fenómeno bajo estudio. Por ejemplo, no es posible usar un modelo clásico de átomos para describir correctamente una reacción química, sino que se necesita hacer algún tipo de cálculo cuántico; en cambio, el modelo clásico de átomo puede

incluso estar sobrado para describir un proceso de separación entre líquidos no miscibles, como el agua y el aceite.

Sin embargo, para muchos sistemas de interés bioquímico y nanotecnológico, no es tan fácil juzgar de antemano la relevancia de distintos efectos; por ejemplo, a diferencia de los electrones, los átomos son objetos clásicos bajo condiciones estándar de presión y temperatura (sus longitudes de onda son suficientemente pequeñas), pero existen detalles finos de comportamiento de fases condensadas cuya descripción cuantitativa requiere de tomar en cuenta la deslocalización cuántica de los átomos ligeros, típicamente de los hidrógenos. ¹ Así que en la práctica del modelado molecular más bien se utiliza un proceso de aproximación sucesiva” (= ensayo y error), común en la investigación científica, en particular en la Física: se comienza utilizando el modelo más simple del que se disponga, de preferencia que ya se haya probado con sistemas similares al de interés, y se realizan los cálculos y las simulaciones necesarias para obtener resultados comparables con datos experimentales del sistema o del fenómeno real. Si resulta un buen acuerdo con los datos experimentales disponibles, entonces se consideran confiables los resultados adicionales del modelo, para los cuales podría no existir contraparte experimental. A este proceso se le conoce como la *validación* del modelo. Desde el punto de vista de un físico, cuanto más simple sea el modelo que se haya usado, tanto mayor será la comprensión del fenómeno. Algunos físicos incluso definen el proceso de entender como el de eliminar selectivamente información poco relevante; una buena explicación debe usar la menor cantidad de datos y de suposiciones para poder reproducir cuantitativamente la mayor cantidad de resultados experimentales. ²

1 El modelado de sistemas reales

Para simular el comportamiento de un *sistema real* se necesitan:

1. Un sistema modelo (una representación adecuada del sistema real)
2. Un conjunto de reglas (las Leyes Naturales del modelo)

¹Incluso en muchos casos, e. g. en la molécula de agua, el hidrógeno cede su electrón al formar el enlace químico, de modo que al romperse el enlace se tiene un protón libre en el sistema.

²Ésta es una versión del principio filosófico atribuido al logicista y monje franciscano inglés del s. XIV Guillermo de Ockham (u Occam): *entia non sunt multiplicanda praeter necessitatem* (se puede consultar en http://en.wikipedia.org/wiki/Occams_Razor)

3. Un estado inicial del sistema modelo
4. Las influencias externas que afecten al sistema.

Dado un propósito, el modelo debe ser *atinado, preciso y eficiente*; pero además debe ser factible con los medios técnicos disponibles. Así que la respuesta (simulada) a la pregunta depende de:

1. El estado de desarrollo de las teorías pertinentes (los modelos y los métodos de solución).
2. Las capacidades computacionales disponibles.
3. Las posibilidades de implementar los métodos de solución en algoritmos.
4. Las posibilidades de validar al modelo.

2 La descripción de propiedades macroscópicas en términos moleculares

La teoría subyacente que conecta el comportamiento molecular con las propiedades macroscópicas de las fases condensadas de la materia, es la Mecánica Estadística, y hay dos modos de abordar la conexión:

Maxwell-Boltzmann: la cantidad *observable* A resulta del promedio sobre toda la trayectoria en el espacio fase, con coordenadas $\vec{q}(t)$ e ímpetus $\vec{p}(t)$,

$$\langle A \rangle_{MB} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^{\tau} A(\{\vec{q}(t)\}, \{\vec{p}(t)\}) dt \quad (1)$$

Gibbs: la cantidad *observable* A resulta del promedio ponderado sobre un *colectivo (ensemble)* de réplicas del sistema, cada una con una energía E_j . La *función de partición* \mathcal{Z} describe la distribución de las energías y depende de las condiciones termodinámicas específicas, impuestas por el entorno. En el caso en que el número de partículas N , el volumen V y la temperatura T permanecen constantes (el *colectivo canónico*, o NVT):

$$\begin{aligned} \mathcal{Z} &= \sum_j e^{-\beta E_j} \\ \langle A \rangle_G &= \frac{\sum_j A_j e^{-\beta E_j}}{\mathcal{Z}} \end{aligned} \quad (2)$$

Ambas formas de abordar la conexión deben producir idénticos resultados para un sistema en equilibrio termodinámico, o sea:

$$\langle A \rangle_{MB} = \langle A \rangle_G . \quad (3)$$

Esta afirmación se conoce como *la hipótesis ergódica* y no tiene demostración matemática, por lo que se ha abandonado la denominación de *teorema ergódico*.

3 Una Jerarquía de Modelado

En la tabla al final de esta contribución se expone esquemáticamente una jerarquía de modelado de la materia “normal” (no consideramos la posible existencia de la *materia oscura*), a partir de su constitución en núcleos atómicos, y hasta la descripción de fluidos continuos.

En este curso se tratan con más detalle los niveles 2 (la dinámica cuántica), 3 (la dinámica atómica cuántica) y 4 (la dinámica molecular), junto con el tratamiento de Monte Carlo.

4 La Química Cuántica

El objetivo de la Química es la descripción de la estructura y de las propiedades de los átomos y las moléculas. La solución de la ecuación de Schrödinger independiente del tiempo generalmente provee la información requerida. Desafortunadamente, sólo existe solución exacta en el caso del átomo de hidrógeno; por lo que es necesario recurrir a soluciones numéricas.

4.1 Soluciones aproximadas de la ecuación de Schrödinger

La ecuación de Schrödinger dependiente del tiempo tiene la forma siguiente:

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H} \Psi$$

Cuando el Hamiltoniano es función explícita sólo de las coordenadas espaciales y de un campo externo estacionario:

$$\begin{aligned} \hat{H} \psi(\vec{r}) &= E \psi(\vec{r}) \\ \Psi_n(\vec{r}, t) &= \psi_n(\vec{r}) \exp\left(-\frac{i}{\hbar} E_n t\right) \end{aligned}$$

La integración numérica se utiliza para resolver la ecuación de Schrödinger en una dimensión,

$$\frac{d^2 \psi(x)}{dx^2} = \frac{2m}{\hbar^2} [V(x) - E] \psi(x)$$

por ejemplo con los métodos de *disparo* (*shooting*), de los que se obtiene tanto el estado base como los estados excitados.

4.2 Optimizando la aproximación

Alternativamente, la solución (incógnita) se propone como combinación lineal de un conjunto base $\{\phi_n\}$ elegido adecuadamente:

$$\psi(x) = \sum_n c_n \phi_n(x)$$

Que transforma el problema a una ecuación de *eigen-valores*, con una *matriz de traslape* (*amontonamiento*)

$$\begin{aligned} \mathbf{Hc} &= \lambda \mathbf{Sc} \\ S_{mn} &= \int \phi_n^* \phi_m dx \end{aligned}$$

El *Teorema Variacional*, para una función con parámetros ajustables $\psi(\vec{r})$:

$$\langle E \rangle = \frac{\int \psi(\vec{r}) \hat{H} \psi(\vec{r}) d\vec{r}}{\int \psi^2(\vec{r}) d\vec{r}} = \left\langle \frac{\hat{H} \psi}{\psi} \right\rangle_{\psi^2} \geq E_0$$

Que puede resolverse para problemas en pocas dimensiones y con dependencia lineal de los parámetros.

Para integración multidimensional se emplea el método de *Monte Carlo*: Partiendo de una configuración inicial \vec{r}_1 con $\psi^2(\vec{r}_1) = P_1$, la “energía inicial” es

$$\epsilon_1 = \frac{\hat{H} \psi(\vec{r}_1)}{\psi(\vec{r}_1)}$$

Se repite N veces el siguiente proceso iterativo:

1. Se genera una nueva configuración \vec{r}_2 con un desplazamiento *aleatorio y corto* respecto de \vec{r}_1 .
2. Se calcula $P_2 = \psi^2(\vec{r}_2)$

3. Se acepta la nueva configuración con probabilidad $\max\{1, P_2/P_1\}$.
4. Se calcula la nueva energía ϵ_2

Se estima el valor de expectación del Hamiltoniano con el promedio:

$$\langle \hat{H} \rangle = \frac{\sum_{i=1}^N \epsilon_i}{N}.$$

4.3 La ecuación de Schrödinger para un sistema molecular

Se aplica la aproximación de Born-Oppenheimer, que proviene de un desarrollo en términos de la masa del electrón m_e dividida entre la masa nuclear (promedio) M_n . A orden cero, $M_n \rightarrow \infty$ y se considera a los núcleos fijos, lo cual se denomina *el límite adiabático*: los electrones se ajustan instantáneamente a las (nuevas) posiciones de los núcleos. El efecto de esta aproximación en el estado base del átomo de hidrógeno alcanza más de 0.7 kJ mol^{-1} ; pero en la disociación de la molécula H_2 sólo alcanza $0.072 \text{ kJ mol}^{-1}$.

Así que primero se busca la función de onda de N electrones en el campo de M núcleos, que depende de $3N$ coordenadas espaciales \vec{r}_i y N coordenadas de *spin* ω_i , que se combinan en $4N$ coordenadas $\vec{x}_i = \vec{r}_i, \omega_i$, y es antisimétrica ante intercambios impares:

$$\hat{H}\Psi = E\Psi$$

$$\hat{H} = -\sum_{i=1}^N \frac{\hbar^2}{2m} \nabla_i^2 - \sum_{i=1}^N \sum_{k=1}^M \frac{z_k e^2}{4\pi\epsilon_0 r_{ik}} + \sum_{i,j=1; i < j}^N \frac{e^2}{4\pi\epsilon_0 r_{ij}}$$

$$\Psi(\vec{x}_1, \dots, \vec{x}_i, \dots, \vec{x}_j, \dots) = -\Psi(\vec{x}_1, \dots, \vec{x}_j, \dots, \vec{x}_i, \dots)$$

El Hamiltoniano puede re-escribirse en *unidades atómicas*:

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_i \sum_k \frac{z_k}{r_{ik}} + \sum_{i < j} \frac{1}{r_{ij}}$$

Y las funciones de onda monoeléctricas se construyen como combinaciones lineales de un conjunto base (finito):

$$\psi_i(\vec{r}) = \sum_{\mu=1}^K c_{\mu i} \phi_{\mu}(\vec{r})$$

donde las funciones $\phi_\mu(\vec{r})$ son de *tipo Slater*, similares a las funciones 1s, 2s, 2p, ..., del hidrógeno, y a su vez se construyen con combinaciones de funciones Gaussianas.

Las funciones monoeléctricas se *ortonormalizan*:

$$\langle \psi_i | \psi_j \rangle = \int \psi_i^*(\vec{r}) \psi_j(\vec{r}) d\vec{r} = 1$$

y se incrementan al doble de *spin-orbitales* χ al añadirles el *spin* α o β

$$\begin{aligned} \chi_{2i-1}(\vec{x}) &= \psi_i(\vec{r}) \alpha(\omega) \\ \chi_{2i}(\vec{x}) &= \psi_i(\vec{r}) \beta(\omega) \end{aligned}$$

La función de onda electrónica antisimétrica total se construye como un *determinante de Slater*

$$\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(\vec{x}_1) & \chi_j(\vec{x}_1) & \dots & \chi_k(\vec{x}_1) \\ \chi_i(\vec{x}_2) & \chi_j(\vec{x}_2) & \dots & \chi_k(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(\vec{x}_N) & \chi_j(\vec{x}_N) & \dots & \chi_k(\vec{x}_N) \end{vmatrix}$$

Que abreviadamente se escribe como $\Psi = |\chi_i \chi_j \dots \chi_k\rangle$

Las funciones monoeléctricas ψ_i son soluciones del *operador de Fock*

$$\begin{aligned} \hat{f}(i) &= -\frac{1}{2} \nabla_i^2 - \sum_k \frac{z_k}{r_{ik}} + \mathcal{V}^{HF}(i) \\ \hat{f}(i) \chi_i(\vec{r}_i) &= \epsilon \chi(\vec{r}_i) \end{aligned}$$

en que $\mathcal{V}^{HF}(i)$ es el potencial de campo promedio que resulta de la combinación de las densidades de carga de todos los otros electrones. La solución, pues, requiere un proceso iterativo hasta que la distribución de densidad electrónica sea consistente con el potencial \mathcal{V} (el *campo autoconsistente SCF*).

Para resolver la ecuación de Fock se aplica el método variacional:

$$\frac{\int \psi'^* \hat{H} \psi' d\vec{r}}{\int \psi'^* \psi' d\vec{r}} \geq E_0$$

Se varían los coeficientes de las combinaciones lineales del conjunto base, sujeto a la condición de normalización, hasta encontrar un mínimo de $\int \chi^* \hat{f} \chi d\vec{r}$.

Al distribuir los electrones del sistema en los spin-orbitales disponibles se tiene una *configuración*.

Al aplicar el operador monoeléctrico de Fock se omitió la *energía de correlación*. Un método para incluirla es el de modificar la distribución de los electrones del sistema entre los spin-orbitales generados, creando distintas configuraciones. Es un método computacionalmente caro, por lo que es más común utilizar una aproximación perturbativa, desarrollada por Møller y Plesset.

4.4 Una alternativa: los funcionales de la densidad

La densidad de onda electrónica $\rho(r)$ determina la energía y la función de onda exactas para el estado base de un sistema de electrones; pero sólo sirve para el estado base, y la forma funcional para describir el intercambio y la correlación no puede derivarse de primeros principios.

4.4.1 La energía del sistema como funcional de la densidad electrónica

La energía y los términos que la constituyen son *funcionales* de la densidad ρ :

$$E[\rho] = V_{ne}[\rho] + K[\rho] + V_{ee}[\rho]$$

donde

$$V_{ne} = \int \rho(\vec{r}) \mathcal{V}_n d\vec{r}$$

es la interacción electrón - núcleo, \mathcal{V}_n el potencial producido por los núcleos, K es la energía cinética de los electrones, y V_{ee} es la interacción electrón - electrón.

El segundo teorema de Hohenberg y Kohn:

$$E[\rho'] \geq E[\rho]$$

Considerando un número par N de electrones *sin interacciones*, distribuidos en $\frac{1}{2}N$ orbitales, las expresiones para la energía cinética y la densidad resultan ser:

$$K_s[\rho] = \sum_{i=1} n_i \int \phi_i^* \left(-\frac{1}{2} \nabla^2 \right) \phi_i d\vec{r}$$

$$\rho[\vec{r}] = \sum_{i=1} n_i \phi_i^* \phi_i(\vec{r})$$

donde $n_i = 1$ ó 2 , es el número de ocupación del orbital ϕ_i . Las funciones de onda son soluciones de:

$$\left\{ -\frac{1}{2}\nabla^2 + \mathcal{V}_s(\vec{r}) \right\} \phi_i = \varepsilon_i \phi_i$$

en que \mathcal{V}_s no se determina aún en este paso.

La energía del sistema no interactivo es:

$$E[\rho] = K_s[\rho] + \int \mathcal{V}_s(\vec{r})\rho(\vec{r})d\vec{r}$$

Al usar la expresión de la energía para el sistema interactivo y el Hamiltoniano de las funciones de onda ϕ_i , se llega a:

$$\mathcal{V}_s(\vec{r}) = \mathcal{V}_n(\vec{r}) + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + \mathcal{V}_{xc}(\vec{r})$$

que se puede resolver para las funciones de onda de Kohn y Sham. El paso crucial es el de proponer el *potencial de intercambio y correlación* \mathcal{V}_{xc} . Hecho esto, se usa un procedimiento iterativo de campo autoconsistente.

Los primeros intentos para determinar \mathcal{V}_{xc} se hicieron a partir de la solución exacta para un gas uniforme de electrones:

$$\mathcal{V}_x^{LDA} = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \rho^{1/3}$$

El funcional de intercambio E_x resultante es:

$$E_x^{LDA}[\rho] = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int \rho^{4/3}(\vec{r})d\vec{r}$$

al que se agregan correcciones más *sofisticadas*, con derivadas de la densidad, para incluir la correlación: BLYP, B3LYP, PBE, . . .

5 La dinámica molecular

Para describir la evolución de un sistema molecular en el tiempo, se requiere de un campo de fuerzas modelo que permita un cálculo *eficiente* y *atinado* de las fuerzas que actúan sobre cada átomo, así como de la energía total del sistema. Esto es *factible* para una muestra de algunas decenas de miles de átomos, quizás

con las computadoras actuales hasta algunas centenas de miles, pero no con 10^{23} . La extensión a un sistema macroscópico se logra a través de las *condiciones a la frontera*. Las más usadas son las condiciones periódicas: se elige una celda unitaria y se consideran repeticiones en el espacio tridimensional; pueden usarse las celdas unitarias cristalinas en el caso de representar sólidos. Para simulaciones de líquidos se usan preferentemente la celda unitaria cúbica (ver Fig. 1), la rectangular, el octaedro truncado y el dodecaedro rómbico.

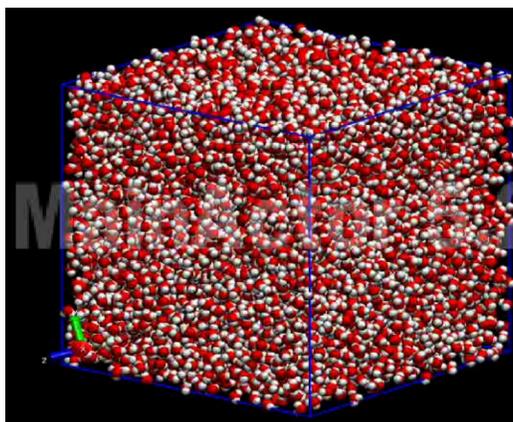


Figure 1: Celda Unitaria Cúbica.

Bajo la aproximación de Born-Oppenheimer adiabática, los núcleos se mueven en el potencial producido por los electrones, además del de los otros núcleos. Esto es razonable porque los núcleos son de 2000 a 200000 veces más masivos que los electrones. Sin embargo, dada la gran repulsión a distancias cortas, se pueden omitir algunos efectos del intercambio, pero no los de simetría.

Las vibraciones intramoleculares tienen altas frecuencias: los estados nucleares rotacionales y vibracionales son apreciables en la espectroscopía Raman y la infrarroja. Aun así, cuando los grados de libertad cuánticos de los núcleos tienen estados excitados muy por arriba del estado base, la evolución del sistema permite la aproximación adiabática: pueden separarse de los movimientos de desplazamiento molecular. Es decir, es *atinado* usar la Mecánica Clásica para describir tanto los desplazamientos como las rotaciones de las moléculas, e incluso las rotaciones de grupos de varios átomos; pero no las vibraciones intramoleculares. El criterio para discriminar cuáles grados de libertad pueden tratarse clásicamente se obtiene del cociente entre el *cuanto* vibracional del movimiento $h\nu$ y la energía térmica disponible $k_B T$: si $h\nu/k_B T \leq 1$, entonces aplica la descripción clásica; si

$h\nu/k_B T > 5$ entonces conviene mantener el grado de libertad en un valor fijo. En el intervalo $1 < h\nu/k_B T < 5$ se trata con ensayo y error.

5.1 Las ecuaciones de movimiento clásico

En la dinámica molecular clásica se usa la Segunda Ley de Newton, aplicada a cada átomo i :

$$\begin{aligned}\dot{\vec{r}}_i &= \vec{v}_i \\ \dot{\vec{v}}_i &= \vec{F}_i/m_i\end{aligned}$$

Y la conservación de la energía total $E_{tot} = K + V$:

$$\begin{aligned}\frac{dE_{tot}}{dt} &= \frac{d}{dt} \sum_i \frac{1}{2} m_i \vec{v}_i^2 + \frac{dV(\vec{r})}{dt} = \sum_i m_i \vec{v}_i \cdot \dot{\vec{v}}_i + \sum_i \frac{\partial V}{\partial \vec{r}_i} \cdot \vec{v}_i \\ &= \sum_i \left(\vec{v}_i \cdot \vec{F}_i + \frac{\partial V}{\partial \vec{r}_i} \cdot \vec{v}_i \right) = 0.\end{aligned}$$

La solución apropiada de estas ecuaciones producirá un *ensemble microcanónico NVE*. Dado que se trata de muchos átomos, no existe solución analítica, por lo que se recurre a un método numérico que considera intervalos de tiempo Δt finitos, pero suficientemente cortos como para mantener buena precisión en la predicción de la evolución temporal del sistema. El algoritmo resultante debe satisfacer tres puntos importantes :

1. La conservación de la reversibilidad en el tiempo
2. El algoritmo debe ser *simpléctico* (debe conservar el volumen en el espacio fase)
3. El costo computacional depende del cálculo de la fuerza, por lo que son preferibles los métodos con una sola evaluación de la fuerza por paso de tiempo Δt

5.2 Algoritmos de diferencias finitas

Los tres esquemas comúnmente empleados en los códigos de dinámica molecular son:

1. El algoritmo de Verlet

$$\ddot{x}_i \approx \frac{x(t - \Delta t) - 2x(t) + x(t + \Delta t)}{(\Delta t)^2}$$

$$f(t) = F_i(x(t))/m_i$$

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + f(t)(\Delta t)^2 + \mathcal{O}((\Delta t)^4)$$

$$v(t) = \frac{v(t + \Delta t) - v(t - \Delta t)}{2\Delta t} + \mathcal{O}((\Delta t)^2)$$

2. El algoritmo del *salto de rana (leapfrog; burro corrido)*, que comienza de $v(t - \frac{1}{2}\Delta t)$ y $x(t)$, con actualizaciones

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + f(t)\Delta t$$

$$x(t + \Delta t) = x(t) + v\left(t + \frac{1}{2}\Delta t\right)\Delta t$$

3. El algoritmo de *velocidades de Verlet*

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}f(t)(\Delta t)^2$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}[f(t) + f(t + \Delta t)]\Delta t$$

Los tres algoritmos producen los mismos resultados. La elección de uno de ellos es más bien una cuestión del estilo de programación.

5.3 El movimiento clásico en un potencial cuántico

Una forma de abordar el problema del campo de fuerzas es el de calcular la contribución electrónica “al vuelo”; es decir, obtener una función de onda electrónica aproximada para una configuración molecular dada. Luego se calculan las fuerzas correspondientes sobre los átomos, se resuelven las ecuaciones para un paso de tiempo y se repite el procedimiento hasta generar una trayectoria “suficientemente” larga.

5.3.1 La dinámica molecular de Car y Parrinello (CPMD)

Se hace una aproximación de la función de onda electrónica, como un desarrollo en ondas planas, en un esquema de funcionales de la densidad. Se acopla la solución cuántica aproximada a la dinámica clásica, asignando una masa ficticia a la propia función de onda electrónica, y se sigue la evolución temporal con uno de los algoritmos de diferencias finitas.

Los pasos de CPMD son los siguientes:

Se usa la densidad electrónica $n(\vec{r})$ en términos de orbitales de Kohn-Sham

$$n(\vec{r}) = \sum_i |\psi_i(\vec{r})|^2$$

teniendo como base funciones de ondas planas $\exp(i\vec{k} \cdot \vec{r})$, por lo que un punto en la superficie de energía potencial *adiabática* proviene de un mínimo del funcional de energía respecto de los orbitales Kohn-Sham:

$$E = -\frac{\hbar^2}{2m_e} \sum_i \int d\vec{r} \psi_i^*(\vec{r}) \nabla^2 \psi_i(\vec{r}) + U[n(\vec{r}); \vec{R}]$$

bajo la condición de normalización

$$\int d\vec{r} \psi_i^*(\vec{r}) \psi_j(\vec{r}) = \delta_{ij}$$

La minimización Car-Parrinello implica un sistema extendido, en que se asigna una *masa ficticia* μ a las funciones de onda y se construye una función Lagrangiana

$$\mathcal{L} = \frac{1}{2} \sum_i \int d\vec{r} |\dot{\psi}_i|^2 + \sum_I M_I \dot{\vec{R}}_I^2 - E(\psi, \vec{R})$$

lo cual origina las siguientes ecuaciones de movimiento:

$$\begin{aligned} \mu \ddot{\psi}_i(\vec{r}, t) &= -\frac{\partial E}{\partial \psi_i} + \sum_k \Lambda_{ik} \psi_k \\ M_i \ddot{\vec{R}}_I &= -\nabla_{\vec{R}_I} E \end{aligned}$$

Dado el alto costo computacional del cálculo cuántico, aun bajo el esquema de funcionales de la densidad, las simulaciones CPMD cubren tiempos de unas pocas decenas de picosegundos, y son aplicables a sistemas de un par de centenares de átomos. Además, todavía se discute la precisión de los funcionales de densidad utilizados en el cálculo de las energías intermoleculares.

5.3.2 Los sistemas mixtos

Un esquema que permite tratar un mayor número de átomos es el de calcular la función de onda electrónica sólo para una parte del sistema en cada paso de la dinámica, y acoplar el potencial cuántico estacionario con el resto del sistema a través de una frontera, o de una región de transición. El resto del sistema se trata con Mecánica Clásica. Este es el método de *Mecánica Cuántica sobre Mecánica Molecular* (QM/MM). En principio, el tratamiento de la parte cuántica puede hacerse con mayor precisión que con los funcionales de la densidad y no requiere de un desarrollo en ondas planas. El punto fino de este método es el del acoplamiento con la parte clásica. En la práctica suelen usarse soluciones cuánticas sólo hasta el nivel de Hartree-Fock, por lo que se omite la contribución de la correlación electrónica. La longitud de la trayectoria está limitada a unas cuantas centenas de picosegundos.

5.3.3 Los potenciales analíticos (*Force-fields*)

En vez de obtener la función de onda electrónica en cada paso de la simulación, es más común utilizar formas funcionales *adecuadas*, que permitan modelar las interacciones moleculares. Dichas funciones tienen parámetros que se ajustan para reproducir los valores experimentales de algunas propiedades macroscópicas del sistema, como pueden ser la densidad de un líquido y su entalpía de evaporación, o bien la entalpía de fusión del sólido correspondiente.

Los campos de fuerzas clásicos sacrifican la reproducción de las propiedades espectroscópicas vibracionales; consideran a los átomos, incluso a algunos grupos de átomos (united atom) como CH₂ y CH₃, como masas puntuales. Las contribuciones a las fuerzas se expresan como función de las posiciones atómicas:

Al contar con funciones bien definidas para representar las energías moleculares, las fuerzas que actúan sobre los átomos pueden calcularse a partir de las derivadas con respecto a las coordenadas correspondientes.

A continuación se describen algunas de las funciones más utilizadas como potenciales analíticos:

5.3.4 Términos intramoleculares

1. Los enlaces químicos se describen con osciladores armónicos

$$V_b(\vec{r}_i, \vec{r}_j) = \frac{1}{2}k_b(r_{ij} - b)^2,$$

o con un potencial cuártico,

$$V_b(\vec{r}_i, \vec{r}_j) = (k_b/8b^2) (r_{ij}^2 - b^2)^2,$$

o con un potencial de Morse

$$V_{Morse}(\vec{r}_i, \vec{r}_j) = D_{ij} [1 - \exp(-\gamma_{ij}(r_{ij} - b))]^2$$

2. Los ángulos de enlace se describen con un potencial angular armónico:

$$V_a(\vec{r}_i, \vec{r}_j, \vec{r}_k) = \frac{1}{2} k_\theta (\theta - \theta_0)^2$$

donde

$$\theta = \text{acos} \frac{\vec{r}_{ij} \cdot \vec{r}_{kj}}{r_{ij} r_{kj}}$$

o con la forma más simple

$$V_a = \frac{1}{2} k' (\cos \theta - \cos \theta_0)^2$$

5.3.5 Ángulos diedros

1. El ángulo diedro ϕ se define con las posiciones de cuatro átomos i, j, k, l , como el ángulo entre las normales \vec{n} y \vec{m} a los dos planos i, j, k y j, k, l :

$$\phi = \text{acos} \frac{\vec{n} \cdot \vec{m}}{nm}$$

donde

$$\begin{aligned} \vec{n} &= \vec{r}_{ij} \times \vec{r}_{kj} \\ \vec{m} &= \vec{r}_{jk} \times \vec{r}_{lk}. \end{aligned}$$

El potencial del diedro se describe con una función periódica:

$$V_d(\phi) = k_\phi (1 + \cos(n\phi - \phi_0))$$

que se afecta con un potencial 1-4, o bien otras funciones, como el potencial de *Ryckaert-Bellemans*

$$V_{RB}(\phi) = \sum_{n=0}^5 C_n \cos^n \phi$$

2. Los *diedros impropios* sirven para mantener grupos planos, y se les asigna un potencial armónico de restricción:

$$V_{imp}(\xi) = \frac{1}{2}k_{\xi}(\xi - \xi_0)^2$$

5.3.6 Las interacciones no ligadas

1. La repulsión de corto alcance y la dispersión de alcance medio suelen representarse con el potencial de Lennard-Jones:

$$\begin{aligned} V_{LJ} &= \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6} \\ &= 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] \end{aligned}$$

O bien con el de Buckingham:

$$V_B = A \exp(-Br) - \frac{C_6}{r^6}$$

2. Las interacciones Coulómbicas:

$$V_C = f_{el} \frac{q_i q_j}{\epsilon_r r}$$

En que generalmente $f_{el} = (4\pi\epsilon_0)^{-1}$ y ϵ_r es una constante dieléctrica relativa, que suele hacerse igual a 1.

6 El método de Monte Carlo

Así como el tratamiento de colectivos de Gibbs es una alternativa al seguimiento de la trayectoria de Maxwell-Boltzmann, el método de Monte Carlo es una alternativa a la dinámica molecular.

6.1 Fundamentos

El teorema del valor medio para integrales: *Si f es una función continua en el intervalo $[a, b]$ entonces existe $z \in [a, b]$ tal que*

$$f(z)(b-a) = \int_a^b f(x)dx$$

El *valor promedio* de la función $f(x)$ sobre su dominio se define como:

$$\langle f(x) \rangle = f(z) = \frac{1}{b-a} \int_a^b f(x) dx$$

Y el problema de la integración se puede sustituir por el de encontrar un estimador válido del promedio: Para calcular la integral

$$I = \int_a^b g(x) dx$$

se la puede representar como el valor de expectación de alguna variable aleatoria:

$$I = \int_a^b \frac{g(x)}{f_X(x)} f_X(x) dx$$

suponiendo que $f_X(x)$ es una función de distribución de probabilidad,

$$I = \left\langle \frac{g(x)}{f_X(x)} \right\rangle$$

6.2 La minimización de la varianza

El objetivo es minimizar la varianza del *estimador* ζ usando la variable aleatoria X :

$$\begin{aligned} \zeta &= \frac{g(X)}{f_X(X)} \\ \text{var} \zeta &= \int \frac{g^2(x)}{f_X(x)} dx - I^2 \end{aligned}$$

El estimador del promedio es:

$$\langle \zeta \rangle = \frac{1}{N} \sum_{i=1}^N \frac{g(X_i)}{f_X(X_i)}$$

Y la varianza mínima (nula) ocurre cuando

$$f_X(x) = \frac{g(x)}{I}$$

En la práctica, se busca que $f_X(x)$ tenga una forma similar a la de $|g(x)|$. Esto es el *muestreo de los estados importantes* (*importance sampling*).

6.3 El algoritmo de Metropolis

Éste es el algoritmo que se emplea para la mecánica estadística de sistemas moleculares. El objetivo es generar un colectivo con probabilidades canónicas:

$$w(\vec{r}) \sim e^{-\beta V(\vec{r})}$$

Considerando el paso de \vec{r} a $\vec{r}' = \vec{r} + \Delta\vec{r}$, con energías potenciales respectivas

$$\begin{aligned} E &= V(\vec{r}) \\ E' &= V(\vec{r}'), \end{aligned}$$

El muestreo debe ser homogéneo sobre el espacio de configuraciones. Se busca que las *probabilidades de transición* W_{\rightarrow} de \vec{r} a \vec{r}' y W_{\leftarrow} de \vec{r}' a \vec{r} cumplan la *condición de balance detallado*:

$$w(\vec{r})W_{\rightarrow} = w(\vec{r}')W_{\leftarrow}$$

Que lleva al cociente

$$\frac{W_{\rightarrow}}{W_{\leftarrow}} = \frac{w(\vec{r}')}{w(\vec{r})} = e^{-\beta(E' - E)}$$

Lo cual sugiere una probabilidad de aceptar la nueva configuración:

$$\begin{aligned} E' - E \leq 0 &\Rightarrow W_{\rightarrow} = p^{ac} = 1 \\ E' - E > 0 &\Rightarrow W_{\rightarrow} = p^{ac} = e^{-\beta(E' - E)} \end{aligned}$$

7 La comparación md vs. MC

La dinámica molecular representa al colectivo *microcanónico*, en que el número de partículas N , el volumen V y la energía total del sistema E se mantienen fijos (colectivo NVE), en tanto que las simulaciones por Monte Carlo representan al colectivo canónico NVT. En cambio los experimentos suelen realizarse bajo condiciones en que la presión P se mantiene constante, o sea, el colectivo *gran canónico* NPT.

En el caso del método de Monte Carlo, el control de la presión se logra agregando al criterio de aceptación de una nueva configuración un término

$$N \ln \frac{V_n}{V_o} - \beta P \Delta V$$

lo cual toma en cuenta tanto el cociente como la diferencia entre el volumen de la nueva configuración V_n y el volumen de la anterior V_o , así como el trabajo de expansión $P\Delta V$, y el control es exacto.

En la dinámica molecular, el control de la temperatura y de la presión requiere de métodos más complicados, como son:

- Los métodos estocásticos para imponer una temperatura al sistema.
- Los métodos de acoplamiento fuerte (p. ej. el escalamiento de velocidades).
- Los métodos de acoplamiento débil (Berendsen, Andersen).
- La dinámica de sistema extendido (Nosé-Hoover).

Parecería entonces que en general MC tiene ventaja sobre md; sin embargo, el muestreo MC se hace generando perturbaciones aleatorias a partir de una configuración inicial arbitraria. Puesto que cada configuración subsecuente es una perturbación, la convergencia al equilibrio es muy lenta, en tanto que en cada paso de md todas las partículas del sistema se mueven en respuesta a las fuerzas que las urgen, logrando converger al equilibrio más rápidamente.

8 Jerarquía de Modelado

NIVEL 1: LA DINÁMICA CUÁNTICA RELATIVISTA		
Sistema		Reglas
Núcleos atómicos (masa, carga, <i>spin</i>)		Mecánica Cuántica Relativista
Electrones (masa, carga, <i>spin</i>)		dependiente del tiempo
Fotones (frecuencia)		La ecuación de Dirac
		Electrodinámica (cuántica)
Aproximación		No aplicable a:
Velocidades de las partículas	⇓	Electrones cerca de
mucho menores que c	⇓	núcleos pesados
	⇓	Plasmas calientes

NIVEL 2: LA DINÁMICA CUÁNTICA

Sistema	Reglas
Núcleos atómicos	Las ecuaciones de Schrödinger no relativistas dependiente e independiente del tiempo
Electrones	
Fotones	Las ecuaciones de Maxwell

Aproximación	↓	No aplicable a:
Electrones mucho más rápidos que los núcleos (Born-Oppenheimer)	↓	Dinámica electrónica
	↓	(e. g. en semiconductores)
	↓	Procesos rápidos de transferencia electrónica
	↓	Comportamiento dinámico de estados excitados
	↓	

NIVEL 3: LA DINÁMICA ATÓMICA CUÁNTICA

Sistema	Reglas
Átomos	Los átomos se mueven en un potencial efectivo producido por los electrones
Iones	
Moléculas	Los átomos pueden obedecer la ec. de Schrödinger dependiente del tiempo
Fotones (gas)	

Aproximación	↓	No aplicable a:
El movimiento atómico es clásico	↓	Transferencia de protones
	↓	H₂ y He a bajas temperaturas
	↓	Reacciones rápidas
	↓	Movimientos de altas frecuencias

NIVEL 4: LA DINÁMICA MOLECULAR

Sistema	Reglas
Materia condensada: (Macro)moléculas, fluidos, soluciones, cristales líquidos, reacciones lentas	La Mecánica Clásica (ecs. de Newton) La Mecánica Estadística La dinámica molecular

Aproximación	↓	No aplicable a:
Reducción del número de grados de libertad	↓	Los detalles de una dinámica rápida Las propiedades de transporte

NIVEL 5: LA DINÁMICA GENERALIZADA DE LANGEVIN EN SISTEMAS REDUCIDOS

Sistema	Reglas
Materia condensada: Agregados moleculares grandes Polímeros Defectos en sólidos Reacciones muy lentas	<i>Superátomos</i> , coordenadas de reacción Promedios sobre el equilibrio local Dinámica con constricciones Energías libres y potenciales de campos promedio

Aproximación	↓	No aplicable a:
Omisión de correlaciones temporales y/o espaciales en las fluctuaciones	↓	Correlación del movimiento Alta precisión en tiempos cortos

NIVEL 6: LA DINÁMICA DE LANGEVIN SIMPLE

Sistema	Reglas
Procesos y reacciones “lentos”, fuera del equilibrio	Aceleraciones deducidas de fuerzas sistemáticas, fricción y ruido Las ecs. de Fokker-Planck

Aproximación	↓	No aplicable a:
Omisión de términos inerciales: “grano grueso” temporal	↓	Los detalles dinámicos

NIVEL 7: LA DINÁMICA BROWNIANA

Sistema	Reglas
Procesos de “grano grueso” fuera del equilibrio	Velocidades que resultan de las fuerzas y la fricción, más el ruido
Sistemas poliméricos	Dinámica Browniana (difusiva)
Sistemas coloidales	Relaciones flujo/fuerza de Onsager

Aproximación	↓	No aplicable a:
Reducir la descripción a densidades continuas de las especies constituyentes	↓	Los detalles de las partículas
	↓	
	↓	

NIVEL 8: LA DINÁMICA MESOSCÓPICA

Sistema	Reglas
Similar al nivel 7: sistemas auto-organizables; sistemas reactivos fuera del equilibrio	Descripción en términos de densidades: conservación de masa, más la ec. de dinámica del flujo, con ruido

Aproximación	↓	No aplicable a:
Promedio sobre un número “infinito” de partículas	↓	La formación espontánea de estructura debida a fluctuaciones
	↓	

NIVEL 9: LA DINÁMICA REACTIVA DE FLUIDOS

Sistema	Reglas
Mezcla macroscópica de diferentes especies fuera del equilibrio (como la atmósfera para la predicción del clima)	Conservación de masa, de energía y de ímpetu Flujos reactivos

Aproximación	↓	No aplicable a:
Reducir a una especie con viscosidad Newtoniana de las especies constituyentes	↓	Procesos reactivos
	↓	Comportamiento no Newtoniano
	↓	

NIVEL 10: LA DINÁMICA DE FLUIDOS

Sistema	Reglas
Fluidos macroscópicos fuera del equilibrio: gases y líquidos	Conservación de masa, de energía y de ímpetu La ec. de Navier-Stokes
Aproximación Fluidos a bajas velocidades (Números de Reynolds bajos)	↓ ↓ No aplicable a: La turbulencia

NIVEL 11: LA DINÁMICA DE FLUIDOS EN FLUJO ESTACIONARIO

Sistema	Reglas
Fluidos macroscópicos fuera del equilibrio, con flujo laminar	La ec. de Navier-Stokes simplificada