ESCUELA DE VERANO EN FÍSICA EN LA UNAM

junio 13-24, 2016

Instituto de Ciencias Físicas, UNAM, 13-17 de junio Instituto de Física, UNAM, 20-24 de junio

Editores

José Récamier Rocío Jáuregui



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO 2017

ESCUELA DE VERANO EN FÍSICA EN LA UNAM, Año 1, Núm. 1, enero-diciembre 2017, es una publicación anual editada por la Universidad Nacional Autónoma de México, Ciudad Universitaria, Delegación Coyoacán, C.P. 04510, Ciudad de México, a través del Instituto de Ciencias Físicas UNAM Campus Morelos, Av. Universidad s/n, Col. Chamilpa, Cuernavaca, Morelos, 62210, México. Teléfono + 52 777 329 17 72, https://www.fis.unam.mx/, pepe@fis.unam.mx. Editor responsable: Dr. José Récamier Angelini. Reserva de Derechos al uso Exclusivo No. 04-2017-110313451400-203, ISSN: en trámite, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número, Instituto de Ciencias Físicas UNAM Campus Morelos, Dr. José Récamier Angelini, Av. Universidad s/n, Col. Chamilpa, Cuernavaca, Morelos, 62210, México, fecha de la última modificación, 31 de noviembre de 2017, 200 ejemplares en CD-ROM.

El contenido de los artículos es responsabilidad de los autores y no refleja el punto de vista de los árbitros, del Editor o de la UNAM. Se autoriza la reproducción total o parcial de los textos aquí publicados siempre y cuando se cite la fuente completa y la dirección electrónica de la publicación.

CONTENIDO

Contribuciones	V
Introducción	VII
Agradecimientos	IX
Profesores participantes	XI
Alumnos participantes	XIII

CONTRIBUCIONES

	Maximino Aldana González, Instituto de Ciencias Físicas UNAM, Redes Complejas: Estructura, Dinámica y Evolución	1
—	F. Castillo, V. U. L. Contreras, B. Campillo, O. Flores y H. Martínez, Instituto de Ciencias Físicas UNAM, <i>Estudio y aplicación de los plasmas de baja temperatura</i>	23
—	Juan Carlos Degollado Daza, Instituto de Ciencias Físicas UNAM, Agujeros negros, superradianza y extracción de energía	49
—	Agustín E. González, Instituto de Ciencias Físicas UNAM, Dinámica de las fronteras de grano en la cristalización coloidal	57
	Juan Carlos Hidalgo Cuéllar, Instituto de Ciencias Físicas UNAM, Formación de estructura con un campo escalar	65
_	Antonio Juárez Reyes, Instituto de Ciencias Físicas UNAM, La relación entre el conocimiento fundamental, generado por la Ciencia y su impacto en la economía de un país	77
	Lucila Juárez y W. Luis Mochán, Instituto de Ciencias Físicas UNAM, <i>Respuesta magnética de metamateriales</i>	89
	Gloria Koenigsberger Horowitz, Instituto de Ciencias Físicas UNAM, La exploración del cosmos y los telescopios espaciales	117
—	Gustavo Martínez Mekler, Instituto de Ciencias Físicas UNAM, Criticalidad en la dinámica de redes de señalización relevantes a la fecundación	143
_	Frederic Masset, Instituto de Ciencias Físicas UNAM, Simular en computadoras la evolución de fluídos: ejemplos y estrategias en astrofísica	159
	Ricardo Román Ancheyta y José Récamier Angelini, Instituto de Ciencias Físicas UNAM,	
	Efecto Casimir Dinámico	165

INTRODUCCIÓN

La XXIV Escuela de Verano en Física fue organizada con el apoyo del 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 Ciencias Físicas en Cuernavaca, Morelos, del 13 al 17 de junio y en las instalaciones del Instituto de Física en Ciudad Universitaria del 20 al 24 de junio.

En esta Escuela se impartieron 8 cursos cortos de cinco horas de duración cada uno y 20 conferencias. Los cursos y conferencias cubrieron un amplio espectro con temas como óptica cuántica, sistemas complejos, óptica de meta-materiales, biofísica molecular, plasmas, redes complejas, física médica, estructuras fotónicas, química cuántica y física atómica, entre otros.

José Récamier, ICF UNAM Mayo, 2017

AGRADECIMIENTOS

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 y del Instituto de Ciencias Físicas. Agradecemos al programa de Becas de Movilidad Santander-Universia y a las Becas ECOES.

PROFESORES PARTICIPANTES

- W. Luis Mochan, Instituto de Ciencias Físicas UNAM, Óptica de meta-materiales
- Iván Ortega Blake, Instituto de Ciencias Físicas UNAM, Biofísica molecular
- Maximino Aldana González, Instituto de Ciencias Físicas UNAM, *Redes neuronales complejas*
- Horacio Martínez Valencia, Instituto de Ciencias Físicas UNAM, Aplicaciones de los plasmas fríos
- Gustavo Martínez Mekler, Instituto de Ciencias Físicas UNAM, *Redes regulatorias de la fecundación*
- Antonio Juárez Reyes, Instituto de Ciencias Físicas UNAM, Aplicaciones de física molecular en medicina y monitoreo del agua
- Jesús Antonio del Río Portilla, Instituto de Energías Renovables UNAM, *Optimización de estructuras fotónicas*
- Alejandro Ramírez Solís, Facultad de Ciencias Universidad Autónoma del Estado de Morelos, *Estudios teóricos de las fases epsilon y dzeta del oxígeno sólido a altas presiones*
- Juan Carlos Hidalgo Cuéllar, Instituto de Ciencias Físicas UNAM, *Cosmología con agujeros negros primordiales*
- Fréderic Masset, Instituto de Ciencias Físicas UNAM, Simular en computadoras la evolución de fluídos: ejemplos y estrategias en astrofísica
- Sergio Cuevas García, Instituto de Energías Renovables UNAM, Inestabilidades magnetohidrodinámicas en flujos de metal líquido
- Remigio Cabrera Trujillo, Instituto de Ciencias Físicas UNAM, Métodos numéricos para resolver problemas de física atómica
- Agustín González Flores, Instituto de Ciencias Físicas UNAM, Cristalización coloidal con fuerzas atractivas de corto alcance
- Gloria Koenigsberger Horowitz, Instituto de Ciencias Físicas UNAM, *La exploración del cosmos con telescopios espaciales*

- Hernán Larralde Ridaura, Instituto de Ciencias Físicas UNAM, Caminatas aleatorias
- Humberto Saint Martin Posada, Instituto de Ciencias Físicas UNAM, *Estudios de metaloides y metales contaminantes de agua*
- José Recamier Angelini, Instituto de Ciencias Físicas UNAM, Efecto Casimir dinámico
- Juan Carlos Degollado Daza, Instituto de Ciencias Físicas UNAM, *Agujeros negros, superradianza y nueva física*
- Rocío Jáuregui, Instituto de Física UNAM, Óptica Cuántica e información cuántica
- Oliver Paz Borbón, Instituto de Física UNAM, Nanomateriales: un enfoque práctico
- Mercedes Rodríguez, Instituto de Física UNAM, Formación de imágenes tomográficas
- Carlos Pineda, Instituto de Física UNAM, Decoherencia e información cuántica
- Juan Adrián Reyes, Instituto de Física UNAM, Fenómenos ópticos en cristales líquidos
- Rosario Paredes, Instituto de Física UNAM, Materia cuántica ultrafría
- Carlos Villagómez, Instituto de Física UNAM, *Espectroscopia óptica y manipulación de moléculas individuales en superficies*
- Genaro Toledo, Instituto de Física UNAM, Símetrías en partículas elementales
- Matías Moreno, Instituto de Física UNAM, Una fuente de luz sincrotrón en México
- Shahen Hacyan, Instituto de Física UNAM, Ondas gravitacionales
- Carlos Villarreal, Instituto de Física UNAM, *Fluctuación y disipación en la física cuántica*
- Daniel Sahagún, Instituto de Física UNAM, Atrapando átomos con microchips
- Laura Serkovic, Instituto de Física UNAM, *El gráfeno y sus posibilidades de investi*gación en el IFUNAM
- Francisco Javier Sevilla Pérez, Instituto de Física UNAM, *Mecánica estadística de materia activa*
- Roberto C. Muñoz, Instituto de Ciencias Físicas UNAM, *Parámetros biofísicos de las bicapas lipídicas celulares y las bicapas modelo*
- G. García Naumís, Instituto de Física UNAM, *Teoría de campos aplicada a la nano*tecnología
- María Ester Brandán, Instituto de Física UNAM, Resta de imágenes médicas
- Rafael Pérez Pascual, Instituto de Física UNAM, Galileo y el pensamiento científico
- Luis Armando Acosta Sánchez, Instituto de Física UNAM, Astrofísica nuclear

ALUMNOS PARTICIPANTES

- José Luis Gómez Morales, Facultad de Ciencias, Universidad Autónoma del Estado de México, jgm0094@gmail.com
- Ronny Abdiel González Hernández, Facultad de Estudios Superiores Aragón, UNAM, agh.69@hotmail.com
- David Valdés Corona, Facultad de Ciencias, UNAM, corona@ciencias.unam.mx
- Luis Alberto Razo López, Universidad de Guadalajara, luis.a.razo@hotmail.com
- Juan Carlos Obeso Jureidini, Facultad de Ciencias, UNAM, juan-e22@ciencias.unam.mx
- Norman Wilfrido Molina González, Universidad de Guadalajara, normanmolina@hotmail.com
- Samuel Hidalgo Caballero, Benemérita Universidad Autónoma de Puebla, alpino.shc@hotmail.com
- Rocío Belén Lucero Pennacchio, Facultad de Ciencias Exactas y Naturales, Universidad Nacional del Cuyo, Argentina, ro lucero@hotmail.com
- Silvia Fernanda Cárdenas López, Facultad de Ciencias, UNAM, silvia.crdns@ciencias.unam.mx
- Selma Kuri Hernández, Benemérita Universidad Autónoma de Puebla, karim1594_@hotmail.com
- Benjamín Medina Carrillo, Universidad Autónoma de Yucatán, benjamin.medina.carrillo@gmail.com
- Reynaldo Ortiz Guerrero, Facultad de Ciencias, UNAM, reyorguer@ciencias.unam.mx
- Rodrigo Itzamná Becerra Deana, Benemérita Universidad Autónoma de Puebla, rodro_d94@hotmail.com
- Iván Kaleb Ramírez Torres, Facultad de Estudios Superiores Cuautitlán, UNAM, rtkaleb@gmail.com
- Evelyn Álvarez Cruz, Facultad de Ciencias, UNAM, evelyn_333@ciencias.unam.mx

- Humberto Torres Bustamante, Facultad de Ciencias, UNAM, lfhumberto@ciencias.unam.mx
- Mireyly Estefania Pérez Hernández, Universidad Autónoma de Yucatán, meph2392@gmail.com
- Julio César Quiñones Pérez, Facultad de Ciencias, UNAM, julio.quinones@correo.nucleares.unam
- Analyz Brito Ruiz, Universidad Autónoma del Estado de Morelos, any.liz.azul@gmail.com
- Leonardo Daniel Orozco Colín, Universidad Autónoma del Estado de México, FC, leonardJan30@gmail.com
- Raúl Maximiliano Urrutia Hernández, Benemérita Universidad Autónoma de Puebla, uhrmax93@hotmail.com
- Esli Daniel Morales Tehuitzitl, Benemérita Universidad Autónoma de Puebla, copercot@gmail.com
- Juan Antonio Rivera, Facultad de Ciencias, UNAM.
- Jorge Emmanuel Chávez Saab, Massachusetts Institute of Technology.
- Oswaldo Martínez Vázquez, Universidad Autónoma del Estado de Morelos.
- Christian Alejandro Mercado Ornelas, Universidad de Zacatecas.

Maximino Aldana

I. INTRODUCCIÓN

f

En los últimos 17 años hemos sido testigos de una explosión en el estudio de las propiedades estructurales y dinámicas de las redes complejas. Durante este tiempo se han publicado cientos de artículos sobre este tema en revistas de investigación científica internacionales de diferentes disciplinas, que abarcan física, biología, sociología, neurología, economía, medicina, por mencionar algunos ejemplos. Este interés en las redes complejas radica en que nos hemos dado cuenta de que dichas redes abundan en la naturaleza, son parte de nuestra vida diaria y se presentan a diferentes niveles de organización. Por ejemplo, algunas redes biológicas que encontramos en el nivel microscópico son las redes de regulación genética, redes de proteínas, redes neuronales, redes metabólicas. Por otro lado, a un nivel de organización mucho mayor, encontramos redes de comunicación e informáticas (la red internet, la red www, redes telefónicas, etc.), redes sociales (amistades, contactos sexuales, colaboradores científicos, propagación de enfermedades, etc.), redes ecológicas (interacciones tróficas en un ecosistema). Las redes complejas son ubicuas, están por todos lados. Incluso se ha estudiado la red de super héroes en el Universo de Marvel, siendo el hombre araña el super héroe más popular con la mayor conectividad [1]. Es un hecho sobresaliente el que todas estas redes, tan diferentes en naturaleza y en tama no, tengan muchas propiedades estructurales similares. Este hecho, tan simple como sorprendente, hace posible que podamos formular modelos matemáticos para entender y explicar las propiedades estructurales (y en algunos casos también las propiedades dinámicas) de las redes complejas.

En estas notas presento algunas de las herramientas matemáticas y conceptuales que se han desarrollado a lo largo de varios a nos para analizar la estructura y dinámica de las redes complejas. Los que estén interesados podrán encontrar un estudio mucho más completo del que presento aquí en las referencias [2–5], las cuales son trabajos de revisión excelentes que sirven como punto de partida para adentrarse en el fascinante mundo de las redes complejas.

II. ALGUNAS DEFINICIONES

A. Sistemas complejos

Comenzemos nuestro estudio describiendo qué son los sistemas complejos. Como ocurre con la gran mayoría de los conceptos científicos, no podemos definir los sistemas complejos en un simple enunciado. En lugar de ésto, vamos a enumerar las características más importantes que son comunes a todos los sistemas complejos:

- 1. Están compuestos de muchas partes que interactúan entre sí. De hecho, el adjetivo "Complejo" en este contexto no significa solamente que el sistema sea complicado, sino también que está compuesto de muchas partes, como un *complejo* industrial.
- 2. Cada parte tiene su propia estructura interna y está encargada de llevar a cabo una función específica.
- 3. Lo que ocurra a una parte del sistema afecta de manera **altamente no lineal** a todo el sistema.
- 4. Presentan **comportamientos emergentes**, de tal manera que el *el todo no es la simple suma de sus partes*.

Como un ejemplo típico de sistema complejo consideremos a la célula. Evidentemente la célula está compuesta de muchas partes (ribosomas, mitocondrias, núcleo, membrana, retículo endoplasmático, ADN, ARN, etc.), v cada una de estas partes se encarga de realizar alguna función específica dentro de la célula. Las partes de la célula responden de forma no lineal ante perturbaciones externas. Por ejemplo, algunas veces una mutación en el ADN no tiene ningún efecto en la célula, mientras que otras veces una sóla mutación puede ser fatal[12]. Además, la célula presenta comportamientos emergentes que no pueden explicarse en términos de las propiedades de sus partes individuales. Así, podemos hablar de una célula enferma, pero no podemos decir que un ribosoma o una proteína estén enfermos. La enfermedad es una propiedad que emerge como resultado de la organización colectiva de todos los constituyentes de la célula.

B. Redes Complejas

Las redes complejas son conjuntos de muchos nodos conectados que interactúan de alguna forma. A los nodos de una red también se les llama vértices o elementos y los representaremos por los símbolos $v_1, v_2, ..., v_N$, donde N es el número total de nodos en la red. Si un nodo v_i está conectado con otro nodo v_j , esta conexión se representa por una pareja ordenada (v_i, v_j) . La definición matemática de una red (también llamada grafo por los matemáticos) es la siguiente:

Definición 1 Una red \mathcal{R} consiste de un conjunto de nodos $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}, y$ un conjunto de parejas ordenadas $\mathcal{E} = \{(v_i, v_j)\} \subset \mathcal{V} \times \mathcal{V}$. Cada pareja ordenada (v_i, v_j) se llama **conexión dirigida** del nodo v_i al nodo v_j . La red \mathcal{R} se llama **no dirigida** si para cada pareja $(v_i, v_j) \in \mathcal{E}$ también existe la pareja $(v_j, v_i) \in \mathcal{E}$. De lo contrario, la red se denomina **dirigida**. Llamaremos a todos los nodos que estén conectados directamente a un nodo v_i , los **vecinos** de v_i . Finalmente, el número k_i de vecinos del nodo v_i (es decir, el número de conexiones de v_i) se llama la **conectividad** de v_i , y el promedio de estas conectividades, $\langle k \rangle = N^{-1} \sum_{i=1}^{N} k_i$, es la **conectividad de la red**.

Aunque la definición formal de una red es útil en el desarrollo matemático de la teoría, para nuestros propósitos basta con considerar que una red es un montón de nodos entre los que existen coneciones. En la naturaleza se pueden encontrar muchos tipos de redes, es decir, muchos tipos de nodos y conexiones. Por ejemplo, en una red social los nodos son las personas y las conexiones pueden ser los lazos de amistad que existan entre ellas: dos personas están conectadas si son amigos. En la misma sociedad podemos definir las conexiones de forma distinta, por ejemplo, dos personas están conectadas si han tenido relaciones sexuales. Claramente, la red definida a través de amistades es diferente a la red definida a través de contactos sexuales, ya que el hecho de que dos personas sean amigas no significa que hayan tenido relaciones sexuales, y viceversa.

Notemos entonces que incluso en un mismo conjunto de nodos podemos definir redes diferentes dependiendo de como hayamos definido las conexiones, lo cual, por supuesto, depende del fenómeno que nos interese estudiar. Por ejemplo, si estuviésemos interesados en analizar como se propaga una enfermedad como el SIDA en una sociedad, claramente nos convendría estudiar la red de interacciones sexuales, mientras que si estamos interesados en encontrar a un asesino, lo que nos conviene es estudiar la red de amistades, ya que son sus amigos los que pueden darnos información sobre su paradero. El Cuadro I muestran diferentes tipos de redes que se encentran en la naturaleza.

En los ejemplos anteriores hay redes dirigidas y redes no dirigidas. Por ejemplo, la red de contactos sexuales es no dirigida, ya que si A tuvo relaciones con B, entonces evidentemente B tuvo relaciones con A. Pero la red de transmisión de la gripe es dirigida, ya que si A contagió de gripe a B, no necesariamente B contagió también a A. Otra red dirigida es la World Wide Web (WWW). En mi página web yo tengo una liga a la página del periódico la jornada, pero en la página de la jornada no hay ninguna liga a mi página web. En este sentido, hay una conexión de mi página hacia la de la jornada, pero no hay una conexión de regreso.

Intuitivamente, una red no dirigida puede pensarse como aquella en la que las conexiones entre los nodos siempre son simétricas (si A está conectado con B, entonces B está conectado con A), mientras que en una red dirigida no todas las conexiones son simétricas, es decir, siempre

Redes sociales						
Sexuales		Dos personas están conectadas				
		si han tenido por lo menos una				
		relación sexual				
Actores		Dos actores están conectados si				
		han aparecido en la misma película				
Amistades		Dos personas están conectadas si				
		son amigas				
Científicos		Dos científicos están conectados si han				
		sido coautores en algún artículo				
Familiares		Dos personas están conectadas si				
		son familiares cercanos				
Enfermedades		Dos personas están conectadas si una				
		contagió de una enfermedad a la otra				
		Redes informáticas				
Internet	Dos	s computadoras están conectadas si				
Internet	hay	un cable que las conecta				
WWW	Dos	Dos páginas web están conectadas si				
	hay	hay un hipervínculo de una a la otra				
	Dos	Dos palabras están conectadas si en				
Palabras	el d	el diccionario una aparece en la definición				
	de l	de la otra				
Dalahwag	Dos	s palabras están conectadas si son				
Falabias	sinónimos					
Redes biológicas						
Protóicos	D	Dos proteínas están conectadas si				
Tioteicas	p	participan en la misma reacción química				
Genéticas I		Dos genes están conectados si uno				
		egula la expresión del otro				
Ecológicas I s		os especies están conectadas				
		una se come a la otra				
Neuronales		os neuronas están conectadas si existe				
		na conexión sináptica entre ellas				

Cuadro I. Diferentes tipos de redes.

existen conexiones asimétricas (A está conectado con B pero B no está conectado con A).

Otro concepto importante es el de *islas* (o *sub redes*) de una red. Notemos que la definición de red que dimos arriba **no dice** que todos los nodos deben estar conectados unos con otros. Ni siguiera dice que todos los nodos deben tener conexiones. La definición matemática nos dice que una red es un conjunto de nodos entre los que existen algunas conexiones. Esto quiere decir que en la red pueden existir nodos que no tengan conexiones, es decir, nodos aislados. También pueden existir grupos de nodos que estén conectados entre sí pero que no estén conectados con el resto de la red. Como un ejemplo concreto pensemos en una red social en la que dos individuos están conectados sin son familiares cercanos (específicamente, hermanos, medios hermanos, primos, padres, hijos, esposos, tíos, sobrinos, abuelos y nietos). Esta es evidentemente una red no dirigida, va que si A es familiar de B, entonces B también es familiar de A. Sin embargo, muy probablemente en una sociedad grande esta red estaría fracturada en islas o sub redes, debido a que claramente no todas las personas en una sociedad son familiares cercanos de todos los demás. En mi caso particular, mi apellido es "Aldana González", por lo que es natural pen-



Figura 1. Una red puede estar compuesta de varias islas, como en el ejemplo mostrado en esta figura. La isla más granda se denomina la *isla gigante*.

sar que estoy conectado directamente con algunas de las familias Aldana y algunas las familias González en México. Pero muy probablemente ningún miembro de mi familia está conectado con alguien de la familia Azcárraga, o con la familia Zabludovsky. Por lo tanto, los miembros de mi familia conforman una isla o sub red dentro de la cual estamos conectados entre nosotros, pero esta isla está desconectada de otras familias de la sociedad. La Fig. 1 muestra un ejemplo de una red compuesta de varias islas.

Las islas en una red pueden tener diferentes tama nos, que van desde 1 (un sólo nodo que no está conectado a nadie) hasta el tama no de toda la red (todos los nodos están conectados con todos), en cuyo caso la red consiste de una sóla isla, que es ella misma. Es importante enfatizar que el hecho de que una isla no esté conectada al cuerpo principal de la red no significa que dicha isla no pertenezca a la red. La red no está determinada sólo por las conexiones, sino también por los nodos que conforman al sistema. Esto puede parecer poco intuitivo, pero desde el punto de vista matemático es conveniente considerar que todos los nodos del sistema pertenecen a la red, independientemente de que haya o no conexiones entre ellos. Como veremos más adelante, las islas juegan un papel importante en la teoría de redes.

III. ESTUDIO DE LAS REDES COMPLEJAS

Ejemplos de redes dirigidas y no dirigidas abundan en la naturaleza. Ahora ustedes pueden comenzar a pensar en la red que más les guste, ya sea una red biológica, social, informática o cualquier otra. Las redes se presentan en diferentes tamaos, colores y sabores. Pero, ¿qué hacemos con las redes? ¿Cómo las estudiamos?



Figura 2. El estudio de las redes complejas puede dividirse en dos partes: (a) el estudio de sus propiedades estructurales y (b) el estudio de sus propiedades dinámicas.

El estudio general de las redes complejas puede dividirse en dos campos diferentes y complementarios: *Estructura* y *Dinámica* (ver la Fig. 2). En el primer campo de estudio uno está interesado en determinar las propiedades estructurales (o topológicas) de la red, es decir, en las propiedades que nos dicen cómo están conectados los nodos unos con otros. Algunas de las propiedades más importantes que determinan la estructura (o topología) de una red son las siguientes:

- 1. La distribución de conexiones (o vecinos) P(k): Es la probabilidad de que un nodo escogido al azar tenga k conexiones (o vecinos). Por ejemplo, en una red de contactos sexuales P(k) es la probabilidad de que una persona escogida al azar en una sociedad haya tenido k parejas sexuales distintas a lo largo de su vida.
- El coeficiente de agregación C: Es la probabilidad de que dos nodos conectados directamente a un tercer nodo, estén conectados entre sí (ver la Fig. 3(a)). Por ejemplo, en una red de amistades, es la probabilidad de que dos de mis amigos sean ellos mismos amigos uno del otro.
- 3. La longitud mínima L_{ij} entre dos nodos v_i y v_j : Es el número mínimo de "brincos" que se tienen que dar para llegar de un nodo v_i de la red a otro nodo v_j de la red. Por ejemplo, en la red mostrada en la Fig. 3(b), aunque existen varios caminos para llegar de v_i a v_j , el camino mínimo consiste de tres pasos (indicado con líneas gruesas).
- 4. La longitud promedio de la red L: Es el promedio de las longitudes mínimas L_{ij} entre todas las posibles parejas de nodos (v_i, v_j) de la red.
- 5. La distribución de tama nos de islas P(s): Es la probabilidad de que una isla esté compuesta por s nodos.



Figura 3. (a) Los nodos v_1 , v_2 y v_3 están conectados al nodo v_4 . Sin embargo, los nodos v_1 y v_2 no están conentados entre sí, mientras que los nodos v_2 y v_3 sí lo están. Esto significa que no todos los "amigos" de v_4 son amigos entre sí, lo cual disminuye el coeficiente de agregación. (b) Aún cuando existen varios caminos para llegar del nodo v_i al nodo v_j , el camino de longitud mínima consiste de tres pasos, indicados con líneas gruesas en la figura.

6. El tama no de la isla más grande, al que denotaremos por S_{∞} .

En una red, los nodos además de estar conectados también interactuan, y las interacciones pueden dar lugar a fenómenos dinámicos muy interesantes. Por lo tanto, además de estudiar las propiedades estructurales de una red también es importante estudiar sus propiedades dinámicas una vez que sabemos de qué manera interactuan los nodos. Por ejemplo, las enfermedades en una sociedad no son estáticas, sino que se propagan por toda la población dando lugar a epidemias. Las neuronas en el cerebro están conectadas físicamente unas con otras por medio de las uniones entre dendritas y axones. A través de dichas uniones las neuronas se transmiten se nales eléctricas que se propagan por todo el cerebro y que dan lugar a una serie de fenómenos dinámicos interesantísimos, entre los cuales destacan el reconocimiento de imágenes y sonido, la motricidad de los músculos, el lenguage, el pensamiento y finalmente la consciencia. Otros ejemplos son la propagación de virus informáticos en la red internet, o la comunicación entre los peces que da lugar a grupos enormes de peces moviendose todos en la misma dirección. En fin, existen tantos fenómenos dinámicos en redes complejas como interacciones físicas, químicas, informáticas, o sociales se puedan imaginar, y cada día aparcen más y más artículos en la literatura científica donde se estudian nuevos procesos dinámicos sobre redes complejas.

En estas notas introductorias nos enfocaremos más al

estudio de las propiedades estructurales de las redes complejas. Sin embargo, en el último capítulo veremos un poco las propiedades dinámicas de redes neuronales.

IV. ESTRUCTURA DE LAS REDES COMPLEJAS

A. Distribución de vecinos

Tal vez la propiedad más importante que caracteriza la estructura de una red compleja es la distribución de vecinos P(k), que nos dá la probabilidad de que un nodo escogido al azar tenga k conexiones (o vecinos). En los trabajos recientes que se han llevado a cabo para caracterizar a las redes complejas se ha encontrado que existen tres tipos de distribuciones P(k) importantes, las que determinan tres estructuras o topologías[13] diferentes:

Topología de Poisson
$$P(k) = e^{-z} \frac{z^k}{k!},$$
 (1)

Topología Exponencial $P(k) = Ce^{-\alpha k}$, (2)

Topología Libre de Escala $P(k) = Ck^{-\gamma}$. (3)

Las redes con topología de Poisson son importantes principalmente por razones históricas, ya que dichas redes fueron las primeras que se analizaron matemáticamente. Este análisis lo llevaron a cabo los matemáticos húngaros Paul Erdös (1913-1996) y Alfréd Rényi (1921-1970) en la década de los 50s. Ellos también reportaron la primera transición de fase topológica observada en redes con topología de Poisson. Por lo tanto, a estas redes también se les conoce como redes tipo Erdös-Rényi. Sin embargo, a pesar de su importancia histórica, las redes con topología de Poisson están lejos de ser una representación realista de las redes reales observadas en la naturaleza. No fue sino hasta 1998 que se comenzó el estudio sistemático de las propiedades topológicas de las redes complejas reales. En este estudio participaron principalmente y de forma independiente Albert Lásló-Barabási, Ricard Solé y Mark J. Newman. Ellos encontraron que la topología exponencial aparece algunas veces en las redes reales. Pero el resultado más sorprendente de sus estudios fue la ubicuidad de las redes con topología libre de escala, la cual aparece prácticamente en todos lados, desde las peque nas redes metabólicas dentro de la célula, hasta las grandes redes informáticas como la red Internet. En el Cuadro II se listan algunas de las redes con topologías libres de escala que se han encontrado en los últimos 10 años.

Seguramente se estarán preguntando qué es lo sorprendente respecto a la topología libre de escala. Bueno, por un lado sorprende que esta topología se encuentre en redes tan diferentes y de tan gran variedad como las lista-

Red	Núm. de nodos	Núm. de conexiones	γ_i
dominio www.nd.edu	325,729	1,469,680	2.1
Páginas de WWW	$2,711 \times 10^{9}$	$2,130 \times 10^{9}$	2.1
encontradas por Altavista	,		
Dominos en	$2,60 \times 10^{5}$		1.9
la WWW			
Nivel de inter-dominio	4,389	8,256	2.2
de la Internet			
Sistemas autónomos	6,374	13,641	2.2
en la Internet			
Nivel de ruteador	150,000	200,000	2.3
en la Internet			
Citas en la base	783,339	6,716,198	3.0
de datos ISI			
Citatas de la revista	24,296	351,872	2.3
Phys. Rev. D			
Red de colaboraciones	212,250	61,085,555	2.3
de actores de Hollywood			
Red de colaboradores en	1,388,989	$1,028 \times 10^{7}$	2.5
las revistas Medline			
Colaboradores en las	70,975	$1,32 \times 10^{5}$	2.1
revistas de matemáticas			
Colaboradores an las	209,293	$1,21 \times 10^{6}$	2.4
revistas de neurociencias			
Red de interacciones	778	$\sim 1500 - 3000$	2.2
metabólicas en $(E. \ coli)$			
Red de interacciones	1,870	2,240	2.5
protéicas en levadura			
Co-ocurrencia de palabras	470,000	17,000,000	2.7
Red de palabras sinónimas	22,311		2.8
Circuitos digitales	2×10^4	4×10^4	3.0
Llamadas telefónicas	47×10^{6}	8×10^{7}	2.1
Red de interacciones	2,810		3.4
sexuales en humanos			
Redes alimenticias	154	405	1.0
(interacciones tróficas)			

Cuadro II. Algunas de las redes libres de escala que se han encontrado en la naturaleza. Sólo se muestra el exponente de entrada para las redes dirigidas. Los cuadros con líneas "——ïndican que yo no tenía el dato correspondiente al momento de escribir estas notas.

das en el Cuadro II. El hecho de que la topología libre de escala aparezca por todos lados sugiere que podría existir un mecanismo simple que genera este tipo de redes a diferentes niveles de organización, desde las peque nas redes intracelulares hasta las grandes redes sociales o informáticas. ¿Podría ser esto posible? ¿Será cierto que la formación de redes tan diferentes como la red de interacciones protéicas de *S. cerevisiae* (levadura), la red Internet y la red de colaboraciones científicas en las revistas de neurociencias, esté governada por la misma ley fundamental? No lo sabemos aún.

Por otro lado, la topología libre de escala es sorprendente porque no se esperaba que existiera. En la naturaleza existen muchos procesos aleatorios que generan distribuciones de Poisson o distribuciones exponenciales, pero existen muy pocos procesos conocidos que generan distribuciones libres de escala como la dada en la Eq. (3)[14]. De hecho, el trabajo de Erdös y Rényi demostró que las redes que se construyen a nadiendo nuevos nodos y conexiones al azar presentan topologías de Poisson o exponenciales. Y ésta es precisamente la paradoja, que la red internet, la red de colaboraciones científicas y la red de contactos sexuales, por ejemplo, son redes que se formaron aleatoriamente a nadiendo nuevos nodos y nuevas conexiones a lo largo del tiempo. Entonces, ¿cómo es posible que estas redes que se formaron al azar no presenten topologías de Poisson o topologías exponenciales como lo habían predicho Erdös y Rényi?

Las redes con topología de Poisson son muy diferentes estructuralmente a las redes con topología libre de escala. La Fig. 4 muestra una red de Poisson (arriba) y una red libre de escala (abajo). Como puede observarse, la red de Poisson se ve más aleatoria y más homogénea que la red libre de escala[15]. En las redes de Poisson todos los nodos tienen más o menos el mismo número de conexiones. Algunos nodos estarán más conectados que



Figura 4. La figura superior muestra una red aleatoria con topología de Poisson. La red mostrada en el panel inferior es la red Internet al nivel de ruteadores, la cual es una red con topología libre de escala.

otros, pero en promedio todos tienen la misma conectividad, es decir, las conexiones en una red de Poisson están distribuidas homogéneamente entre sus nodos. Por el contrario, la característica más importante de las redes libres de escala es su alta heterogeneidad, ya que existen nodos con muy pocas conexiones, nodos medianamente conectados y nodos extremadamente conectados. Los nodos altamente conectados se denominan los *núcleos* o *centros* de la red[16]. Con una red libre de escala uno no puede decir que todos los nodos tienen "más o menos" la misma conectividad. Por el contrario, hay nodos con una sóla conexión y también hay nodos con miles de conexiones.

Todavía no sabemos cuáles son los procesos que conducen a la formación de redes libres de escala. Sin embargo, en los últimos 8 a nos se han llevado a cabo avances sustanciales en la comprensión de estas redes. En las secciones que siguen veremos algunos de los formalismos matemáticos de crecimiento de redes que se han desarrollado para generar las tres topologías mencionadas con anterioridad (Poisson, exponencial y libre de escala).

V. REDES DE TIPO ERDÖS-RÉNYI

Imaginemos un conjunto de N botones de pantalón distribuídos aleatoriamente sobre una mesa e inicialmente desconectados. Al tiempo t = 0 escogemos aleatoriamente una pareja de botones y los hilvanamos con un hilo. Después de haber enlazado a esta pareja, la dejamos sobre la mesa y escogemos aleatoriamente otra pareja para hilvanar. Podemos escoger botones que están conectados con otros botones, pero si la pareja que escogemos ya está conectada entre sí, la descartamos y escogemos otra pareja. Lo que no se vale es hilvanar más de una vez a la misma pareja de botones. Repetimos este proceso sucesivamente M veces, escogiendo aleatoriamente una pareja de botones cada vez. Al final del proceso habremos establecido M enlaces entre M parejas diferentes de botones, generando así una red de botones. Intuitivamente es claro que si M (el número total de enlaces) es peque no comparado con N (el número total de botones), entonces la red resultante estará desmembrada en varias islas peque nas. Dentro de cada isla los botones estarán hilvanados entre sí, pero estarán desconectados de las otras islas. Sin embargo, si M es muy grande comparado con N, terminaremos con casi todos los botones hilvanados unos con otros. Probablemente haya islas muy peque nas desconectadas de la red principal, pero seguramente la gran mayoría de botones formarán parte de una isla principal: la isla gigante.

Después de haber hilvanado M parejas en un conjunto total de N botones, ¿cuál es la distribución de conexiones P(k) en la red resultante? Como veremos en un momento, la red que resulta de este proceso tiene una distribución P(k) de Poisson. Pero antes de dar la prueba de este resultado, vale la pena mencionar que durante muchos a nos se pensó que este mecanismo de formación de redes en el cual parejas de nodos se enlazan aleatoriamente, era adecuado para describir el origen de ciertas redes sociales como las redes de amistades o las redes de contactos sexuales. Después de todo, las amistades o los contactos sexuales se dan por el encuentro casual y aleatorio de las personas que viven en una sociedad. Por lo tanto, era natural pensar que el mecanismo de "hilvanar parejas de botones escogidas al azar" reproducía lo que realmente ocurre en las redes sociales. Sí, era "natural" pensarlo, pero era incorrecto.

Calculemos ahora la probabilidad P(k) para nuestra red de botones hilvanados. Para comenzar notemos que el número *total* N_p de parejas que se pueden formar en un conjunto de N botones es

$$N_p = \frac{1}{2}N(N-1)$$

Como enlazamos M parejas de botones, la probabilidad p_e de que una pareja arbitraria seleccionada al azar esté enlazada es

$$p_e = \frac{M}{N_p} = \frac{2M}{N(N-1)}$$
 (4)

Ahora enfoquemos nuestra atención sobre un nodo particular v_j de la red, escogido al azar. El número total de parejas que podrían contener a v_j es N-1, ya que v_j se podría haber hilvanado con los N-1 nodos restantes de la red. Sin embargo, en los M enlaces que se llevaron a cabo, no necesariamente escogimos al nodo v_j todas las veces posibles que se podría haber escogido. Supongamos entonces que de las M parejas que se escogieron, el nodo v_j estaba solamente en k de ellas. La probabilidad de que v_j esté contenido en k parejas de las N-1 posibles es

$$P(k) = \binom{N-1}{k} (p_e)^k (1-p_e)^{N-1}$$
(5)

Esta es una distribución binomial para N y Mfinitas. Pero si consideramos ahora que la red es muy grande y tomamos el límite $N\to\infty$ y $M\to\infty$ de tal forma que la cantidad

$$z = \frac{2M}{N}$$

permanezca finita, entonces la distribución (5) se transforma en

$$P(k) = e^{-z} \frac{z^k}{k!} \tag{6}$$

lo cual es la distribución de Poisson con promedio z. En el apéndice A muestro los pasos algebráicos que conducen de la ecuación (5) a la ecuación (6).

VI. CRECIMIENTO DE REDES

En la sección anterior suponíamos que teníamos una población fija de N nodos y un número fijo M conexiones a nadidas aleatoriamente. Sin embargo, en la realidad esto no ocurre, las redes no están fijas. Por el contrario, las redes complejas evolucionan y crecen en el tiempo a través de la adición simultánea tanto de conexiones como de nodos. Pensemos por ejemplo en la red internet. En octubre de 1969 la "redïnternet consistía de sólo dos computadoras, una en la Universidad de California en Los Angeles (UCLA) y la otra en el Instituto de Investigaciones de Stanford (SRI). El primer mensaje que se transmitieron estas computadoras fue "LOGWIN" [17]. A lo largo de los a nos más y más computadoras se sumaron a la red internet, y para el a no 2000, las primeras dos computadoras de la UCLA y el SRI que se dijeron "LOGWIN", se habían convertido ya en sistemas que conectaban a 170 países y a más de 300 millones de personas.

Así como la red internet nació siendo peque na y después creció con el paso del tiempo, también las redes metabólicas y las redes genéticas dentro de la célula, y muchas otras redes en la naturaleza, han crecido y evolucionado a lo largo del tiempo. Por lo tanto, es importante que nuestros modelos de formación de redes incorporen el hecho de que nuevos nodos y nuevas conexiones se pueden a nadir a la red. También debe tomarse en cuenta el que nodos y conexiones ya existentes pueden eliminarse.

En el modelo más simple de crecimiento de redes a nadimos un nuevo nodo en cada paso de tiempo. Este nuevo nodo puede conectarse con alguno de los nodos ya



Figura 5. Crecimiento de redes. Al tiempo t = 0 hay un sólo nodo v_0 . En cada paso de tiempo subsecuente a nadimos un nuevo nodo que se conectará a alguno de los nodos ya existentes con una probabilidad $\Pi(k,t)$ que depende de la conectividad k al tiempo t del nodo con el que se pretende establecer la conexión.

existentes. Cada uno de los nodos ya existentes pueden ser seleccionados para la conexión con una probabilidad $\Pi(k_i, t)$, siendo k_i la conectividad al tiempo t del *i*-ésimo nodo ya existente.

El proceso comienza con un único nodo inicial v_0 al tiempo t = 0 (ver la Fig. 5). Al tiempo t = 1 a nadimos un nuevo nodo v_1 , que se conectará al único nodo ya existente v_0 con probabilidad 1. Después, al tiempo t = 2a nadimos el nodo v_2 que se conectará con cualquiera de los nodos ya existentes v_1 y v_2 con la misma probabilidad 1/2. En este momento los nodos v_0 , v_1 y v_2 ya no tienen todos la misma conectividad: alguno de ellos tendrá dos conexiones mientras que los otros dos nodos tendrán sólo una conexión. Al tiempo t = 3 a nadimos al nodo v_3 , que se conectará con alguno de los nodos ya existentes v_0, v_1 o v_2 con una probabilidad que es función de sus conectividades k_0 , k_1 y k_2 . Continuando con este proceso, al tiempo t+1 a nadimos al nodo v_{t+1} que se conectará a cualquiera de los nodos ya existentes v_0, v_1, \ldots, v_t , el cual será seleccionado con una probabilidad $\Pi(k_i, t)$, siendo k_i la conectividad al tiempo t del nodo seleccionado.

Denotemos por P(k, t) a la probabilidad de que al tiempo t un nodo arbitrario de la red tenga k conexiones. Es claro que esta probabilidad depende del tiempo. Sin embargo, si continuamos a nadiendo nodos por un tiempo suficientemente largo esperamos que la función P(k, t)alcance un estado estacionario independiente del tiempo. Esto no significa que la *red* alcance un estado estacionario. La red sigue creciendo mientras continuemos a nadiendo nodos. Es únicamente la distribución de conectividades P(k, t) la que llega a un estado estacionario en el cual P(k, t + 1) = P(k, t) = P(k)[18].

Existen varios métodos para calcular la distribución de conectividades P(k, t). En estas notas veremos el método de la *ecuación maestra*, pero existen por lo menos otros dos métodos distintos: el *método continuo*, inventado por Barabási, y el *método cinético*, introducido por Krapivsky, Redner, y Leyvraz [2]. Los tres métodos dan resultados equivalentes.

Para aquellos que no conocen el formalismo de la ecuación maestra, en el apéndice B doy una breve introducción al planteamiento de dicha ecuación, la cual es muy fácil de escribir pero, la mayoría de las veces, muy difícil de resolver.

Denotemos por P(n, k, t) la probabilidad de que el nodo v_n tenga k conexiones al tiempo t. Notemos que esta probabilidad está asociada al nodo específico v_n . Sin embargo, podemos obtener la probabilidad P(k, t) de que un nodo *arbitrario* tenga k conexiones al tiempo t a trevés de la siguiente ecuación:

$$P(k,t) = \frac{1}{N(t)} \sum_{n=0}^{t} P(n,k,t),$$
(7)

donde N(t) es el número total de nodos de la red al tiempo t.

La ecuación que determina la evolución temporal de P(n, k, t) se obtiene notando que en cada paso de tiempo hay dos contribuciones a dicha probabilidad:

- 1. Al tiempo t el nodo v_n tenía k-1 conexiones y fue seleccionado (con probabilidad $\Pi(k-1,t)$) para conectarse con el nuevo nodo a nadido a la red. Por lo tanto, al tiempo t+1 el nodo v_n tendrá k conexiones.
- 2. El nodo v_n ya tenía k conexiones al tiempo t y no fue seleccionado para conectarse con el nuevo nodo a nadido (lo cual ocurre con probabilidad $1 \Pi(k,t)$). Por lo tanto, al tiempo t + 1 el nodo v_n seguirá teniendo k conexiones.

Tomando en cuenta estas dos contribuciones, la ecuación maestra que determina la evolución temporal de P(n,k,t) es

$$P(n,k,t+1) = \underbrace{P(n,k-1,t)}_{\substack{k-1 \text{ conexiones}\\al \text{ tiempo } t}} \underbrace{\prod_{\substack{v_n \text{ fue}\\seleccionado}}^{v_n \text{ fue}}$$

$$+ \underbrace{P(n,k,t)}_{\substack{k \text{ conexiones}\\al \text{ tiempo } t}} \underbrace{\left[1 - \Pi(k,t)\right]}_{seleccionado}.$$
 (8)

Como el nodo v_n "nació" al tiempo t = n con una sóla conexión, la condición inicial para resolver la ecuación anterior es

$$P(n,k,t)|_{t=n} = \delta_{k,1}.$$
 (9)

Sumando sobre n en la Eq. (8) desde n = 0 hasta n = ty tomando en consideración la Eq. (7), obtenemos

$$\sum_{n=0}^{t} P(n,k,t+1) = N(t) \Big\{ P(k-1,t)\Pi(k-1,t) + P(k,t) \left[1 - \Pi(k,t)\right] \Big\}$$
(10)

Ahora bien, como en cada paso de tiempo a nadimos un nuevo nodo (comenzando con el nodo v_0 al tiempo t = 0), entonces N(t) = t + 1. Por lo tanto, la ecuación anterior puede escribirse como

$$\sum_{n=0}^{t} P(n,k,t+1) = (t+1) \Big\{ P(k-1,t) \Pi(k-1,t) + P(k,t) \left[1 - \Pi(k,t) \right] \Big\}$$
(11)

En el lado izquierdo de la ecuación anterior podemos completar la suma hasta t + 1 de la siguiente manera:

$$\sum_{n=0}^{t} P(n,k,t+1) = \sum_{n=0}^{t+1} P(n,k,t+1) - P(t+1,k,t+1)$$
$$= \sum_{n=0}^{t+1} P(n,k,t+1) - \delta_{k,1}$$
$$= N(t+1)P(k,t+1) - \delta_{k,1}$$
$$= (t+2)P(k,t+1) - \delta_{k,1}$$

donde hemos utilizado el hecho de que

$$P(t+1,k,t+1) = \delta_{k,1}$$

tal y como lo establece la Eq. (9). También utilizamos la Eq. (7) (evaluada en t + 1) y el hecho de que N(t + 1) = t + 2. Sustituyendo estos resultados en la Eq. (11) obtenemos

$$(t+2)P(k,t+1) - \delta_{k,1} = (t+1) \Big\{ P(k-1,t)\Pi(k-1,t) + P(k,t) \left[1 - \Pi(k,t)\right] \Big\}$$
(12)

Notemos que para poder resolver la Eq. (12) necesitamos conocer explícitamente la función $\Pi(k,t)$, que es la probabilidad de que un nodo ya existente con conectividad k sea seleccionado para conectarse con el nuevo nodo que se a nade al tiempo t. Como veremos en las siguientes secciones, diferentes formas de la probabilidad $\Pi(k,t)$ conducen a topologías diferentes.

VII. TOPOLOGÍA EXPONENCIAL

En el caso en que cualquiera de los nodos ya existente pueda escogerse con la misma probabilidad para conectarse con el nuevo nodo a nadido, la probabilidad $\Pi(k, t)$ es independiente de k y queda dada por

$$\Pi(k,t) = \frac{1}{N(t)} = \frac{1}{t+1}$$

donde N(t) = t + 1 es el número total de nodos al tiempo t. Sustituyendo esta forma de $\Pi(k)$ en la Eq. (12) obtenemos

$$(t+2)P(k,t+1) - \delta_{k,q} = P(k-1,t) + tP(k,t)$$

Después de un tiempo muy largo, la función P(k, t) alcanza un estado estacionario en el que P(k, t+1) = P(k, t) = P(k). En dicho estado estacionario la última ecuación se transforma en

$$P(k) = \frac{1}{2} \left(P(k-1) + \delta_{k,1} \right).$$
(13)

Como todos los nodos de la red tienen por lo menos una conexión, es claro que P(0) = 0. Con esta condición inicial, la ecuación de recurrencia anterior tiene la solución

$$P(k) = 2^{-k} \tag{14}$$

que no es más que la distribución exponencial. Es importante se nalar que esta distribución aparece en el contexto del crecimiento de redes como el resultado de lo que se llama *enlace igualitario*, es decir, en una situación en la cual cada nodo nuevo que se a nade a la red se puede enlazar a cualquiera de los nodos ya existentes con la misma probabilidad. En este sentido el nuevo enlace que se forma es igualitario, porque no discrimina entre los nodos ya existentes.

VIII. TOPOLOGÍA LIBRE DE ESCALA

En la vida real las conexiones entre diferentes nodos no se dan de manera igualitaria. Por ejemplo, si tenemos una computadora nueva y queremos conectarla a internet, no vamos a contratar el servicio de internet de alguna compa nia elegida al azar, sino que buscaremos la compa nia que ofrezca el mejor servicio y al mejor precio, y probablemente será esta compa nia la que tenga más clientes. En una escuela los varones no buscan a su pareja al azar, sino que buscarán salir con la chica más bonita, o tal vez con la más inteligente, y será esta muchacha la que tenga más pretendientes. Por esta razón, Barabási inventó el concepto de enlace preferencial en el cual los nuevos nodos que se a naden a la red se conectarán preferentemente con los nodos ya existentes que tengan el mayor número de conexiones. Intuitivamente podemos pensar que el enlace preferencial consiste en que uno siempre trata de estar conectado con los nodos más "populares", es decir, con los nodos de mayor conectividad.

Para incorporar este comportamiento, Barabási sugirió que la probabilidad de enlace $\Pi(k,t)$ debe tomar la forma

$$\Pi(k,t) = \left(\sum_{n=0}^{N(t)} k_n\right)^{-1} k$$
 (15)

donde k_n es la conectividad del *n*-ésimo nodo ya existente al tiempo *t*. El factor $\left(\sum_{n=0}^{N(t)} k_n\right)^{-1}$ es simplemente

para garantizar que la probabilidad $\Pi(k, t)$ esté normalizada. Al hacer que $\Pi(k, t)$ sea proporcional a k, como lo propuso Barabási, tenemos enlace preferencial, ya que de esta forma, entre más grande sea la conectividad k de un nodo, mayor será la probabilidad de conectarse con él. Como en cada paso de tiempo a nadimos un nuevo nodo con una conexión, comenzando con cero conexiones al tiempo t = 0, entonces para cualquier tiempo t > 0 se tiene

$$\sum_{n=0}^{N(t)} k_n = 2t$$

y por lo tanto

$$\Pi(k,t) = \frac{k}{2t} \tag{16}$$

Substituyendo este resultado en la Eq. (12) obtenemos

$$(t+2)P(k,t+1) - \delta_{k,q} = \frac{t+1}{2t} \Big\{ (k-1)P(k-1,t) + [2t-k]P(k,t) \Big\}$$

En el límite $t \to \infty$ el sistema alcanza el estado estacionario. Por lo tanto, la distribución de conexiones estacionaria P(k) se obtiene de la ecuación anterior tomando el límite $t \to \infty$, lo que conduce a

$$P(k) + \frac{1}{2} \left[kP(k) - (k-1)P(k-1) \right] = \delta_{k,1}$$
(17)

Como todos los nodos tienen por lo menos una conexión, para resolver la ecuación (17) utilizamos la condición inicial P(0) = 0, lo cual nos da la solución

$$P(k) = \frac{4}{k(k+1)(k+2)}$$
(18)

Aun cuando esta no es una ley de potencias "perfecta" (ver la Fig. 6), para valores grandes de k la distribución P(k) se comporta como $P(k) \sim k^{-3}$. Por un lado, este es un resultado muy interesante: el enlace preferencial genera distribuciones de conectividades con colas libres de escala. Son estas colas largas las responsables de que existan elementos altamente conectados. Por otro lado, este resultado también es desalentador ya que el proceso de enlace preferencial siempre da el exponente $\gamma = 3$, el cual no se encuentra frecuentemente en la naturaleza (ver los exponentes en el Cuadro II). En otras palabras, aun cuando el método de enlace preferencial nos da una ley de potencias, es incapaz de reproducir la gran variedad de exponentes encontrados en la naturaleza. Por esta razón, se han propuesto diferentes formas de la función $\Pi(k, t)$ que corresponden a diferentes tipos de enlace preferencial. Por ejemplo, Krapivsky, Redner, y Leyvraz propusieron una función de enlace $\Pi(k, t)$ no lineal (ver la referencia [6]), de la siguiente forma

$$\Pi(k,t) = \left(\sum_{n=0}^{N(t)} k_n^{\alpha}\right)^{-1} k^{\alpha}$$
(19)



Figura 6. Gráfica log-log de la distribución de conectividades P(k) dada en la ecuación (18). La línea punteada roja es la gráfica de la ley de potencias "perfecta" $f(k) = 4k^{-3}$. Nota que $P(k) \approx 4k^{-3}$ para valores grandes de k.

donde α es un exponente arbitrario. Desafortunadamente, sólo para $\alpha = 1$ se tiene que la distribución P(k) tiene una cola libre de escala. El trabajo de Krapivsky, Redner, y Leyvraz mostró que la naturaleza libre de escala de la red queda destruida cuando el enlace preferencial obedece una regla no lineal como la dada en la Eq. (19) cuando $\alpha \neq 1$. Es importante mencionar que todavía no tenemos modelos de crecimiento de redes que generen todos los exponentes listados en el Cuadro II.

IX. ¿NODOS QUE SE HACEN VIEJOS?

Otro problema con el modelo de enlace preferencial propuesto por Barabási es que predice que los nodos más viejos, es decir lo que se a nadieron primero a la red, son los que eventualmente adquirirán el mayor número de conexiones. En otras palabras, en una red libre de escala podríamos identificar facilmente a los nodos más viejos: son aquellos altamente conectados. Y entre más viejo sea un nodo, mayor será su conectividad.

Para ver que esto efectivamente es un resultado del modelo de enlace preferencial, consideremos un nodo particular v_n de la red. Su conectividad k_n va a cambiar a una tasa que es proporcional a la probabilidad $\Pi(k_n, t)$ de que este nodo adquiera más conexiones. Es decir,

$$\frac{dk_n}{dt} = \Pi(k_n, t)$$

Para el caso particular en el que $\Pi(k, t)$ está dada como en la Eq. (16), tenemos

$$\frac{dk_n}{dt} = \frac{k_n}{2t}$$

Como el nodo v_n nació al tiempo t = n con una conexión, la condición inicial para la ecuación anterior es $k_n(n) = 1$,



Figura 7. (A) Gráfica log-log de la distribución de conectividades P(k) para una red WWW consistente en 260,000 páginas. (B) Conectividad k de los nodos de la red como función de su edad. Las figuras fueron tomadas (sin permiso) de la referencia [7].

lo cual conduce a la solución

$$k_n(t) = \left(\frac{t}{n}\right)^{1/2} \tag{20}$$

De la ecuación anterior, es claro que a cualquier tiempo t se cumple la siguiente desigualdad

$$k_1(t) > k_2(t) > k_3(t) > \dots > k_n(t)$$
 (21)

lo cual significa que los nodos que nacieron primero tendrán, en promedio, conectividades mayores que los nodos que llegaron después.

Sin embargo, este comportamiento predicho por el modelo del enlace preferencial, en el que los nodos más viejos son los más conectados, no siempre se observa en la naturaleza. Por ejemplo, Lada A. Adamic y Bernardo Huberman estudiaron una red que consistía en 260,000 páginas www [7], considerando que dos páginas estaban conectadas si una contenía un hipervínculo a la otra. Adamic y Huberman encontraron que, aún cuando esta red tiene topología libre de escala con un exponente $\gamma \sim 2$ (ver la Fig. 7(a)), no existe correlación entre la conectividad de los nodos y su edad. La Fig. 7(b) muestra la gráfica de la conectividad de los nodos como función de su edad. Como puede observarse, estas dos cantidades no parecen estar relacionadas.

Probablemente el contra ejemplo más contundente que contradice los resultados de las Eqs. (20) y (21) es Google, el robot buscador de la red WWW. Recordemos que Altavista y Yahoo ya existían y estaban bien establecidos en 1998, a no en que apareció Google. Rápidamente Google tomó la delantera, convirtiéndose en el robot buscador más popular en el mundo informático. ¿A qué se debió este éxito repentino de Google? Probablemente a que estaba mejor dise nado, era más rápido y más eficiente que sus competidores Altavista y Yahoo. En otras palabras, Google nació con características que lo hacían un robot "mejor adaptado" que sus competidores.

Esta observación hizo que Barabási propusiera el concepto de adaptabilidad en el contexto de las redes complejas. Así, cada uno de los nodos v_n , además de tener una conectividad k_n , también tenia un parámetro de adaptabilidad asociado w_n . Este parámetro es una medida de qué tan bien estaba adaptado el nodo v_n a su entorno: entre más grande es el valor de w_n , mayor es el grado de adaptabilidad de v_n . Específicamente, lo que Barabási hizo fue proponer una función de enlace preferencial $\Pi(k, w, t)$ que, además de ser proporcional a la conectividad k de los nodos ya existentes, también es proporcional a su adaptabilidad w (ver la referencia [8]). De esta forma, la probabilidad de que un nuevo nodo a nadido a la red se conecte con el nodo v_n ya existente, cuya conectividad es k_n y cuya adaptabilidad es w_n , queda dada por

$\Pi(k_n, w_n, t) = C w_n k_n$

donde C es una constante de normalización. Las adaptabilidades w_n asociadas a los nodos de la red son variables aleatorias que se escogen de una distribución A(w). Cada nodo nace con su propia adaptabilidad, la cual no cambia en el tiempo.

Un resultado sorprendente del modelo de enlace preferencial con adaptabilidad es que puede ser mapeado exactamente a un gas de Bose-Einstein. En este mapeo cada nodo de la red representa un nivel energético del gas, mientras que las conectividades de los nodos representan los números de ocupación de los respectivos niveles de energía. El mapeo entre el modelo de red y el gas de Bose-Einstein es más que una simple matáfora. Es un mapeo matemático preciso y, bajo ciertas condiciones, la red en crecimiento puede presentar "condensación de Bose-Einstein", lo cual significa que repentinamente un nodo puede adquirir la mayoría de las conexiones de la red, independientemente de su edad (como sucedió con Google). Los que estén interesados podrán encontrar los detalles de este trabajo en la referencia [8]. Hasta donde yo sé, no existe acuerdo cuantitativo entre los resultados del modelo de enlace preferencial con adaptabilidad, y los datos experimentales que se tienen de redes reales. Sin embargo, la condensación de Bose-Einstein predicha en este modelo podría ser el primer paso para entender mejor el "fenómeno Google".

X. LA ISLA GIGANTE

Regresemos al modelo de los botones descrito en la sección V. Inicialmente todos los botones están sobre la mesa desconectados, pero conforme vamos hilvanando parejas de botones, se comienzan a formar islas de botones conectados. Al princípio las islas son peque nas, pero al ir a nadiendo más y más conexiones entre los botones, las islas crecen y se conectan entre sí. Eventualmente se formará una isla gigante, es decir, una isla que es mucho más grande que todas las demás.

El tama no de la isla gigante es importante en el estudio de la propagación de epidemias en una sociedad, en donde en lugar de tener botones hilvanados, tenemos personas que se contagian unas a otras. Las islas consisten en los grupos de personas que están infectadas, y la enfermedad se convierte en una epidemia cuando se forma una isla gigante que abarca a la mayoría de la sociedad.

La pregunta importante es: dado un conjunto de elementos, ¿cuántos enlaces (contagios) tienen que establecerse para que se forme la isla gigante? Esta pregunta fue contestada por Erdös y Rényi, quienes fueron los primeros en mostrar la existencia de una transición de fase en la teoría de redes. Dicha transición de fase consiste precisamente en la formación de la isla gigante. Aunque su trabajo original lo llevaron a cabo para redes con topología de Poisson, es fácil generalizar el análisis que hicieron para extenderlo a redes con topologías arbitrarias.

Para calcular el tama no de la isla gigante utilizaremos un razonamiento de consistencia. Sea q la probabilidad de que un nodo v_n escogido aleatoriamente *no pertenezca* a la isla gigante. Supongamos que v_n tiene k vecinos. Claramente, v_n no pertenece a la isla gigante si y sólo si ninguno de sus k vecinos tampoco pertenece a la isla gigante. Por lo tanto, la probabilidad q debe satisfacer la ecuación de consistencia

$$q = \sum_{k} P(k)q^k \tag{22}$$

El lado izquierdo de esta ecuación es simplemente la probabilidad q de que v_n no pertenezca a la isla gigante. El lado derecho es la probabilidad P(k) de que v_n tenga kvecinos, multiplicada por la probabilidad q^k de que ninguno de estos k vecinos pertenezca a la isla gigante. La suma sobre k toma en cuenta todos los posibles vecinos que v_n pudiera llegar a tener.

En el caso en que P(k) es una distribución de Poisson, la Eq. (22) se transforma en

$$q = \sum_{k=0}^{\infty} e^{-z} \frac{(zq)^k}{k!} = e^{z(q-1)}$$

Denotando como p = 1 - q la probabilidad de que un nodo



Figura 8. Probabilidad p de pertenecer a la isla gigante como función de la conectividad promedio z de la red, pare el caso de la topología de Poisson.

arbitrario si pertenezca a la isla gigante, obtenemos

$$p = 1 - e^{-zp} \tag{23}$$

La ecuación anterior es trascendental y no puede resolverse de forma analítica. Sin embargo, es fácil resolverla de forma numérica y encontrar p para cada valor de z. La Fig. 8 muestra p como función de z. Notemos que para z < 1 la probabilidad de pertenecer a la isla gigante es cero. En otras palabras, no hay isla gigante para z < 1. Conforme aumentamos el valor de z (lo cual es equivalente a a nadir más conexiones a la red), la isla gigante comienza a formarse v justo en z = 1 aparece. Este es un resultado interesante: se necesita, en promedio, sólo una conexión por nodo para que aparezca la isla gigante. Si seguimos aumentando el valor de z, el tama no de la isla gigante aumenta más y más, como se puede observar en la Fig. 8, la cual muestra el comportamiento característico de una transición de fase de segundo orden que ocurre en z = 1.

Consideremos ahora el caso de las redes con topología libre de escala, para las que la distribución P(k) toma la forma

$$P(k) = \frac{1}{\zeta(\gamma)} k^{-\gamma}$$

donde $\zeta(\gamma) = \sum_{k=1}^\infty k^{-\gamma}$ es la función zeta de Riemman la

cual, en este caso simplemente juega el papel de la constante de normalización. Para esta distribución la ecuación de consistencia Eq. (22) se convierte en

$$q = \frac{1}{\zeta(\gamma)} \sum_{k=1}^{\infty} \frac{q^k}{k^{\gamma}}$$
$$= \frac{1}{\zeta(\gamma)} \text{Li}_{\gamma}(q)$$

donde Li_{γ} $(q) = \sum_{k=1}^{\infty} k^{-\gamma} q^k$ es la función poli-logarítmica de orden γ . En términos de la probabilidad p = 1 - q, la ecuación anterior queda como

$$p = 1 - \frac{1}{\zeta(\gamma)} \operatorname{Li}_{\gamma}(1-p)$$

Como $\operatorname{Li}_{\gamma}(0) = 0$, es claro p = 1 siempre es olución de la ecuación anterior. Por lo tanto, en una red libre de escala *la componente gigante siempre existe y la probabilidad de pertenecer a ella es p* = 1. Este interesante resultado es una consecuencia de la existencia de los elementos altamente conectados en redes libres de escala. Dichos elementos mantienen a toda la red conectada, impidiendo que se fracture en peque nas islas.

El resultado anterior puede tener implicaciones importantes en la propagación de epidemias en una sociedad, en la que podemos considerar que la isla gigante es el conjunto de personas que se han contagiado de alguna enfermedad. Los resultados presentados en esta sección muestran que en redes con topología de Poisson siempre podemos vacunar a un número suficiente de personas para garantizar que la conectividad promedio z de la red de individuos infectados se mantenga inferior a 1. En tal caso podemos detener la epidemia, ya que no habrá isla gigante, sólo peque nas islitas de personas infectadas.

Sin embargo, las redes sociales no tienen topología de Poisson, sino que exhiben topologías libres de escala. Para tales topologías siempre existe una isla gigante, y la probabilidad de pertenecer a ella es p = 1. Por lo tanto, en redes libres de escala las enfermedades tarde o temprano se propagan a todas las personas que no hayan sido vacunadas. Es decir, la única forma de detener la propagación de una epidemia en redes libres de escala es vacunando a *todas* las personas de la sociedad. Los efectos aterradores del resultado anterior pueden observarse claramente en áfrica, donde comunidades enteras han desaparecido a causa del SIDA.

Más información sobre la propagación de epidemias en redes sociales puede encontrarse en las referencias [9, 10].

XI. PROCESOS DINÁMICOS EN REDES

Hasta ahora hemos visto las propiedades estructurales o topológicas de las redes complejas. Estas propiedades nos dicen cómo están conectados los elementos de la red y su estudio nos muestra de qué forma procesos aleatorios y aparentemente independientes pueden dar lugar a estructuras complejas e inesperadas (como la topología libre de escala). No obstante, sabemos que una vez que los elementos de una red están conectados, éstos pueden comenzar a interactuar generando procesos dinámicos que se propagan a través de toda la red. Consideremos, por ejemplo, las neuronas en el cerebro, las cuales están conectadas unas con otras por medio de axones y dendritas. A través de estas conexiones las neuronas interaccionan enviándose impulsos eléctricos que se propagan por todo el cerebro y que generan patrones de actividad eléctrica en las diferentes capas cerebrales responsables de la visión, el olfato, el tacto, etc. Son precisamente estos patrones de actividad eléctrica neuronal, asociados con nuestros sentidos y con nuestra consciencia, lo que llamamos "procesos dinámicos" en el cerebro.

Otro ejemplo interesante de un proceso dinámico en redes es la formación de opiniones en una sociedad. En este caso, la red social la componen los individuos de la sociedad ligados entre sí por medio de interacciones de amistad, camaradería, parentesco, etc. Los individuos intercambian opiniones con sus amigos, colegas, compañeros o familiares respecto a un determinado tema, y es gracias a este intercambio de opiniones que los individuos se van formando su propia opinión. Las opiniones se propagan a través de toda la sociedad, y eventualmente en la sociedad existirá una opinión mayoritaria respecto a dicho tema. El libre intercambio de opiniones es la base de la democracia en una sociedad moderna, en la que se supone que los individuos eligen, por mayoría, a sus gobernantes. Este es precisamente el tema que abordaremos en este capítulo: la formación de una opinión mayoritaria en una sociedad.

A. El modelo de votantes

Supongamos que en una sociedad existe un tema respecto al cual se pueden tener sólo dos opiniones. Por ejemplo, el tema del aborto, respecto al cual se puede estar a favor o en contra. O la privatización de la industria eléctrica, respecto de la cual cada uno de nosotros podemos estar a favor o en contra. (No valen opiniones indeterminadas como el típico "no sé"). Cada individuo tiene su propia opinión respecto a dicho tema. Representemos la opinión de un individuo con la variable σ , la cual puede tomar sólo dos valores: $\sigma = +1$ si el individuo está a favor, y $\sigma = -1$ si el individuo está en contra. Como en una sociedad existen N individuos, entonces tenemos un conjunto de N variables $\sigma_1, \sigma_2, \ldots, \sigma_N$, cada una representando la opinión de un individuo en la sociedad. De esta forma, $\sigma_n = +1$ si el n-ésimo indivi

duo está a favor, mientras que $\sigma_n = -1$ si el *n*-ésimo individuo está en contra.

Lo interesante es que cada individuo intercambia opiniones con sus amigos y, en base a dicho intercambio, se forma su propia opinión. Supongamos, por tanto, que la opinión de cada individuo está influenciada por sus amigos (o conocidos, o familiares, etc.). Sea k_n el número de personas que tienen influencia sobre la opinión del *n*-ésimo individuo, y sean $\sigma_{n_1}, \sigma_{n_2}, \ldots, \sigma_{n_{k_n}}$ las k_n personas que tienen influencia sobre el individuo σ_n . Estríctamente hablando, las variables σ_n representan las *opiniones* de las personas, no a las personas mismas. Sin embargo, utilizaremos un lenguaje simple y diremos indistintamente que σ_n es la opinión del *n*-ésimo individuo, o bien σ_n es el *n*-ésimo individuo.

Decíamos entonces que cada individuo σ_n recibe opiniones de otros k_n individuos en la sociedad, a los cuales denotamos como $\sigma_{n_1}, \sigma_{n_2}, \ldots, \sigma_{n_{k_n}}$. Llamaremos a este conjunto los inputs de σ_n . En otras palabras, los inputs de σ_n son todas las personas que tienen influencia en su opinión, ya sean amigos, conocidos, parientes, maestros, etc. Cada individuo tiene su conjunto particular de inputs, lo cual significa que dos individuos diferentes σ_n y σ_m tendrán, en general, inputs distintos (aunque algunos de sus inputs pueden ser comunes). La red social se forma estableciendo quién recibe opiniones de quién, es decir, estableciendo para cada individuo el conjunto de personas que influencian su opinión. Notemos que esta red es dirigida, ya que si σ_i es un input de σ_j , entonces no necesariamente σ_j será un input de σ_i .

Una vez que sabemos cuáles son los inputs de cada individuo en la sociedad, tenemos que establecer la regla dinámica a través de la cual los individuos influyen en las opiniones de los demás. Dicha regla está basada en la siguiente observación sencilla: en general, un individuo tiende a ser de la misma opinión que la mayoría de sus amigos (inputs). Si lo pensamos un poco nos daremos cuenta de que esta regla es bastante cierta. En general, tendemos a juntarnos con personas que tienen (más o menos) la misma forma de pensar que nosotros. El objetivo de este capítulo es ver como esta sencilla regla de interacción, en la que un individuo "tiende" a ser de la misma opinión que la mayoría de sus inputs, puede generar estados de orden colectivo en toda la sociedad.

Establezcamos matemáticamente la regla dinámica. Sean $\sigma_{n_1}, \sigma_{n_2}, \ldots, \sigma_{n_{k_n}}$ las opiniones de los k_n inputs de σ_n . Estas opiniones cambian a lo largo del tiempo, y supondremos que el tiempo lo medimos en unidades discretas que pueden ser días o semanas. Por ejemplo, cada semana hacemos una encuesta para monitorear las opiniones de las personas en la sociedad. Nuestra primer regla dinámica consiste en que el valor de la variable σ_n al tiempo t+1 está determinado por el valor de sus inputs al tiempo t de acuerdo con la siguiente ecuación:

$$\sigma_n(t+1) = \operatorname{Sgn}\left[\frac{1}{k_n} \sum_{j=1}^{k_n} w_{n,n_j} \sigma_{n_j}(t)\right], \qquad (24)$$



Figura 9. Los individuos de la sociedad están representados por los círculos negros. Los colocamos en una red cuadrada de tal manera que cada individuo tiene 5 inputs, sus cuatro primeros vecinos y él mismo. Las flechas indican los inputs del individuo particular seleccionado. La misma configuración se tiene para todos los individuos.

donde la función Sgn[x] se define como

$$\operatorname{Sgn}[x] = \begin{cases} +1 & \operatorname{si} & x \ge 0\\ -1 & \operatorname{si} & x < 0 \end{cases}$$

Por otro lado, la cantidad $\frac{1}{k_n} \sum_{j=1}^{k_n} w_{n,n_j} \sigma_{n_j}(t)$ que aparece

dentro de la función Signo es el promedio ponderado de las opiniones de los inputs de σ_n al tiempo t. Los factores w_{n,n_i} son números reales que representan el peso de las opiniones de los inputs. Estos factores toman en cuenta el hecho de que no todos los amigos de un individuo tienen la misma influencia sobre su opinión. Habrá algunas personas a las que dicho individuo respetará más que a otras. Por lo tanto, en esta suma ponderada, entre más grande sea el valor de algún w_{n,n_i} , más importancia tendrá el correspondiente input σ_{n_j} para determinar el valor de σ_n . Más aún, algunos de los factores w_{n,n_i} podrían incluso ser negativos, lo que significaría que el individuo σ_{n_i} tiene una influencia *negativa* sobre la opinión del individuo σ_n . (Nunca falta el baboso que expresa puras opiniones desatinadas. Nada más de oirlo nos dan ganas de hacer lo contrario). En general los factores w_{n,n_i} se asignan de forma aleatoria tomándolos de una distribución de probabilidad P(w) que representa estadísticamente la manera en que las pesonas influyen sus opiniones unas a otras. En la siguiente sección veremos como elegiendo adecuadamente los factores w_{n,n_i} podemos hacer que la red reconozca imágenes. Por el momento vamos a suponer simplemente que estos factores están asignados de forma aleatoria.

La Eq. (24) nos dice entonces que $\sigma_n = +1$ si la mayoría (ponderada) de sus inputs tienen opinión +1, mientras que $\sigma_n = -1$ si la mayoría (ponderada) de sus inputs tienen opinión -1. A cada instante de tiempo, todos los individuos de la red actualizan su valor de acuerdo con la



Figura 10. Cada cuadrito representa un individuo, negro si el individuo tiene opinión -1 y claro si el individuo tiene opinión +1. El cuadro grande de la izquierda es la condición inicial con 50 % de opiniones positivas y 50 % de opiniones negativas repartidas aleatoriamente en la sociedad. El cuadro de la derecha es el estado al que llega el sistema después de 100 pasos de tiempo iterando la Eq. (24). Aun cuando se forman grupos de opinión, los porcentajes de opiniones positivas y negativas no cambian con el tiempo y permanecen iguales a lo que se tenía en la condición inicial.

Eq. (24). Ilustremos la dinámica generada por esta ecuación en un caso sencillo.

Supongamos que colocamos a los individuos en una red cuadrada de tal modo que cada individuo recibe inputs de sus cuatro primeros vecinos y de sí mismo (porque su opinión también cuenta), como se indica en la Fig. 9. Como todos los individuos tienen 5 inputs, entonces $k_n = 5$ para todos. Además, vamos a suponer que todos los pesos w_{n,n_i} son positivos, es decir, los inputs de un individuo tienen todos una influencia positiva sobre su opinión. En esta caso, la dinámica generada por la Eq. (24) es muy aburrida. La Fig. 10 muestra un sistema con 10000 individuos acomodados en una red cuadrada de 100×100 . Cada cuadrito es un individuo. Aquellos con opinión -1 se pintan con color negro, mientras que los individuos con opinión +1 se pintan en un color más claro. Al tiempo t=0las opiniones +1y -1 están distribuidas al azar en la sociedad de tal forma que aproximadamente la mitad de los individuos tienen opinión +1 y la otra mitad tienen opinión -1. Después de iterar la Eq. (24) cien veces (t = 100), el sistema alcanza un estado estacionario en el que se forman *arupos de opinión*, es decir, individuos con la misma opinión tienden a agregarse en cúmulos. Aunque dejemos correr el tiempo más y más, la configuración mostrada en el panel derecho de la Fig. 10 ya no cambia, se queda "congelada" en el tiempo. Es interesante notar que, aunque se formaron grupos de opinión, los porcentajes de opiniones positivas y negativas siguen siendo los mismos que al principio, 50 % y 50 % respectivamente. Es decir, la Eq. (24) no genera una opinión mayoritaria en la sociedad, simplemente preserva los porcentajes iniciales de opiniones, aunque lleva al sistema a un estado en donde se forman cúmulos de opinión.

Este ejemplo muestra que la vida sería muy sencilla si nuestras opiniones estuvieran regidas por la Eq. (24), es decir, si todo mundo hiciéramos lo que la mayoría de las personas nos dicen que hagamos. Sin embargo, somos





Figura 11. En este caso la dinámica del sistema está regida por la Eq. (25) con una probabilidad $\eta = 0.05$ de violar la regla de la mayoría. El primer cuadro al tiempo t = 0 es la condición inicial, con 50% de opiniones positivas y 50% de opiniones negativas repartidas aleatoriamente en la sociedad. Vemos que conforme pasa el tiempo se va formando una opinión mayoritaria en la sociedad, hasta que claramente al tiempo t = 10000 la gran mayoría de los individuos en la sociedad tienen opinión +1.

necios en el sentido de que, incluso cuando la mayoría de nuestros amigos tengan una opinión respecto a un tema, con una determinada probabilidad nosotros podemos tener la opinión contraria. Por lo tanto, modificaremos la Eq. (24) para introducir el *libre albedrío* de las personas. Esto lo haremos suponiendo que cada individuo en la sociedad adquiere la opinión contraria a la de la mayoría de sus inputs con probabilidad η , o bien, adquiere la opinión de la mayoría de sus inputs con probabilidad $1 - \eta$. Es decir, la regla dinámica es ahora

$$\sigma_n(t+1) = \begin{cases} \operatorname{Sgn}\left[\sum_{j=1}^{k_n} w_{n,n_j} \sigma_{n_j}(t)\right] & \operatorname{prob.} \quad 1-\eta \\ -\operatorname{Sgn}\left[\sum_{j=1}^{k_n} w_{n,n_j} \sigma_{n_j}(t)\right] & \operatorname{prob.} \quad \eta \end{cases}$$
(25)

El parámetro η es entonces la probabilidad de que cada individuo vaya en contra de la opinión mayoritaria de sus inputs.

La Fig. 11 muestra la dinámica generada por la Eq. (25) para el caso de la red cuadrada mostrada en la Fig. 9 (otravez, con todos los pesos $w_{n_j} = 1$). Escogimos para este ejemplo una probabilidad $\eta = 0.05$ muy pequeña de ir en contra de la opinión mayoritaria de los inputs. Como muestra la Fig. 11, comenzamos al tiempo t = 0 con una condición inicial con la mitad de opiniones positivas y la otra mitad de las opiniones negativas, distribuidas aleatoriamente en la sociedad. Conforme pasa el tiempo se forman los grupos de opinión, es decir, individuos con la misma opinión tienden a agregarse (como en el panel correspondiente al tiempo t = 100). Sin embargo, al transcurrir el tiempo notamos que ocurre algo verdade-

Figura 12. Ejemplo de la dinámica generada por la Eq. (25) con un valor relativamente alto de ruido: $\eta = 0.15$. En este caso nunca se genera orden en el sistema ya que el ruido es tan alto que lo destruye.

ramente importante: emerge una opinión mayoritaria en *la sociedad* de tal forma que al tiempo t = 10000 la gran mayoría de los individuos tienen la misma opinión (que en este ejemplo es +1). Vemos entonces que el *ruido* que introducen las personas que hacen lo contrario de lo que dicen la mayoría de sus inputs (los famosísimos "contreras", o "anarquistas", o "revoltosos", etc.), representados por el parámetro η , son fundamentales para la emergencia de orden en la sociedad. Notemos que la Eq. (24), la cual no genera una opinión mayoritaria, es un caso particular de la Eq. (25) con $\eta = 0$. Por lo tanto, si en nuestra sociedad no hay "revoltosos", si todo el mundo hace lo que dicen los demás, no se genera orden. Es fundamental la presencia de unos cuantos "revoltosos" para la emergencia de orden, es decir, para la formación de una opinión común en toda la sociedad. Esto es sorprendente porque generalmente tendemos a pensar que el ruido (los "revoltosos") destruye el orden, y lo consideramos como algo no deseado. Sin embargo, este ejemplo muestra que el ruido muchas veces es necesario para tener orden en un sistema. Sorprendente, ¿no?

B. Transición de fase

En el ejemplo mostrado en la Fig. 11 el valor del parámetro η que determina la cantidad de ruido presente en el sistema es muy pequeño: $\eta = 0.05$ (esto quiere decir que hay muy pocos revoltosos). En tal caso la presencia del ruido es fundamental para generar un orden global, es decir, una opinión mayoritaria en toda la sociedad. Sin embargo, mucho ruido tampoco es bueno porque entonces el orden se destruye. Esto se ilustra en la Fig. 12, la cual se generó utilizando el valor $\eta = 0.15$. Como podemos ver, el sistema nunca se ordena. Nunca emerge una



Figura 13. Parámetro de orden $\psi(t)$ como función del tiempo para para la red cuadrada de la Fig. 9 con todos los pesos $w_{n,n_j} = 1$ y un valor de ruido $\eta = 0.05$. La condición inicial corresponde a un estado en el que la mitad de las opiniones son positivas y la mitad son negativas. Por lo tanto, $\psi(0) \approx 0$. Nótese que después de un tiempo transitorio, aproximadamente en t = 5000, el parámetro de orden alcanza un valor constante que ya no cambia en el tiempo (salvo por pequeñas fluctuaciones producidas por el ruido).

opinión global común en la sociedad. (Incluso si comenzaramos con una condición inicial totalmente ordenada en la que todos los individuos tienen la misma opinión, la presencia de muchos revoltosos destruiría este orden.) Unos cuantos revoltosos ayudan a generar orden, pero muchos lo destruyen.

Para ver qué tanto ruido se necesita para destruir el orden en la sociedad vamos a definir un parámetro que mida la cantidad de orden. A este parámetro se le llama (no muy imaginativamente) *el parámetro de orden*, y se define como

$$\psi(t) = \frac{1}{N} \left| \sum_{n=1}^{N} \sigma_n(t) \right|.$$
(26)

De acuerdo con esta definición, $\psi(t)$ es el valor absoluto de la opinión promedio de la sociedad al tiempo t (la suma se hace sobre todos los individuos en la sociedad). Si aproximadamente la mitad de las opiniones son positivas y la otra mitad negativas, entonces $\psi(t) \approx 0$ y no hay orden en el sistema. Por el contrario, si casi todos los individuos tienen la misma opinión (ya sea positiva o negativa), entonces $\psi(t) \approx 1$. En tal caso decimos que el sistema está muy ordenado.

Cuando $t \to \infty$, el sistema alcanza un estado estacionario en el cual el valor de $\psi(t)$ ya no cambia (excepto por fluctuaciones), como se muestra en la Fig. 13. La gráfica mostrada en esta figura se generó simulando el sistema de la Fig. 9 en la computadora para una red cuadrada con N = 10000 individuos y utilizando un valor de ruido $\eta = 0.05$. Dado que en el límite $t \to \infty$ el valor del parámetro de orden ya no cambia, definimos el valor estacionario del parámetro de orden como[19]

$$\psi = \lim_{t \to \infty} \psi(t).$$



Figura 14. Valor estacionario del parámetro de orden ψ como función del ruido η , para el sistema de la red cuadrada mostrado en la Fig. 9. Obsérvese que al aumentar η , la cantidad de orden disminuye, hasta que para el valor crítico $\eta_c \approx 0.149$ el orden en el sistema se destruye completamente.

Este valor de ψ depende de la cantidad de ruido η presente en el sistema, es decir, de la cantidad de "revoltosos" que existan en la sociedad. La Fig. 14 muestra el valor estacionario ψ como función del ruido η . Como puede observarse, la cantidad de orden en el sistema es grande para valores pequeños de η . Sin embargo, conforme η aumenta, el orden en el sistema disminuye hasta que, para un valor crítico $\eta_c \approx 0.149$ el orden en el sistema desaparece completamente. Lo que la Fig. 14 está mostrando es una transición de fase dinámica de segundo orden (análoga a la mostrada en la Fig. 8). Esta transición de fase confirma las observaciones que hicimos en la sección anterior: Pequeñas cantidades de ruido son necesarias para generar orden en el sistema. Sin este ruido simplemente no se genera una opinión mayoritaria en la sociedad. Pero mucho ruido destruye el orden, dejándonos con una sociedad indecisa.

En este capítulo vimos los conceptos básicos referentes a los procesos dinámicos en redes. Utilizamos una red muy simple, la red cuadrada en la que cada nodo tiene 5 inputs (los 4 primeros vecinos y él mismo). El análisis se puede extender a redes con topologías más realistas, como la red libre de escala o la de mundo pequeño, y utilizando pesos w_{n,n_j} todos distintos entre sí. De hecho, el modelo de votantes puede resolverse exactamente de forma analítica, pero la solución es complicada y va más allá de los objetivos de estas notas introductorias. Sin embargo, aún cuando utilicemos diferentes topologías, o factores de peso distintos, los conceptos fundamentales presentados aquí no cambian. Los que estén interesados pueden encontrar un análisis matemático completo del modelo de votantes en la referencia [11].



Figura 15. Comenzando con una condición inicial aleatoria, como el cuadro de la izquierda que corresponde al estado de la red en t = 0, quisiéramos que la red eventualmente converja a un patrón predefinido, como la imagen del perro a la derecha que corresponde al estado de la red en el tiempo t = 10.

XII. REDES NEURONALES Y RECONOCIMIENTO DE PATRONES

El modelo de votantes discutido en la sección anterior y cuya dinámica está dada por la Eq. (24) o la Eq. (25) también se conoce como red neuronal, pues describe de forma cualitativa la dinámica de neuronas conectadas que se prenden y apagan. En este caso σ_n representa el estado de activación de la *n*-ésima neurona, teniendo $\sigma_n = 1$ si la neurona está prendida y $\sigma_n = -1$ si la neurona está apagada. Los pesos w_{nj} reflejan entonces la intensidad de la conexión sináptica entre las neuronas σ_n y σ_i . Este modelo fue propuesto por primera vez en 1943 por Warren McCulloch y Walter Pitts para describir la dinámica neuronal, pero posteriormente fue criticado por los neurocientíficos diciendo que el modelo de McCulloch-Pitts no contiene el "realismo" suficiente para describir a las neuronas biológicas. Aún así, el modelo se ha utilizado mucho en ciencias de la computación porque resultó tener propiedades muy interesantes para la codificación de algoritmos. Uno de los ejemplos más espectaculares está en el reconocimiento de imágenes. Seguramente han utilizado programas como Facebook o Google Picasa para compartir sus fotografías familiares. Si lo han hecho, se habrán dado cuenta de que estos programas tienen la capacidad de reconocer rostros de personas. La primera vez que suben una foto, por ejemplo la foto de "Juan", el programa pregunta quién es esa persona, y entonces uno contesta que es "Juan". A partir de ese momento, cada vez que se sube una foto de "Juan.^{el} programa reconoce que es "Juanz lo etiqueta como "Juan". Incluso si las fotos son diferentes y en algunas Juan está de perfil, en otras de frente, en otras con lentes, con pelo largo o pelo corto, con barba o sin barba, etc., el programa rara vez se equivoca y cada vez que se sube una foto de "Juan"lo reconoce bien y lo etiqueta como "Juan". ¿Cómo hace el programa para reconocer las fotos de "Juan" (o de cualquier otra persona)? O mejor dicho, ¿Qué algoritmo utilizaron los programadores para que el programa reconozca imágenes de la misma persona incluso en posiciones distintas?

Precisamente, el modelo de McCulloch-Pitts dado en

la Eq. (24) es el que está detrás del reconocimiento de imágenes. Como vimos en la Fig. 10, partiendo de una condición inicial arbitraria en t = 0, después de un tiempo el sistema llega a un estado congelado que muestra un patrón de dominos prendidos y apagados sin forma definida (como el panel que corresponde a t = 100 en la Fig. 10). Pero imaginemos ahora que quisiéramos que el sistema, en lugar de congelarse en un patrón amorfo, se congele en una imagen bien definida, como se muestra en la Fig. 15. ¿Qué tendríamos que hacerle a la red neuronal para que, partiendo de una condición inicial arbitraria, después de un tiempo el sistema llegue a una imagen predefinida por nosotros? La respuesta a esta pregunta la dio John Hopfield en 1982 y tiene que ver con seleccionar adecuadamente los pesos sinápticos w_{n,n_i} .

Supongamos que digitalizamos una imagen (como la del perro mostrada en la Fig. 15) convirtiéndola en una cadena binaria de 1's y 0's, de tal manera que después de digitalizar la imagen la podemos representar como una cadena binaria del tipo $\{0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, \ldots, 0, 1\}$. De hecho, para nuestros propósitos, es más conveniente digitalizar la imagen utilizando 1's y -1's (en lugar de 1's y 0's) de tal manera que la imagen digitalizada sea de la forma $\{-1, 1, 1, -1, -1, 1, -1, -1, -1, -1, 1, \ldots, -1, 1\}$. A esta cadena la vamos a llamar *patrón* y lo vamos a representar como

$$\vec{\xi} = \{\xi_1, \xi_2, \xi_3, \dots, \xi_N\}$$

donde cada ξ_n es uno de los bits de la imagen digitalizada, la cual consiste en total de N bits. Podemos ahora considerar que estos bits corresponden al estado estacionario de las neuronas del modelo de McCulloch-Pitts. Es decir, que esta secuencia de bits corresponde al patrón congelado al que llega la red después de un cierto tiempo t_0 . Si esto es así, entonces tendríamos $\sigma_n(t) = \xi_n$ para todo $t > t_0$. Esto significa que el patron $\vec{\xi} = \{\xi_1, \xi_2, \xi_3, \dots, \xi_N\}$ se convierte en un *punto fijo* de la Eq. (24), de tal manera que si alimentamos el lado derecho de la Eq. (24) con los bits del patrón $\vec{\xi}$ lo que obtenemos del lado izquierdo son los bits del mismo patrón $\vec{\xi}$:

$$\xi_n = \operatorname{Sgn}\left[\frac{1}{k_n} \sum_{j=1}^{k_n} w_{n,n_j} \xi_{n_j}\right], \qquad (27)$$

La pregunta es: dado un patrón $\vec{\xi} = \{\xi_1, \xi_2, \xi_3, \dots, \xi_N\}$ predefinido por nosotros, ¿cómo se deben escoger los pesos w_{nj} y la conectividad de cada nodo k_n para que se cumpla la ecuación anterior? En cuanto a la conectividad, Hopfield trabajó con redes totalmente conectadas en las que cada neurona σ_n está conectada con todas las demás, de tal manera que $k_n = N$ para todas las neuronas. Sin embargo, esto no es necesario y se puede hacer que cada neurona esté conectada solo con otras K neuronas, con K < N, aunque la eficiencia de la red para reconocer imágenes aumenta entre más grande sea K. Lo que es más interesante, es la forma de escoger los pesos sinápticos w_{n,n_j} . Hopfiel se dio cuenta de que para satisfacer la Eq. (27) los pesos se deben escoger como

$$w_{n,n_j} = \xi_n \xi_{n_j} \tag{28}$$

pues de esta manera se tiene

$$\operatorname{Sgn}\left[\frac{1}{K}\sum_{j=1}^{K}w_{n,n_{j}}\xi_{n_{j}}\right] = \operatorname{Sgn}\left[\frac{1}{K}\sum_{j=1}^{K}\xi_{n}\xi_{n_{j}}\xi_{n_{j}}\right]$$
$$= \operatorname{Sgn}\left[\frac{\xi_{n}}{K}\sum_{j=1}^{K}\xi_{n_{j}}^{2}\right]$$
$$= \operatorname{Sgn}\left[\frac{\xi_{n}}{K}\sum_{j=1}^{K}1\right]$$
$$= \operatorname{Sgn}\left[\xi_{n}\right]$$
$$= \xi_{n}$$

por lo que la Eq. (27) se satisface.

Por lo tanto, si queremos que la imagen codificada en el patrón $\vec{\xi} = \{\xi_1, \xi_2, \xi_3, \dots, \xi_N\}$ sea un estado estacionario de la dinámica dada en la Eq. (24), entonces se tienen que definir los pesos sinápticos w_{n,n_i} como en la Eq. (28).

La Fig. 16 muestra diferentes condiciones iniciales que convergen a la imagen del perro. En este caso, no importa con qué condición inicial comencemos la dinámica, la red siempre va a converger a la imagen del perro porque es elúnico patrón codificado en los pesos sinápticos. Bueno, de hecho no es el único. También está codificado el *negativo* del patrón: $\vec{\xi'} = -\vec{\xi}$, como puede observarse en el último panel de la Fig. 16. Esta es un propiedad interesante: si el patrón $\vec{\xi}$ es un estado estacionario de la diámica determinada por la Eq. (24), entonces el negativo de este patrón también lo es. Lo interesante de la Fig. 16 es que muestra que si comenzamos la dinámica con una condición inicial suficientemente parecida al patrón codificado $\vec{\xi},$ entonces la red va a converger a dicho patrón. Solo cuando comenzamos la dinámica con una condición incial totalmente aleatoria, a veces la red converge al patrón $\vec{\xi}$ mientras que otras veces la red converge al patrón negativo $-\vec{\xi}$.

 ${}_{\dot{c}}$ Qué pasa si que
remos codificar más de un patrón en la red? Supongamos que tenemos entonces un conjunto de
 M patrones

$$\vec{\xi}^{1} = \{\xi_{1}^{1}, \xi_{2}^{1}, \xi_{3}^{1}, \dots, \xi_{N}^{1}\}$$
$$\vec{\xi}^{2} = \{\xi_{1}^{2}, \xi_{2}^{2}, \xi_{3}^{2}, \dots, \xi_{N}^{2}\}$$
$$\vdots$$
$$\vec{\xi}^{M} = \{\xi_{1}^{M}, \xi_{2}^{M}, \xi_{3}^{M}, \dots, \xi_{N}^{M}\}$$

y queremos que cada uno de estos patrones sea un estado estacionario, es decir queremos que se cumpla la ecuación

$$\xi_n^{\mu} = \operatorname{Sgn}\left[\frac{1}{K} \sum_{j=1}^K w_{n,n_j} \xi_{n_j}^{\mu}\right]$$
(29)



Figura 16. Diferentes condiciones iniciales (cuadros en la columna izquierda en t = 0) convergen a la imagen del perro (cuadros a la derecha en t = 10). En el primer caso una condición inicial aleatoria converge a la imagen del perro. En el segundo caso, la condición inicial es la imagen del perro con un 20% de los pixeles modificados aleatoriamente. En el tercer caso, la condición inicial es la imagen del perro modificada, con lentes dibujados e incluso incompleta pues le faltan las esquinas superior izquierda e inferior derecha. Aún así, la red converge a la imagen del perro. En el último caso, una condición inicial aleatoria converge a la imagen opuesta (el negativo de la imagen original).

En un momento de inspiración casi divina, Hopfield generalizó la regla Eq. (28) que define los pesos sinápticos de la siguiente manera:

$$v_{n,n_j} = \frac{1}{M} \sum_{\mu=1}^{M} \xi_n^{\ \mu} \xi_{n_j}^{\ \mu} \tag{30}$$

De esta forma, Hopfield logró que cada patrón ξ^{μ} sea un estado estacionario de la red (Eq. (29)). Como ejemplo, la Fig. 17 muestra el caso de una red neuronal que tiene cuatro patrones codificados (imágenes derechas). Las



Figura 17. Las imágenes a la derecha son los cuatro patrones codificados en los pesos sinápticos de la red. Las imágenes de la izquierda son las condiciones iniciales de la red, las cuales se han escogido parecidas a los patrones codificados pero con distintas modificaciones y deformaciones. En cada caso, la red converge al patrón más parecido a la condición inicial correspondiente. Se dice que en este caso la red está *reconociendo* adecuadamente los patrones que tiene codificados.

imágenes en la izquierda son condiciones iniciales con las que se alimenta la red, y las flechas indican el patrón al que convergió la condición inical después de 10 pasos de tiempo. En esta figura puede verse que si la condición inicial es suficientemente parecida al patrón $\vec{\xi}^{\mu}$, entonces la red converge al patrón $\vec{\xi}^{\mu}$. En otras palabras, si la condición inicial es parecida al patrón "perro", entonces la red converge al patrón "perro". Si la condición inicial es parecida al patrón "niño, entonces la red converge al patrón "niño", y así sucesivamente. Se dice entonces que la red está *reconociendo* los patrones codificados en los pesos sinápticos w_{n,n_j} . Este es el principio de reconocimiento facial que opera en los programas Facebook y Google Picasa. Estos programas tienen redes neuronales que hacen el trabajo de reconocimiento de la imagen. La primera vez que uno sube la imagen de "Juan" al programa, la red neuronal codifica en sus pesos sinápticos dicha imagen. A partir de entonces, cada que uno sube una foto distinta de "Juan" (pero suficientemente parecida a la imagen original), la red neuronal toma la nueva imagen como condición inicial y converge al patrón $\vec{\xi}^{\ \mu}$ más parecido que está codificado en los pesos sinápticos (que en este caso sería la imagen de "Juan" codificada originalmente). Se dice entonces que la red está *reconociendo* la imagen de "Juan".

Cada que subimos la imagen de una nueva persona al Facebook, la red neuronal subvacente codifica dicha imagen en sus pesos sinápticos de acuerdo con la Eq. (30). Podemos hacer entonces que la red reconozca las imágenes de "Juan", "Pedro", "Luisa", etc. Sin embargo, cuando ya hay muchos patrones que se tienen que reconocer (codificados en los pesos sinápticos), la red comienza a cometer errores. En tal caso, podría ocurrir que al subir una nueva foto de "Juan" la red se confunda y converja al patrón de "Pedro". Hopfield entonces se preguntó cuál es el número máximo de patrones que se pueden codificar en una red neuronal sin que la red se confunda. O expresado de manera más cuantitativa, ¿cuál es el número máximo de patrones, M_{max} , que se pueden codificar en una red neuronal de tamaño N de tal forma que al dar una condición inicial parecida al patrón $\vec{\xi}^{\mu}$, la red converja a este patrón con una probabilidad del 98 %? La respuesta, que calculó el mismo Hopfield, es

$$M_{max} = 0,138K$$

donde K es el número de conexiones de cada neurona. Recordemos que Hopfield trabajó con redes completamente conectadas en las que K = N. Este es el caso que permite guardar la mayor cantidad de patrones en la red y reconocerlos satisfactoriamente. Para K < N el número máximo de patrones que se pueden reconocer sin cometer errores es menor. Resulta que al ir aumentando el número de patrones codificados en la red, se pasa bruscamente de un régimen en donde la red puede reconocer casi todos los patrones sin cometer errores, a otro régimen en donde no puede reconocer prácticamente nada. Es una transición discontinua.

Para medir la capacidad que tiene la red para reconocer los patrones se da una condición inicial $\vec{\Sigma}_0 = \{\sigma_1(0), \sigma_2(0), \ldots, \sigma_N(0)\}$ parecida al patrón $\vec{\xi}^{\mu} = \{\xi_1^{\mu}, \xi_2^{\mu}, \ldots, \xi_N^{\mu}\}$ (digamos con un 10% de bits diferentes). Se hace evolucionar la red de acuerdo con la Eq. (24) y se ve a qué estado llega después de un tiempo lo suficientemente largo t. Se define el traslape entre el patrón $\vec{\xi}^{\mu} = \{\xi_1^{\mu}, \xi_2^{\mu}, \ldots, \xi_N^{\mu}\}$ y el estado $\vec{\Sigma}_t = \{\sigma_1(t), \sigma_2(t), \ldots, \sigma_N(t)\}$ al que llegó la red como:

$$m_{\mu} = \lim_{t \to \infty} \frac{1}{N} \sigma_n(t) \xi_n^{\mu}$$

(Se toma el límite $t \to \infty$ para gatantizar que la red llegó al estado estacionario). Claramente, si $m_{\mu} \sim 0$ quie-



Figura 18. Traslape promedio ψ como función del número M de patrones codificados en la red. Existe un valor M_{max} tal que si $M < M_{max}$ la red reconoce los patrones casi perfectamente ($\psi \approx 1$). Sin embargo, cuando $M > M_{max}$, la red comienza a fallar terriblemente en el reconocimiento de patrones ($\psi \approx 0$). El valor de M_{max} está dado por $M_{max} = 0,138K$, siendo K la conectividad de cada neurona en la red. Hopfield trabajo con redes totalmente conectadas para las cuales K = N.

re decir que el estado $\vec{\Sigma}_t$ al que llegó la red es muy diferente del patrón $\vec{\xi}^{\ \mu}$ al que queríamos que llegara. En este caso la red no está reconociendo al patrón que debería llegar. Por otro lado, si $m_{\mu} \sim 1$ quere decir que el estado $\vec{\Sigma}_t$ se parece mucho al patgrón $\vec{\xi}^{\ \mu}$ y por lo tanto la red está haciendo bien el reconocimiento. El parámetro de orden que determina si la red está reconociendo bien o no los patrones codificados se define como

$$\psi = \frac{1}{M} \sum_{\mu=1}^{M} m^{\mu}$$

es decir, ψ es simplemente el promedio del traslape para cada patrón. La Fig. 18 muestra como cambia ψ al aumentar en número de patrones en la red. Vemos que cuando $M \leq M_{max} = 0,138N$, el reconocimiento es casi perfecto pues $\psi \approx 1$. Sin embargo, para $M > M_{max}$ el reconocimiento cambia discontinuamente y se hace pésimo, ya que en este caso $\psi \approx 0$. Por lo tanto, cuando se va aumentando el número de patrones codificados en la red, la capacidad de reconocimiento sufre una transición de fase discontinua: O es muy buena para $M < M_{max}$, o es muy mala para $M > M_{max}$.

Cuando vean que Facebook o Google Picasa (o cualquier otro programa para compartir y organizar fotos) ya no puede reconocer bien los rostros de sus amigos o familiares, es porque ya excedieron la capacidad de la red para recnocer patrones. En tal caso deberán borrar algunos de los patrones codificados en la red seleccionando sólo aquellos que realmente les importan (no llenen sus álbumes con la foto del amigo del amigo del conocido del primo del vecino). Sin embargo, los algoritmos de reconocimiento de imágenes en estos programas son tan eficientes, que llegar a saturar la red neuronal subyacente requeriría un gran esfuerzo de su parte (jaunque no dudo que algunos lo puedan lograr!).

XIII. CONCLUSIONES

La teoría de redes complejas es un campo de gran actividad científica en la actualidad. Basta con hacer una búsqueda en Google con la frase "Complex Networks" para darnos cuenta de la gran cantidad de investigadores que se encuentran trabajando en este campo. En estas notas introductorias hemos visto sólo algunas de las propiedades importantes de las redes complejas, pero la teoría no se termina aquí. Actualmente uno de los principales desafios es caracterizar las propiedades dinámicas de las redes complejas dado que conocemos su topología. Por ejemplo, la estructura de la red de regulación genética de Escherichia coli es conocida. Sin embargo, aún no podemos predecir los diferentes fenotípos que resultan de la dinámica de dicha red. Probablemente la persona que resuelva este problema se ganará el premio Nobel. El bioquímico Albert L. Lehninger dijo alguna vez lo siguiente: "La materia no sólo interactua, también se organiza. Conocemos rasonablemente bien las leyes de interacción de la materia, pero desconocemos por completo sus leyes de organización". Las redes complejas plantean la posibilidad, por primera vez en nuestra historia científica, de encontrar las leves de organización de la materia, si es que tales leves existen. El hecho de que redes tan diferentes como la red genética de una bacteria y la red de actores de Holywood presenten el mismo tipo de topología libre de escala sugiere que puede existir una ley fundamental que determine este tipo de estructuras. O que un modelo para explicar la formación de opiniones en una sociedad también pueda utilizarse, casi sin cambios, para reconocer imágenes en una red neuronal, sugiere que las mismas leves dinámicas podrían estar actuando en muy diversas escalas. Dichas leyes no se han encontrado aún y tal vez nunca se encuentren, pero el viaje hacia su descubrimiento ha resultado más que divertido. Espero con estas notas haberlos motivado para que se unan al estudio del fascinante mundo de las redes complejas y a la búsqueda de las leves de organización de la materia.

APÉNDICES

Apéndice A: Distribución de Poisson como límite de una distribución binomial

En este apéndice mostramos cómo obtener la Eq. (6) a partir de la Eq. (5), tomando el límite $N \to \infty$ y $M \to \infty$. Para esto utilizaremos dos resultados bien conocidos: la fórmula de Stirling y la definición de e^a :

$$n! \approx e^{-n} n^n$$
 para n grande,
 $e^a = \lim_{n \to \infty} \left(1 + \frac{a}{n} \right)^n$

Utilizando la fórmula de Stirling, el coeficiente binomial $\binom{N-1}{k}$ que aparece en la ecuación (5) puede escribirse, para N grande, como

$$\binom{N-1}{k} = \frac{(N-1)!}{k!(N-1-k)!}$$
$$\approx \frac{e^{-(N-1)}(N-1)^{N-1}}{k! \ e^{-(N-1-k)}(N-1-k)^{N-1-k}}$$
$$= \frac{e^{-k}(N-1-k)^k}{k!} \left(\frac{N-1}{N-1-k}\right)^{N-1}$$

Por lo tanto, la distribución binomial (5) queda, para ${\cal N}$ grande, como

$$P(k) = \frac{e^{-k}(N-1-k)^k}{k!} \times \left(\frac{N-1}{N-1-k}\right)^{N-1} (p_e)^k (1-p_e)^{N-1} (A1)$$

Para continuar, notemos que la cantidad z = 2M/N es la conectividad promedio de cada nodo en la red. En términos de esta cantidad, la probabilidad p_e puede escribirse como $p_e = z/(N-1)$, con lo cual la ecuación (A1) queda como

$$P(k) = \frac{e^{-k}(N-1-k)^k}{k!} \left(\frac{N-1}{N-1-k}\right)^{N-1} \\ \times \left(\frac{z}{N-1}\right)^k \left(1-\frac{z}{N-1}\right)^{N-1}$$

lo cual, después de arreglar algunos términos, puede escribirse como

$$P(k) = \frac{e^{-k}}{k!} z^k \left(1 - \frac{k}{N-1} \right)^k \left(\frac{1 - z/(N-1)}{1 - k/(N-1)} \right)^{N-1}$$

tomando el límite $N \to \infty$ y $M \to \infty$ de tal forma que la conectividad promedio z = 2M/N permanezca constante, y utilizando el hecho de que $e^a = \lim_{n \to \infty} (1 + a/n)^n$, obtenemos finalmente

$$P(k) = e^{-z} \frac{z^k}{k!}$$

Por lo tanto, para una red muy grande, el proceso de "hilvanar" parejas de botones escogidas al azar genera una distribución de conexiones de Poisson.

Apéndice B: Ecuación Maestra

Supongamos que tenemos un sistema que puede estar en cualquiera de los estados E_1, E_2, \ldots, E_N . Sea $W_{m\to n}(t)$ la probabilidad condicional de que el sistema, dado que estaba en el estado E_m al tiempo t, "brinque" al estado E_n . Las probabilidades $W_{m\to n}(t)$ se denominan probabilidades de transición.

Sea P(n, t) la probabilidad de que el sistema se encuentre en el estado E_n al tiempo t. Esta probabilidad puede cambiar en el tiempo por dos factores:

- 1. El sistema estaba en el estado E_n al tiempo t y brincó a otro estado E_m , lo cual claramente disminuye la probabilidad P(n, t+1) de encontrar al sistema en el estado E_n al tiempo t+1.
- 2. El sistema estaba en algún estado E_m al tiempo t y brincó al estado E_n , lo cual aumenta la probabilidad P(n, t+1) de encontrar al sistema en el estado E_n al siguiente instante de tiempo.

Tomando en cuenta estos dos factores, la probabilidad de que el sistema se encuentre en el estado E_n al tiempo t + 1 está dada por

$$P(n,t+1) = \sum_{m=1}^{N} P(m,t) W_{m \to n}(t) - \sum_{m=1}^{N} P(n,t) W_{n \to m}(t)$$
(B1)

El primer término del lado derecho nos dice que al tiempo t el sistema pudo estar en el estado E_m y brincar al estado E_n , aumentando así la probabilidad de estar en E_n al tiempo t+1. El segundo término toma en cuenta el hecho de que al tiempo t el sistema pudo estar en el estado E_n y salirse de allí, brincando al estado E_m y disminuyendo la probabilidad de estar en E_n al tiempo t+1. Las sumas sobre m toman en cuenta todos los posibles estados hacia o desde los que puede brincar el sistema.

La ecuación (B1) es la ecuación maestra del sistema. El objetivo es resolverla para encontrar P(n,t). Si las probabilidades de transición $W_{m\to n}$ son independientes del tiempo, la ecuación se puede resolver más o menos fácilmente. Pero la cosa se complica enormemente si las probabilidades de transición dependen del tiempo. Afortunadamente, las ecuaciones maestras que aparecen en la teoría de crecimiento de redes, al menos en los casos sencillos presentados en estas notas, son relativamente fáciles de resolver.

^[1] En la red del universo de Malvel, dos super héroes están conectados si han aparecido por lo menos una vez en el

mismo comic. De este estudio, que puede encontrarse en http://dmi.uib.es/~joe/marvel.html, resulta que es el hombre

araña el super héroe con el máximo número de conexiones.

- [2] Statistical Mechanics of Complex Networks. Réka Albert y Albert-László Barabási. *Reviews of Modern Physics* 74(1):47-97 (2002).
- [3] The Structure and Function of Complex Networks. Mark E.J. Newman. SIAM Review 45(2):167-256 (2003).
- [4] Evolution of Networks. S.N. Dorogovtsev and J.F.F. Mendes. Advances in Physics 51(4):1079-1187 (2002).
- [5] Evolution of Networks: From Biological Nets to the Internet and WWW. S.N. Dorogovtsev and J.F.F. Mendes. Oxford University Press, Oxford. ISBN: 0198515901. (2003). Este libro puede encontrarse GRATIS en http://www.fyslab.hut.fi/~sdo/
- [6] Connectivity of Growing Random Networks. P. L. Krapivsky, S. Redner, y F. Leyvraz. Phys. Rev. Lett. 85:4629-4632 (2000).
- [7] Power-Law Distribution of the Word Wide Web. L. A. Adamic y B. Hubeerman. *Science* 287(24):2115a (2000).
- [8] Bose-Einstein condensation in complex networks. G. Bianconi y A.-L. Barabsi. *Phys. Rev. Lett.* 86:5632-5635 (2001).
- [9] Epidemic Spreading in Scale-Free Networks. Romualdo Pastor-Satorras y Alessandro Vespignani. *Phys. Rev. Lett.* 86:3201-3203 (2001).
- [10] Spread of Epidemic Disease on Networks. Mark E. J. Newman. Phys. Rev. E. 66:16128 (2002).
- [11] Dynamical phase transition in a neural network model with noise: An exact solution. Cristián Huepe y Maximino Aldana.

Jour. Stat. Phys. 108:527-540 (2002).

- [12] La anemia falciforme, por ejemplo, es una enfermedad de los glóbulos rojos de la sangre que se origina por una sóla mutación en uno de los más de 600 aminoácidos que conforman a la proteína β -globina encargada de capturar oxígeno.
- $\left[13\right]$ En el contexto de las redes complejas, la palabra "topología.
- [14] A las distribuciones libres de escala también se les llama distribuciones de potencia.
- [15] Las redes con topología exponencial se parecen mucho a las de Poisson.
- [16] En inglés se les llama hubs.
- [17] LOGWIN es la concatenación de "log.
- [18] Análogamente, la distribución de estaturas en los habitantes del D.F. se ha mantenido estacionaria por muchos a nos aún cuando la población de chilangos no ha dejado de aumentar.
- [19] Estríctamente hablando, la definición correcta es $\psi = \left\langle \lim_{t \to \infty} \psi(t) \right\rangle$, donde los paréntesis angulares $\langle \cdot \rangle$ denotan el promedio sobre todas las posibles condiciones iniciales del sistema. Al tomar este promedio, las fluctuaciones que se observan el la Fig. 13 desaparecen. Otra manera de definir a ψ , que es muy útil para calcularlo numéricamente, es $\psi = \lim_{T \to \infty} \frac{1}{T} \int_0^T \psi(t) dt$. La integral lo único que hace es "limpiar" las fluctuaciones producidas por el ruido.
Estudio y aplicación de los plasmas de baja temperatura.

F. Castillo, V. U. L. Contreras, B. Campillo, O. Flores y H. Martínez.

Durante las últimas décadas, la generación de plasmas para la gran diversidad de aplicaciones existente se realiza a través de descargas eléctricas en gases. Los mecanismos presentes en las descargas eléctricas en gases, han sido estudiados y clarificados en términos de los procesos fundamentales entre los electrones, átomos o moléculas, iones, fotones, paredes de la cámara, los electrodos, etc. En la Figura 1 se muestran los diferentes tipos de descarga como función del voltaje de operación y la densidad de corriente en la descarga entre dos electrodos paralelos. Dónde: $-\cdot - \cdot -$ indica la corriente con fotoemisión del cátodo. - - (CD') es la trayectoria positiva de la descarga de Townsend; donde el potencial de rompimiento (V_B) es menor que el potencial de inicio de la descarga luminosa (V_G). Y ——— (CD) es la trayectoria negativa de la descarga de Townsend; donde $V_B = V_G$.



Figura 1. Esquema de las características de la descarga entre electrodos paralelos [1].

En la construcción de este diagrama se ha considerado el producto de la presión (p) y la separación entre los electrodos (d) como un valor fijo, para gases raros es de: 2 < pd < 20 mm·cm.

La descarga para formar plasmas de baja temperatura se realiza típicamente en la zona de descarga luminosa normal. Donde es posible realizarla en varias condiciones de presión, y donde los diferentes procesos son fundamentales en la comprensión de los mecanismos de descarga. El límite entre alta y baja presión no es claro, básicamente depende de las condiciones experimentales, desde este punto de vista este se ubica entre los 20 y 200 mm de Hg de presión. Mientras que la diferenciación basada en la temperatura se establece a través de las diferencias de temperatura del gas (T_g), de los iones (T_i) y de los electrones (T_e), todas estas especies resultantes de los procesos fundamentales de ionización y excitación en el plasma. Se establece una división en tres tipos y se mencionan algunos ejemplos:

- Plasmas no térmicos o fríos: $T_i \cong T_g \cong 300-400$ °K; $T_i \ll T_e < 10^5$ °K (10 eV) Descarga de barrera, descarga de corona, microplasma, jets de plasma.
- Plasmas calientes "no térmicos": $T_i \ll T_e < 10^4 10^5 \,^{\circ}\text{K}$; $T_i \cong T_g < 4.10^3 \,^{\circ}\text{K}$ Antorcha de plasma, plasma inducidos por microondas, plasmas de superfície.
- Plasmas térmicos: T_i ≅ T_g ≅ T_e; T_x>5·10³ °K
 Plasma de arco, plasma de arco-jet, plasma focus.

El Laboratorio de Espectroscopia del Instituto de Ciencias Físicas de la UNAM ha adquirido experiencia en investigaciones en plasmas de baja temperatura. Las líneas de investigación de este laboratorio y en las que ya se tiene experiencia son:

- Diagnóstico de plasmas de resplandor.
- Estudio de plasmas de resplandor de interés atmosférico y astrofísico.
- Estudio de plasmas de resplandor de líquidos.
- Estudio de la interacción de plasmas con superficies y degradación de hidrocarburos por plasmas.
- Modificación superficial producida por los plasmas de películas delgadas con aplicaciones en celdas solares.

- Estudios de oxidación con plasmas.
- Estudio teórico-experimental de interacciones ión-átomo.
- Estudio de compuestos orgánicos volátiles (VOC's).

Es de gran interés el estudio de la interacción de las diversas fuentes de plasmas de baja temperatura con los materiales (metales, polímeros, polvos, semiconductores, etc.); así como el estudio del efecto de los diferentes substratos en los procesos químicos establecidos en el plasma. Los proyectos de investigación se realizan en diferentes cámaras de descarga, las cuales han sido adquiridas y equipadas en el Laboratorio de Espectroscopia del Instituto de Ciencias Físicas de la UNAM. En la Figura 2 se muestran dos esquemas de las cámaras construidas en el Laboratorio de Espectroscopia. Básicamente se requiere de una cámara, un sistema de vacío (10⁰-10⁻⁷ torr) y una fuente de poder (CA o CC). Las variables a controlar durante la generación del plasma son: presión, voltaje, corriente y composición química del gas o mezcla gaseosa. Estas cámaras pueden ser reconfiguradas para los diferentes estudios realizados. La primera se emplea en estudios de interacción de plasma con materiales sólidos, la segunda se emplea en estudios de descargas en gases.



Figura 2. Diagramas de cámaras de descarga de geometrías distintas.

En el estudio del plasma, es de interés dos áreas; ambas de vital importancia para comprender y aplicar esta tecnología:

1. Estudio de las características del plasma generado, este se realiza con la adecuada instrumentación de la cámara de descarga. La instrumentación instalada permite realizar

estudios de: temperatura y densidad de electrones, temperatura y densidad iónica, identificación y análisis de concentración de fragmentos moleculares producidos e identificación de especies producidas en equilibrio en el plasma. En la Figura 3 se muestra la cámara de descargas equipada con los tres sistemas de diagnóstico del plasma generado: sonda de Langmuir, espectrometría de masas y espectrometría de emisión óptica (OES, por sus siglas en ingles).

2. El estudio de la modificación fisicoquímica de la superficie producida por la interacción del plasma con las superficies de los materiales expuestos a ellos. Como ejemplo, esto se ha realizado en películas delgadas con aplicaciones en celdas solares y como recubrimientos duros, áreas de interés científico y tecnológico.



Figura 3. Cámara esférica de descargas.

En el desarrollo de lo anterior se requieren técnicas de análisis como son: microscopia óptica, microscopia electrónica, difracción de rayos x (XRD, por sus siglas en ingles), transformada de Fourier de espectroscopia de infrarrojo (FT-IR), ángulo de mojado, pruebas mecánicas (tensión, compresión, dureza, microdureza, nanodureza, etc.). Estas son algunas de las técnicas de análisis más comunes empleadas en el estudio de la modificación experimentada por los materiales expuestos a un plasma.

Algunas de las áreas de interés y de estudio realizados se ilustran a continuación.

Asfaltenos.

La producción mundial de crudo se enfoca principalmente en los aceites pesados, los cuales son ricos en compuestos pesados (resinas y asfáltenos). Estos compuestos son importantes en las propiedades de los crudos, pues son capaces de cambiar las condiciones de flujo al depositarse en las paredes de las tuberías como resultado de los cambios de presión y temperatura. La formación de depósitos de asfáltenos tiene serias implicaciones en el transporte de crudo, requiriéndose esfuerzos adicionales para mantener la eficiencia de producción. Una mayor cantidad de fracciones pesadas en los crudos ha forzado el desarrollo de la tecnología y de las estrategias para superar las limitaciones impuestas en los reactores de fraccionamiento, procesos de hidrogenación y catálisis. Los cambios químicos en las resinas y asfáltenos han sido estudiados al ser expuestos a diferentes gases, tratamientos térmicos en diferentes condiciones de presión y temperatura. Existe poca información de las comportamiento al ser expuestos a plasma de baja temperatura, donde la energía de los electrones es de pocos electrón-volts (eV), energía menor a la energía de ionización de las moléculas de los asfáltenos. El interés es conocer los cambios químicos que sufren los asfáltenos al ser expuestos a plasmas de baja temperatura de diferentes gases.

Se han estudiado los efectos sobre los asfáltenos al ser expuestos a plasmas de oxígeno, metano, hidrógeno y recientemente en argón. Cuando se exponen los asfáltenos al tratamiento con plasma se obtienen varias ventajas tecnológicas. Primero, el tratamiento favorece el rompimiento económico de las moléculas pesadas, en condiciones de baja energía, sin requerir de altas presiones y altas temperaturas. Segundo, los productos de la fragmentación molecular pueden ser usados en las últimas etapas de los procesos de refinación, incrementando la producción y mejorando el valor económico de las fracciones pesadas. Y finalmente, estas tecnologías de plasma de baja temperatura pueden ser empleadas en cualquier fracción de crudo independientemente de la naturaleza química del crudo disponible. Se requiere de mayor conocimiento de los efectos de esta tecnología y del desarrollo de tecnología para la implementación de este tipo de procesos. Además de mayores esfuerzos en la investigación para tratar de comprender la complejidad de los

procesos que ocurren a nivel molecular en el tratamiento de asfáltenos en plasma de baja temperatura.

La investigación en esta área se realiza desde la obtención de muestras adecuadas para este estudio. En el que se siguen métodos bajo norma de la American Society for Testing of Materials (ASTM) D6560-12,17. Para una vez obtenidos los asfáltenos, exponerlos a plasma de hidrogeno. Tanto el plasma como las especies químicas resultantes de la descomposición de las muestras son estudiados mediante espectroscopia de emisión óptica. La eficiencia de los procesos de descomposición y los cambios químicos se analizan como función del tiempo de exposición al plasma.

Inicialmente el plasma de hidrogeno (H_2) se caracteriza en su composición, identificándose la presencia de electrones, protones (H^+) , y H_2 , con energía cinética del orden de 30 eV. Estas especies pueden interactuar con los asfáltenos y producir cambios en la estructura química por diferentes mecanismos, rompimiento de enlaces, recombinación y condensación.

Las muestras de asfáltenos se exponen hasta por 120 minutos en el plasma de H_2 , estudiándose el proceso de interacción plasma-asfáltenos. La caracterización por OES identificó y comprobó la presencia de las siguientes especies: C⁺, CH⁺, C₂⁺, NH⁺ y V⁺. Estos resultados se muestran en la Figura 4.



Figura 4. Espectros de OES del sistema asfáltenos-plasma de H₂ [11].

Una técnica simple pero muy útil para seguir el proceso de interacción de los asfáltenos con el plasma, es la pérdida de peso al ser expuestos al plasma de H₂. En la Figura 5 se muestra la evolución del peso de la muestra con el tiempo de exposición. Inicialmente la perdida es relativamente rápida y casi a una tasa constante (0.63 mg/min) durante los primeros 45 min, seguido de una estabilización del peso después de los 120 min, donde el comportamiento mostrado coincide con un cambio insignificante de los espectros de OES.



Figura 5. Perfil de la pérdida de peso de los asfáltenos [11].

Los posibles cambios en la estructura de los asfáltenos residuales es posible analizarlos mediante la técnica de difracción de rayos X en las estructuras ordenadas en la fase remanente. En la Figura 6 se muestra el análisis diferencial realizado en los espectros de difracción de rayos X realizados en el material residual obtenido a los diferentes tiempos de exposición al plasma. La línea negra corresponde con el espectro de la muestra de asfáltenos sin tratamiento. En la Tabla I se muestran los resultados de los parámetros calculados a partir de los espectros de difracción de rayos X. Los parámetros que son posibles calcular son los siguientes:

factor de aromaticidad	$A_{002}/(A_{002} + A_{\gamma})$
distancia entre planos aromáticos	$\lambda/2 \ sen \ \theta$
distancia entre cadenas parafinadas	$\lambda/2 \ sen \ \theta$
diámetro promedio de los planos aromáticos, incluidos C-alfa	$0.92/B_{1/2}$ [110]
diámetro promedio de los cúmulos	$0.45/B_{1/2}[002]$
número de los planos aromáticos por cúmulo	$L_c/d_m + 1$
carbonos aromáticos por molécula	$L_a^2/2.62$
numero de carbonos aromáticos por unidad aromática	$(L_a + 1.23)/0.615$
incluyendo C-alfa.	
número de centros aromáticos por molécula	$C_{A/}C_{AU}$
	factor de aromaticidad distancia entre planos aromáticos distancia entre cadenas parafinadas diámetro promedio de los planos aromáticos, incluidos C-alfa diámetro promedio de los cúmulos número de los planos aromáticos por cúmulo carbonos aromáticos por molécula numero de carbonos aromáticos por unidad aromática incluyendo C-alfa. número de centros aromáticos por molécula



Figura 6. Patrones de difracción de rayos X. La línea negra corresponde a los asfáltenos sin tratamiento de plasma y los demás corresponden a la diferencia con los tratados a diferentes tiempos [11].

Parámotro	Tiempo de exposición al plasma de H _z				
Parametro	0	15	30	60	120
f_a	0.14	0.14	0.13	0.12	0.12
$d_{\rm m}$ (Å)	3.53	3.53	3.54	3.54	3.53
d_{γ} (Å)	5.87	5.86	5.93	5.94	5.88
L_a (Å)	20.46	20.42	20.44	15.86	15.15
L_{c} (Å)	28.76	28.92	28.65	31.04	30.72
M	9.15	9.19	9.10	9.77	9.69
CA	159.8	159.2	159.5	96.0	87.6
CAU	35.3	35.2	35.2	27.8	26.6
N	4.53	4.52	4.53	3.45	3.29

Tabla I. Parámetros estructurales obtenidos de los espectros DRX [11].

Se observan cambios en los parámetros estructurales como función del tiempo de exposición al plasma. La distancia interplanar de 3.53 Å de los planos aromáticos no muestran un cambio significante. Mientras que los planos parafinados se mueven alrededor del valor promedio de 5.85 Å, este incremento puede ser atribuido a impedimentos estéricos de las cadenas. Los cambios más apreciables se presentan en los diámetros de los

planos aromáticos, las distancias se reducen de 20.46 a 15.15 Å, indicando que el hidrogeno se inserta en los núcleos aromáticos. Además se puede presentar el rompimiento de los enlaces C-C de la parafina al evacuarse la cámara de descargas.

En la Figura 7. Se muestra el aspecto superficial de las partículas de asfáltenos, así como su análisis por espectroscopia de energía dispersa de rayos X (EDX, por sus siglas en inglés). Se aprecia un aspecto de fractura frágil y polvo fino adherido a la superficie.



Figura 7. Imágenes de electrones secundarios del microscopio electrónico de barrido (SEM, por sus siglas en inglés) del aspecto superficial de las partículas de asfáltenos.

El análisis químico semi-cuantitativo en porciento en peso obtenido por EDX se muestra en la tabla II. La composición química es similar a las otras muestras de otros asfáltenos. Durante el tiempo de exposición se encontró que las partículas de asfáltenos mostraban un el crecimiento de las partículas comparado con el aspecto presentado por el material de llegada. La apariencia que presentan es de haber cristalizado y las superficies parecen de un material fracturado de manera frágil.

Elemento	1	2
СК	90.42	91.00
O K	5.39	5.11
S K	4.00	3.68
V K	0.19	0.21
Totales	100.00	100.00

Tabla II. Composición química obtenida por EDX.

Los principales hallazgos de este trabajo que pueden establecerse son:

La pérdida de masa experimentada por los asfáltenos está asociada a dos procesos. Desorción térmica de los compuestos de baja volatilidad en el vacío de la cámara de descargas, resultado del calentamiento de la muestra durante la primera etapa del proceso debido a las interacción plasma-asfáltenos. La segunda corresponde con la degradación química de los asfáltenos, debido al rompimiento de los enlaces cuando el plasma interactúa con los asfáltenos. En esta etapa pueden presentarse diferentes procesos, hidrogenación de los núcleos aromáticos, apertura de los anillos de nafteno y activación de otros procesos de disociación debido a las colisiones entre el plasma y los asfáltenos. Se identificadas fueron: C⁺, CH⁺, C₂⁺, NH⁺, V⁺. Mediante la comparación de los productos de descomposición en el plasma, las especies identificadas fueron: C⁺, CH⁺, C₂⁺, NH⁺, V⁺. Mediante la comparación de los productos de descomposición en el plasma con el plasma de hidrógeno puro, las especies reportadas corresponden exclusivamente a la descomposición de los asfáltenos.

El análisis de difracción de rayos X es útil en la caracterización de los cúmulos de asfáltenos y su modificación física o química producida por el plasma. El análisis se realiza en base a los factores de forma determinados a partir de los espectros de DRX, como son los parámetros L_a y L_c . Se obtuvo evidencia de los cambios experimentados por las unidades aromáticas. Los cambios químicos y moleculares experimentados por los asfáltenos pueden ser obtenidos mediante la compilación de la información estructural obtenida por diferentes técnicas espectroscópicas (XDR, FT-IR, masas, EDX).

El plasma puede ser generado en un sinfín de condiciones experimentales, por ejemplo, es posible modificar: la presión de trabajo, tipo y energía de las partículas que pueden ser usadas satisfactoriamente para recuperar parte de la fracción pesada en forma de gases de fragmentos más ligeros. Se requiere más investigación en esta área para encontrar la forma de aplicar estas técnicas de manera industrial.

Compuestos orgánicos volátiles (VOC's).

Los VOC's se definen como los compuestos químicos orgánicos con una presión de vapor (en condiciones normales de presión y temperatura) suficientemente alta para vaporizarse en la atmosfera, el más común de estos compuestos es el metano (CH₄, gas de efecto invernadero, ver Figura 8. Los demás VOC's están formados por un amplio rango de moléculas orgánicas basadas en carbono, ejemplo de ellos son: aldehídos, cetonas, hidrocarbonos aromáticos y ligeros (xileno, tolueno), alcoholes, acetatos, bencinas, glicoles, formaldehidos, benceno, butadieno, hexano y muchos otros.



Figura 8. Moléculas de VOC's, CH₄ y C₈H₈.

Las fuentes de los VOC's son de diversa índole; hay fuentes naturales, como son las emisiones de gas natural, los árboles y los humedales por mencionar algunos; dentro de las fuentes artificiales de VOC's se encuentran las industrias de la energía, la manufactura, el transporte, la química, la agricultura, las industrias de la pintura (solventes de pintura, limpiadores de pinturas, etc.), biomasa, fugas de constituyentes de combustibles (gasolinas, diésel, gas natural, etc.). Otras fuentes a considerar son de tipo residencial y comercial, transporte de combustibles, fotocopiadoras, decoración de casas y oficinas, degradación de pinturas, humo de tabaco, etc.

Muchos de estos VOC's son altamente tóxicos y se han mejorado las técnicas de detección para controlar y limitar las concentraciones en el medio ambiente. La Administración de la Salud y Seguridad Ocupacional (Occupational Safety and Health Administration, OSHA) ha establecido valores límites permisibles de exposición de muchos de estos compuestos (permissible exposure limits, PELs), establecido para proteger a los trabajadores en las diferentes fuentes de VOC's. El PEL es la concentración máxima del contaminante al cual puede exponerse sin protección durante una jornada. Los PELs están listados en el estándar de contaminantes del aire 29 CFR 1910.1000, en su tabla Z-1, algunos de ellos se muestran en la Tabla III.

El Instituto Nacional para la Seguridad y Salud Ocupacional (National Institute for Occupational Safety and Health, NIOSH) ha establecido los límites de exposición recomendados (Recommended Exposure Limit, REL). Además, en la Conferencia Americana de higienistas Industriales Gubernamentales (American Conference of Governmental Industrial Hygienists) se establecio el Valor Límite Umbral (Threshold Limit Value®, TLV®) para diversos compuestos, muchos estándares de salud establecen el cumplimiento de estos TLV's como límites de exposición. En el caso de los VOC's altamente combustibles se ha establecido el valor mínimo de explosión (Lower Explosión Limit, LEL), este valor corresponde con la concentración mínima para formar una mezcla combustible con el aire.

Table III. Valores Límite y Constantes Físicas de VOC's.						
Contaminante	LEL (% Vol)	Temperatura de inflamación (°F)	PEL OSHA	REL NIOSH	TLV	5% LEL ppm
Acetona	2.5%	-4° F	1000 ppm TWA	250 ppm TWA	500 ppm TWA; 750 ppm STEL	1250 ppm
Butano	1.9%	-76° F	No Listado	800 ppm TWA	1000 ppm TWA	950 ppm
Cloroformo		> 1000 °C	25000 ppm/5 min		50 mg/m ³ (10 ppm TWA)	
Diésel (No. 2) vapor	0.6%	125° F	No Listado	No Listado	15 ppm	300 ppm
Etanol	3.3%	55° F	1000 ppm TWA	1000 ppm TWA	1000 ppm TWA	1650 ppm
Gasolina	1.3%	-50° F	No Listado	No Listado No Listado		650 ppm
Hexano	1.1%	-7º F	500 ppm TWA	50 ppm TWA	50 ppm TWA	550 ppm
Alcohol Isopropilico	2.0%	53° F	400 ppm TWA	400 ppm TWA; 500 ppm STEL	200 ppm TWA; 400 ppm STEL	1000 ppm
Queroseno/Gas avión	0.7%	100-162° F	No Listado	100 mg/m ³ TWA (approx. 14.4 ppm)	200 mg/m ³ TWA (approx. 29 ppm)	350 ppm
MEK	1.4%	16° F	200 ppm TWA	200 ppm TWA; 300 ppm STEL	200 ppm TWA; 300 ppm STEL	700 ppm
Pentane	1.5%	-40° F	1000 ppm TWA	120 ppm TWA; 610 ppm máx	600 ppm TWA	750 ppm
Fenol	1.8%	175° F	5 ppm TWA	5 ppm TWA; 15.6 ppm máx	5 ppm TWA	900 ppm
Estireno	0.9%	88° F	100 ppm TWA; 200 ppm máx; 600 ppm pico máx. Arriba del máx. 5 min/3 horas	50 ppm TWA; 100 ppm STEL	20 ppm TWA; 40 ppm STEL	450 ppm
Tolueno	1.1%	40° F	200 ppm TWA; 300 ppm máx.; 500 ppm pico máx. Arriba del máx. 10 min/8-hr	100 ppm TWA; 150 ppm STEL	50 ppm TWA	550 ppm
Turpentino	0.8%	95° F	100 ppm TWA	100 ppm TWA	20 ppm TWA	400 ppm
Xilenos (o, m & p isomeros)	0.9- 1.1%	81-90° F	100 ppm TWA	100 ppm TWA; 150 ppm STEL	100 ppm TWA; 150 ppm STEL	450-550 ppm

La descomposición química de los VOC's se realiza tradicionalmente por tres rutas principalmente: procesos térmicos, procesos de filtrado/adsorción y procesos de oxidación no térmica, cada una de estas con variantes.

En los últimos años se ha dado gran impulso a tres técnicas:

 Procesos asistidos por plasma (tratamiento de flujo de gas con un haz de electrones, inyección de ozono (LoTOx), Pre-tratamiento húmedo (ECO)). Oxidación de NO a NO₂, N₂O₄ y N₂O₅ y reducción con NH₃)

- Procesos basados en plasma (Precipitación electrostática (mecánica-filtración de polvos), deodorización, remoción de VOC's por plasma no-térmicos)
- Incineración basada en plasma (Destrucción térmica de desechos sólidos y gases peligrosos por plasma caliente).

El uso de plasma además de ser capaz de descontaminar, también se emplea para evaluar los productos que se forman en la naturaleza al interactuar los VOC's con la atmósfera y posteriormente analizar los riesgos de los productos formados. Los VOC's han sido estudiados mediante plasma poniendo énfasis en la química del plasma, de estos se cuenta con algunos ejemplos de estudio.

Descarga de plasma en una atmosfera de cloroformo (CHCl₃).

El objetivo de este trabajo es investigar el comportamiento del plasma de cloroformo. Esto permitiría proponer formas de degradar o descomponer el cloroformo en especies más simples o menos dañinas. El principal problema con los VOC's clorinados, es su alta toxicidad y carcinogénesis. El protocolo de Kyoto obliga a los países industrializados participantes a disminuir los gases de efecto invernadero. Esto forma parte de la motivación para este tipo de estudios, para encontrar métodos alternos para eliminar este toipo de contaminantes de las emisiones industriales.

El cloroformo es líquido, incoloro y con un olor característico y por supuesto muy volátil (presión de vapor (mm de Hg): 0.825 (a -60 °C), 2.03 (a -50 °C), 4.73 (a -40 °C), 9.98 (a - 30 °C). Generalmente contiene pequeños porcentajes de etanol (1-5 %) como estabilizador. Es soluble parcialmente en agua (10.62 g/kg (a °C), 8.22 g/kg (a 20 °C) y 7.76 g/kg (a 30 °C) y con mayor densidad. Es no inflamable, pero los productos de su oxidación son muy peligrosos (fosgeno, cloruro de hidrógeno, cloro y óxidos de carbono y cloro. Todos ellos productos corrosivos y muy tóxicos). Es peligroso por inhalación e ingestión. Se obtiene por medio de cloración controlada del metano o por tratamiento de acetona con polvos de CaOCl₂ y ácido sulfúrico. Fue descubierto en 1847 y se utilizó como anestésico por inhalación, como insecticida y en la industria farmacéutica, dada su toxicidad ha sido

reemplazado por otras sustancias. Actualmente, es utilizado como intermediario en síntesis orgánica, especialmente en la obtención de fluorocarbono 22, el cual es utilizado como refrigerante, propelente y en la fabricación de tetrafluoroetileno y su polímero (PTFE).

Se descompone a temperatura ambiente por acción de la luz del sol en ausencia de aire y en la oscuridad en presencia de este último, siendo uno de los productos de esta descomposición el fosgeno, el cual es muy tóxico. Está clasificado como moderadamente tóxico, sin embargo está considerado como posible carcinogénico humano. Una probable dosis letal para humanos es de 0.5 a 5 g/Kg. Sin embargo, se sospecha que es carcinógeno para humanos. Puede causar una muerte rápida, atribuida a paro cardiaco y una muerte lenta por daño al hígado y riñón.

Los análisis por OES del plasma de cloroformo generado entre electrodos de Cu, mostraron la presencia de las siguientes especies en el plasma: Cl^+ , HCl^+ , CH^+ , C^+ , Cl, C and C₂, ver Figura 9. Una alternativa a la sonda de Langmuir para evaluar la temperatura (T_e) y densidad electrónica (n_e), es emplear combinadas las ecuaciones de ionización de Saha y las ecuaciones de distribución de excitación de Boltzman [21,22]. La T_e para dos diferentes líneas de emisión óptica del mismo átomo está dada por:

$$T_e = \{ (E_2 - E_1) / k \} [ln (I_1 \lambda_1 g_2 A_2) / (I_2 \lambda_2 g_1 A_1)]^{-1}$$

Donde los subíndices 1 y 2 corresponden a cada línea de emisión observada. Mientras que la ecuación de equilibrio de Saha para n_e es:

$$n_e = (2g_2A_2I_1\lambda_1 / g_1A_1I_2\lambda_2) (2\pi m_e kT_e / h^3) \exp\{(E_1 - E_2) / kT_e\}$$

dónde:

 g_1 y g_2 son los pesos estadísticos o degeneración de los niveles 1 y 2;

 E_1 y E_2 son las energías de los niveles 1 y 2;

 λ_1 y λ_2 son las longitudes de onda de las líneas emitidas;

 I_1 y I_2 son las intensidades de emisión de las líneas 1 y 2;

 A_1 y A_2 son las probabilidades de transición de Einstein;

 T_e es la temperatura de los electrones

 n_e es la densidad de electrones;

 m_e es las masa del electrón y k is the Boltzmann constant, $k = 8.6173 \times 10^{-5} \text{ eV}/^{\circ}\text{K}$.

Los valores de energía, la degeneración y los valores de probabilidad de transición de Einstein son tomadas de la base de datos de espectros atómicos de NIST (Atomic Spectra Database) [23]. Las líneas empleadas para evaluar T_e fueron C 504.9 nm and C 1038.9 nm de los espectros de OES obtenidos del plasma de cloroformo. Los resultados obtenidos para el plasma de CHCl₃ son: $T_e = 5.74\pm0.57$ eV y $n_e = (4.48\pm0.45) \times 10^{10}$ cm⁻³



Figura 9. OES del plasma de cloroformo a 1 torr de presión.

En este caso se formó un depósito de color oscuro sobre los electrodos de cobre. Depósito que fue recuperado después de 30 horas de exposición al plasma de cloroformo y fue analizado por SEM (ver Figura 10), EDX (ver inserto en Figura 10) y DRX (ver Figura 11).

En el análisis por EDX se identificaron los elementos Ca, Cu, O y Cl. El análisis semicuantitativo obtenido mediante la regresión del espectro por el método ZAF se muestra en la tabla IV. Este análisis muestra que existe una relación atómica entre el Cu y el Cl con valores de entre 2 y 4.



Figura 10. Imagen de electrones secundarios del SEM, Se observan las partículas formadas por el depósito formado sobre el electrodo de Cu.

	Región 1		Región 2		
Elemento	% en	%	% en	%	
	peso	Atómico	peso	Atómico	
0	13.41	34.69	9.46	27.47	
Cl	16.70	19.50	10.49	13.75	
Ca	0.72	0.74	0.56	0.65	
Cu	69.18	45.07	79.49	58.13	

Tabla IV. Análisis semi-cuantitativo de dos regiones del depósito.

Debido a la presencia de Cl y Cl⁺, especies identificadas por OES en el plasma de CHCl₃, el Cu reacciona con estas para formar CuCl. La reacción que ocurre es Cu + Cl \rightarrow CuCl, compuesto formado sobre el electrodo de Cu. Este compuesto es fotosensitivo y cambia a Cu₂Cl (color oscuro), posteriormente, una vez extraídos los electrodos de la cámara y expuestos al aire y a la humedad del medio ambiente, el Cu₂Cl (polvo oscuro) reacciona en presencia del H₂O y el O₂ para formar Cu₂Cl(OH)₃ (polvo verde). Esta reacción correlaciona bien con los resultados obtenidos por OES, ver Figura 9 y Tabla IV. En el espectro de OES puede verse que la especie dominante en el plasma de cloroformo es el C y que el Cl y Cl^+ estan presentes en menor proporción. Esto se debe a la reacción entre el Cu y el Cl para formar el CuCl y posteriormente el Cu₂Cl sobre el electrodo de Cu. Esta reacción reduce la concentración de Cl en el plasma de cloroformo.



Figura 11. Espectro de DRX del depósito recuperado de los electrodos de Cu.

Finalmente en condiciones atmosféricas se forma el Cu₂Cl(OH)₃, este compuesto presenta cuatro 4 diferentes formas cristalinas; el primero con estructura monoclínica (botalaquita) la cual se transforma en una estructura monoclínica-pseudo romboédrica (clinoatacamita), que a su vez se transforma a una estructura romboédrica (paratacamita) and / or ortorrómbica (atacamite) bajo casi todas las condiciones ambientales naturales [24,25]. Las últimas dos estructuras cristalinas son reportadas extensamente como productos de corrosión en procesos de hidrolisis. En la Figura 11 se muestra el espectro de DRX obtenido del polvo verde, este corresponde al compuesto Cu₂Cl(OH)₃ con estructura cristalina ortorrómbica (Atacamita: grupo espacial PMAN, a = 6.0479 Å, b = 9.1084 Å, c = 6.8602 Å) [24], reportada como la fase estable en condiciones atmosféricas.

En resumen, las especies identificadas en el plasma de CHCl₃ fueron C₂, C, Cl, C⁺, CH⁺, HCl⁺ and Cl⁺. Se calcularon por una ruta alterna al empleo a la sonda de Langmuir la $T_e = 5.74\pm0.57$ eV y $n_e = (4.48\pm0.45) \times 10^{10}$ cm⁻³. El depósito formado sobre los electrodos de Cu es inicialmente el CuCl, que reacciona para formar el Cu₂Cl en el interior de la cámara para posteriormente transformarse en el compuesto Cu₂Cl(OH)₃ en condiciones atmosféricas. Esta reacción en la cámara, hace probable reducir las emisiones de Cl, convirtiendo el Cl en un compuesto más estable e inocuo.

Plasma atmosférico

Recientemente en el Laboratorio de Espectroscopia se está implementando la generación y aplicación de plasma a presión atmosférica. Uno de los nichos abiertos a esta tecnología es el desarrollo de sistemas con mejor adhesión en las industrias automotriz y aeronáutica. Lograr una mejor adhesión al emplear adhesivos se lograr una mejor preparación superficial, básicamente una superficie limpia y libre de contaminantes que impidan una máxima adhesión entre la superficies y el adhesivo empleado. Esta preparación es todo un reto en condiciones de producción masiva.

Los avances en el tratamiento de superficies con plasma atmosférico ha mostrado la factibilidad de poner en marcha estos dispositivos a nivel industrial. Pero aún se requiere de estudios sistemáticos de nuevos polímeros y de los nuevos adhesivos desarrollados en condiciones reproducibles. En esta parte una de las técnicas sencillas y de gran utilidad es la medición del ángulo de mojado de las superficies tratadas con el plasma atmosférico, método empleado en las líneas de producción. En la Figura 12 se indica cómo se mide el ángulo de contacto de cualquier superficie, sin tratamiento o tratada con plasma. La variación de las superficies tratadas con plasma, en este caso la superficie tratada es un acero sin tratamiento y el segundo es el mismo acero después de ser tratado por tres horas en un plasma de una mezcla de H_2/N_2 . El cambio va de los 95° in tratamiento a los 70° una vez tratada en plasma.



Figura 12. Medición del ángulo de contacto de un material.

El tratamiento con plasma atmosférico puede ser sintonizado para acondicionar cualquier superficie de prácticamente cualquier material. Las áreas de aplicación de esta tecnología es amplia y muy versátil entre ellos se pueden mencionar los dispositivos ópticos, empaque de componentes electrónicos, ensamble, automotriz, medica, etc. Los procesos que pueden ser implementados son: limpieza de superficies, adhesión, recubrimientos, sellado, pintado, imprimidos, etc.

Endurecimiento superficial.

En el área de modificación de las propiedades fisicoquímicas de los materiales, tiene aplicación el tratamiento con plasma. En aceros se puede difundir nitrógeno a partir de un plasma de una mezcla gaseosa de H_2 y N_2 , procedimiento conocido como nitruración iónica. Aun cuando existen otras técnicas a nivel industrial basados en procesos difusivos (cementación, líquida y gaseosa). El espesor de la capa formada es función de la temperatura, el tiempo y el coeficiente de difusión del material tratado. Las especies químicas más empleadas son el C y el N. El incremento en las propiedades mecánicas se debe a la formación de carburos y nitruros sobre la superficie durante el proceso difusivo.

El objetivo del tratamiento con plasma de N_2 de los aceros es el de lograr una capa superficial de mayor dureza que el material sin tratamiento, obteniendo una mayor dureza, resistencia al desgaste y a la corrosión.

La caracterización de los resultados obtenidos son básicamente de tipo mecánico. En la Figura 13 se muestra el espesor de la capa nitrurada como función del tiempo de

tratamiento con plasma de un acero experimental de alta resistencia. Con este tratamiento se busca mejorar su dureza superficial para impartir, resistencia al desgaste y probablemente resistencia a la corrosión.



Figura 13. Espesor estimado de la capa nitrurada como función del tiempo de tratamiento.

El perfil de espesores fue estimado mediante el estudio de las superficies de fractura a temperatura de nitrógeno líquido. Este procedimiento se basa en la diferente respuesta del acero en la matriz y de la capa endurecida en la superficie. En la figura 14 se muestra la superficie de fractura del acero, donde se aprecia la capa endurecida con la superficie de fractura completamente frágil.

En la Figura 15 se muestra el perfil de microdureza obtenido de la capa endurecida por 18 horas de tratamiento con plasma. Debe notarse que existe una discrepancia entre el espesor estimado por fractografia (34 ± 9 µm) y el estimado por el perfil de microdureza Vickers (~100 µm).



Figura 14. Superficie de la fractura a temperatura de nitrógeno líquido. La zona completamente plana corresponde con el espesor de la capa nitrurada.



Figura 15. Perfil de Microdureza de la muestra tr
tada por 18 horas, se estima un espesor de ${\sim}100~\mu m.$

La dureza superficial de un acero microaleado experimental de alta resistencia se incrementó con en el proceso de nitruración, obteniendo una dureza máxima en la muestra nitrurada por 12 horas. Se encontró una mínima diferencia entre las durezas obtenidas de las capas nitruradas, en las muestras de 12 y 18 horas. El espesor de la capa nitrurada es función del tiempo de nitruración.

Agradecimientos

Los autores agradecen el apoyo de: Héctor Hugo Hinojosa, Iván Puente, Itzel Chaparro, Gerardo Aramburo, Sergio García y Ciro Márquez.

Referencias

1. The mechanism of electrical discharges in gases of low pressure. M. J. Druyvesteyn and F. M. Penning. Reviews of Modern Physics, Vol. 12, No. 2, April 1940.

2. Degradation of the textile dye AB 52 in an aqueous solution by applying a plasma at atmospheric pressure. J. Vergara, C. Torres, E. Montiel, A. Gómez, P. G. Reyes and H. Mrtínez, IEEE Transactions on Plasma Science 45 (3), 479-484 (2017) ISSN: 0093-3813.

3. Effects of Na incorporation and plasma treatment on Bi_2S_3 ultra-thin layers. Moreno-Garcia H., Messina S., Calixto-Rodriguez M., Martínez H. Thin Solid films, 2016.

4. Characterization of Ethanol Plasma Glow Discharge, Decomposition in Several Species and Solid Film Formation. P. G. Reyes, A. Gómez, H. Martínez, O. Flores, C. Torres, and J. Vergara. IEEE Transactions on Plasma Science, Vol. 44, NO. 12, December 2016, 2995-3000.

5. Mechanical performance of thermally post-treated ion-nitrided steels, Rosales I., Martinez H., Guardian R. Applied Surface Science 371 (2016) 576–582.

6. Characterization of Ar-O₂ DC Discharge employing Langmuir probe in conjunction with photo-detachment. Yousif F. B.; Solid films B., Vasquez F., Martinez H., Rivera M., Rodriguez J. IEEE Transactions on Plasma Science 44 (2016) 1150-1154. ISSN: 0093-3813.

7. Low energy ionization and fragmentation cross sections for H+ impact on N2 and O2. López-Patíño, Fuentes B.E., Yousif F.B., Martínez. International Journal of Mass Spectrometry 405 (2016) 59-63. ISSN: 1387-3806.

8. Corrosion behavior of 18Cr-18Mn hot-forged and plasma-nitrided steel. Hernández R.R., Torres-Islas A., Serna S., Bedolla A., Molina A., Colín J., Martínez H. Journal of Advanced Electrochemistry 2 (4), 117–120 (2016).

9. The Hydrophilic to Superhydrophilic Change Induced by Polyhydroxybutyrate in Polyethylene glycol: Polyhydroxybutyrate Electrospun Samples by Plasma Treatment. Rojano-Molina Ma. G., Domínguez-Díaz M., Martínez-Valencia H., Escocia-García J. and Fabiola Balderas-Valadez R., MRS Advances 1 (29), 2125-2131 (2016).

10. Ionization and fragmentation of CH_4 by proton impact, Fuentes B.E., López-Patiño J., Yousif F.B., Martínez H. International Journal of Mass Spectrometry 411, 21-26 (2016) ISSN: 1387-3806.

11. Molecular Changes in Asphaltenes within H₂ Plasma. Juan C. Poveda, Daniel Molina, Horacio Martínez, Oswaldo Florez, and Bernardo Campillo. Energy Fuels 2014, 28, 735–744. dx.doi.org/10.1021/ef401773t.

11. Low pressure CH_2Cl_2 plasma discharge. Journal of Advances in Physics Vol.8, No.3, may 2015. J. C. Poveda, O. Flores, H. Martinez, B. Campillo and F. B. Yousif. Electrical and optical characterization of pulsed plasma of N_2 -H₂. Martínez H. and Yousif F. B. European Physics D46, 493 (2008).

12. Diagnostics of parameters by optical emission spectroscopy and Langmuir probe in a discharge of $Ar/N_2/CH_4$ ternary mixture. Guerrero A., Salazar-Flores L., Torres-Segundo C., Martínez H., Reyes P. G. and Castillo F., Journal of Physics: Conference Series 370, 012047-4, (2012).

13. DC discharge experiment in a Ar/N₂/CO₂ ternary mixture: a laboratory simulation of the Martian ionosphere's plasma environment, García-Cosió G., Martínez H., Calixto-Rodríguez M., Gomez A., Journal of Quantitative Spectroscopy and Radiative Transfer 112 (2011) 2787-2793.

14. Langmuir probe and optical emission spectroscopy studies of low-pressure gas mixture of CO_2 and N_2 , Méndez-Martínez E. F., Reyes P. G., Osorio-González D., Castillo F. and Martínez H., Plasma Science and Technology 12(3), 314-319 (2010).

15. CH₂Cl₂ thin film formation on low-pressure DC plasma discharge, H. Martinez, O. Flores, B. Campillo, A. Gomez, L. Salazar-Flores and J.C. Poveda, Radiation Effects and Defects in Solids, en prensa (2012).

16. Asphaltene erosion process in air plasma: emission spectroscopy and surface analysis for air-plasma Reactions, Martínez H., Flores O., Poveda J.C. and Campillo B., Plasma Science and Technology 14, 303-311 (2012).

17. Asphaltene surface erosion in air plasma, Villa M., Calixto-Rodríguez M., Martínez H., Poveda J.C., Reyes P.G., and Altuzar P., Plasma Science and Technology 12(1), 81-86 (2010).

18. Oxidation Performance of Mo₃Si with Al Additions, Rosales I., Martínez H., Bahena D., Ruiz J.A., Guardian R., Colin J., Corrosion Science 51, 534–538 (2009).

19. Absolute differential and total cross sections for neutral fragments from dissociative collisions of triatomic hydrogen like ions in He", Yousif F.B., Fuentes B.E., Martínez H., Journal of Physics B: Atomic, Molecular & Optical Physics. 43 (2010) 235206.

20. Double capture cross sections in p^+Ar collisions, Martinez H., Alarcon F. B. and Amaya-Tapia A., Physical Review A, 062715 (2008).

21. G. J. Meeusen, E. A. Ershov-Pavlov, R. F. G. Meulenbroeks, M. C. M. Van Sanden and D. C. Schram; J. App. Phys. 71 (1992) 4156-4162.

22. W. Levy and M. Goreaud; Bull. Soc. Chim. Fr. 8 (1969) 2623-2634.

23. NIST Atomic Spectra Database Lines. National Institute of Standards and Technology (NIST), USA.

24. JCPDS 25-269 and ICDD 2077-116.

25. S. Wolf and C. Felmann; J. Mater. Chem. 20 (2010) 7694-7699.

Agujeros negros, superradianza y extracción de energía

Juan Carlos Degollado

1 Introducción

La relatividad especial presentada por Eintein en 1905 dio origen a una revolución en nuestras ideas del espacio y tiempo. Al incorporar los efectos gravitacionales surgió uno de los pilares de la física moderna: La teoría de la relatividad general

Einstein postuló que la gravitación debía identificarse con la curvatura del espacio tiempo. En términos matemáticos esto quiere decir que la teoría de la gravedad debe ser una *teoría métrica* en la cual los efectos gravitacionales se manifiestan a través de una distorsión en la geometría del espacio tiempo. Así, toda la información del espacio tiempo, independientemente del observador, está contenida en la métrica y la descripción de la gravitación está vinculada a las propiedades del espacio tiempo mediante la métrica.

La métrica del espacio tiempo describe completamente el campo gravitatorio por medio de sus geodésicas o curvas de longitud extrema. En relatividad general, la presencia de un campo gravitatorio se refleja en el hecho de que las geodésicas de la métrica que originalmente eran paralelas, no se mantienen paralelas. En otros términos, el espacio tiempo es curvo y la gravedad no se ve como una fuerza externa, sino como una distorsión de la geometría. Una descripción matemática del espacio tiempo es como sigue: Sean x^{μ} las coordenadas de un evento en donde la componente x^0 representa el tiempo y las componentes x^1 , x^2 , x^3 , representan las coordenadas espaciales ¹.

Para calcular las distancias infinitesimales entre dos eventos x^{μ} y $x^{\mu} + \delta x^{\mu}$, se utiliza el elemento de línea ds^2 que es una generalización de como se miden las distancias utilizando la geometría de Euclides. El elemento de línea se relaciona con las componentes de la métrica con $g_{\mu\nu}$, como:

$$ds^{2} = g_{\mu\nu}dx^{\mu}dx^{\nu} , \qquad (\mu,\nu=0,1,2,3).$$
(1)

En la expresión anterior hemos utilizando la convención de suma de Einstein que dice que hay una suma implícita siempre que haya índices repetidos. Así $v^{\mu}v_{\mu} = v^{0}v_{0} + v^{1}v_{1} + v^{2}v_{2} + v^{3}v_{3}$ y los índices tienen valores de 0 a 3.

Para dar una descripción completa del espacio tiempo hay que especificar qué geometría o métrica esta asociada a una determinada configuración de la materia. Einstein lo hizo al postular un conjunto de ecuaciones que dicen de forma general que:

Curvatura del espacio tiempo = densidad energética de materia.

De esta forma, tenemos una teoría de la gravitación en la cual los efectos gravitacionales se expresan plenamente desde el punto de vista de la estructura del espacio tiempo y a la estructura del espacio tiempo se la relaciona con la distribución de materia mediante la ecuaciones de Einstein.

Una de las predicciones mas interesantes de la teoría de la relatividad general es la existencia de los agujeros negros que describiremos brevemente a continuación.

2 Agujeros negros

Los agujeros negros son los objetos macroscópicos mas simples en la naturaleza. Un agujero negro, por definición, es una región del espacio tiempo en la cual la gravedad es tan intensa que nada, ni siquiera la luz, puede escapar de su atracción.

Los agujeros negros nacen de la muerte de una estrella –aunque no todas las estrellas al terminar su vida formarán un agujero negro. El destino de la estrella dependerá de su masa. Si la masa de la estrella no excede unas seis veces la masa del Sol, la estrella expulsará sólo una fracción de su masa y se convertirá en una enana blanca. Si la masa es mayor, la estrella expulsa gran parte de su masa

¹En todo el trabajo utilizaremos unidades geométricas en las cuales la velocidad de la luz c = 1 y la constante de gravitación universal G = 1. Como consecuencia de utilizar estas unidades, el tiempo, la distancia y la masa tienen unidades de longitud.

quedando como remanente su núcleo. Ese núcleo, dependiendo de su masa (lo que quedó de la estrella) se transformará en una estrella de neutrones o en un agujero negro. El proceso de la formación de los agujeros negros utilizando la teoría de la relatividad general fue descrito por primera vez hace casi 80 años por Oppenheimer y Snyder [6]. Ellos mostraron que los agujeros negros surgen como una consecuencia natural de la relatividad general y brindan una herramienta muy poderosa para examinar las propiedades macroscópicas y microscópicas de nuestro universo.

La propiedad que define a un agujero negro es el horizonte de eventos, la frontera que separa la parte del universo visible y la región a partir de la cual nada puede escapar. En este sentido, los agujeros negros son absorbedores perfectos.

Cuando un agujero negro se forma por el colapso de una estrella, parte del momento angular de la estrella lo conserva el agujero negro, dando lugar a un agujero negro con rotación. La rotación introduce otro ingrediente importante en el espacio tiempo alrededor de un agujero negro, una región de energía negativa, llamada ergo-región. La ergo-región esta delimitada por el límite estático o ergósfera y es una región en donde todos los objetos, incluso el propio espacio tiempo, son obligados a rotar en dirección del agujero negro. Análisis detallados de la ergoregión han mostrado que en ella pueden existir partículas con energía negativa. Las partículas con energía negativa junto con un horizonte da lugar a un fenómeno muy interesante en la física de agujeros negros. Según explicaremos más adelante, podemos utilizar estas partículas para extraer energía del agujero negro.

3 Extracción energética y superradianza

El espacio tiempo en las cercanías de un agujero negro tiene una influencia muy grande en el movimiento de los cuerpos cercanos. Los cuerpos que se acercan lo suficiente al agujero para atravesar la ergósfera son forzados a girar en la misma dirección que el agujero. En la ergósfera, aún cuando es posible escapar del intenso campo gravitatorio del agujero negro es imposible permanecer inmóvil.

En 1969 Roger Penrose sugirió un método para extraer energía de un agujero negro en la ergoregión [7]. El experimento (pensado) que planteó, es como sigue: Consideremos un agujero negro de masa M. Lejos del agujero negro, existe una partícula de energía E_0 . Como la partícula está fuera de la ergósfera su energía es positiva. Hágase que la partícula caiga hacia la ergósfera y dispóngase de un mecanismo que cuando la partícula llegue a la ergósfera se separe en dos fragmentos de manera que el primero tenga energía negativa E_1 y caiga hacia el agujero, mientras que el segundo fragmento escape con energía E_2 lejos del agujero. Mediante la conservación de energía tendremos:

$$E_0 = E_1 + E_2 , (2)$$

puesto que E_1 es negativa, la energía del fragmento que escapa E_2 , es mayor que la energía E_0 de la partícula incidente. Además, cuando el primer fragmento entra al agujero, reduce la masa energía de éste, por tanto al término del proceso, nos queda un agujero negro con masa $M - |E_1|$. En otras palabras, hemos extraído masa-energía del agujero negro.

Resulta que una partícula de energía negativa en la ergósfera debe tener también un momento angular orbital opuesto al del agujero negro, de tal modo que en el proceso de Penrose, cuando la partícula de energía negativa penetra en el agujero negro, reduce también el momento angular del mismo. Si se fuese a repetir indefinidamente el proceso de Penrose, al final, el momento angular del agujero negro se reduciría a cero. En el momento en que esto suceda, ya no habrá ergósfera y no se podrá extraer más energía. Esto es, según la visión clásica, el fin del proceso de Penrose.

Si se envía una onda electromagnética hacia un agujero negro, parte de ella será absorbida y otra parte será dispersada. En circunstancias comunes, la onda dispersada tendrá una amplitud menor a la incidente. No obstante, si una onda con frecuencia y dependencia espacial seleccionadas adecuadamente es enviada hacia la ergósfera de un agujero negro, la parte de la onda que es es absorbida por el agujero tendrá energía negativa debido a la ergo-región. La onda dispersada tendrá mayor energía y por consiguiente mayor intensidad que la onda incidente. Este proceso fue propuesto por primera vez en 1971 por Y. Zeldovich y se conoce como dispersión superradiante [9, 10].

La ganancia de energía por una onda se obtiene a expensas de la energía del agujero negro y ocurre solamente si la onda tiene una frecuencia

$$\omega \le m \,\Omega_+ \,\,, \tag{3}$$

en donde Ω_+ es la velocidad angular del horizonte del agujero negro y m es el número azimutal.

La dispersión superradiante puede utilizarse para extraer energía de un agujero negro utilizando diferentes tipos de ondas. Quizá el método mas simple de hacerlo consiste en la dispersión de una onda bosónica en un agujero con rotación. Desafortunadamente los factores típicos de amplificación son muy pequeños y se necesitan muchas dispersiones para que la extracción sea significativa.

Otro método simple de extraer energía de un agujero negro en rotación vía superradianza es encerrar al agujero negro dentro de una cavidad completamente reflectante. Cualquier perturbación inicial sera sucesivamente amplificada cerca del agujero negro y reflejada por el espejo, creando un crecimiento en cascada y como consecuencia una inestabilidad. A este proceso se le conoce como black hole bomb y fue sugerido por primera vez por Press y Teukolsky en la década de los 70's [8, 1]. La idea de colocar una superficie reflectante en la vecindad de un agujero negro no resulta muy atractiva desde un punto de vista práctico, sin embargo, algunas veces en la naturaleza podemos encontrar un mecanismo reflectante. Por ejemplo, si uno considera un campo escalar con masa μ , que es dispersado por un agujero negro, la masa funciona efectivamente como un mecanismo reflectante [3, 4].

En este trabajo nos enfocaremos en la dispersión superradiante utilizando un campo escalar. Para un campo escalar, el proceso superradiante puebla todos los niveles que satisfacen la condición de superradianza, siendo los mas cercanos al agujero negro los que crecen mas rápido. Para bosones, el número de ocupación crece exponencialmente formando una nube alrededor del agujero negro y deja de crecer cuando el agujero negro deja de rotar puesto que la condición de superradianza deja de ser válida.

El tiempo necesario para que el número de partículas en un nivel se duplique es típicamente 10^7 veces el radio gravitacional (r = 2M) del agujero negro. Para un agujero negro de una masa solar, esto puede ser tan corto como 100 segundos.

La idea de agujeros negros superradiantes tomó fuerza recientemente porque para que exista superadianza en agujeros negros de una pocas masas solares se requiere la existencia de partículas mas allá del modelo estándar y por lo tanto un fenómeno astrofísico como el *black hole bomb* puede utilizarse para probar teorías de partículas fundamentales. Por otro lado, en la galaxia hay varios millones de agujeros negros cuya masa es comparable con la masa del sol y por lo tanto es posible obtener datos para hacer estadística.

Una vez que se forma el agujero negro, si esta girando lo suficientemente rápido y si un campo bosónico con una masa adecuada existe, entonces el fenómeno superradiante automáticamente comenzará a poblar los niveles con estas partículas. Resolver las ecuaciones de Einstein para describir como evoluciona la inestabilidad de un agujero negro negro rodeado por un campo bosónico es muy complicado. Se requiere resolver un conjunto de 10 ecuaciones diferenciales de segundo orden altamente acopladas para la métrica, mas una ecuación de segundo orden para el campo de materia. Por fortuna, la situación mejora de forma sustancial cuando hacemos la suposición de simetría esférica. En simetría esférica, las ecuaciones se hacen manejables y podemos analizar con detalle lo que sucede. Sin embargo, un agujero negro con simetría esférica no puede tener rotación y por lo tanto no se observa superradianza. Para conservar la simetría esférica y tener superradianza debemos recurrir a un tipo diferente de agujeros negros: Los agujeros negros cargados. En la siguiente sección describiremos su geometría y algunos aspectos generales.

4 Agujeros negros cargados

Si una partícula cargada eléctricamente cae en un agujero negro estacionario en simetría esférica, este adquiere carga. Para describir tal agujero negro, es necesario encontrar una solución a las ecuaciones de Einstein acopladas a las ecuaciones de Maxwell y tomar en cuenta el tensor de energía momento del campo electromagnético. Tal solución existe y se conoce como el espacio tiempo de Reissner Nordstrom en honor a quienes la encontraron. En esta solución, el espacio tiempo está caracterizado por dos parámetros: La masa M y la carga Q del agujero negro.

La métrica que describe tal solución es

$$ds^{2} = f dt^{2} + f^{-1} dr^{2} + d\theta^{2} + \sin^{2} \theta d\varphi \, . \qquad f = 1 - \frac{2M}{r} + \frac{Q^{2}}{r^{2}} \, . \tag{4}$$

El radio del horizonte del agujero negro cargado es

$$r_{+} = M + \sqrt{M^2 - Q^2} \ . \tag{5}$$

Aunque los agujeros negros astrofísicos no tienen carga, el sistema es interesante desde el punto de vista conceptual porque la escala de tiempo para el desarrollo de la inestabilidad superradiante es muy corto comparado con la escala de tiempo del caso con rotación. Esta característica hace del espacio tiempo de Reissner Nordstrom un sistema muy adecuado para probar los estudios en el régimen no lineal y por lo tanto describir el desarrollo de la inestabilidad de un agujero negro en una cavidad.

Para entender como funciona el mecanismo de superradianza e inestabilidad en el espacio tiempo de un agujero negro cargado, consideremos primero el caso en el que la masa de la configuración de campo es pequeña comparada con la masa del agujero negro. De este modo la geometría del espacio tiempo no se ve afectada por su presencia pero si influye en su dinámica. Consideremos también que el campo bosónico es un campo escalar cargado que satisface la ecuación de Klein-Gordon.

$$(\nabla_{\nu} - iqA_{\nu})(\nabla^{\nu} - iqA^{\nu})\Phi - \mu^2\Phi = 0 , \qquad (6)$$

en donde μ y q son la masa y la carga del campo escalar respectivamente, $A_{\nu} := (-Q/r, 0, 0, 0)$ es el potencial vector del campo electromagnético y ∇_{ν} es la derivada compatible con la métrica².

Podemos resolver la ecuación (6) mediante separación de variables de la forma

$$\Phi(t, r, \theta, \varphi) = \frac{\psi(r)}{r} Y_{lm}(\theta, \varphi) e^{-i\omega t} , \qquad (7)$$

en donde Y_{lm} son los armónicos esféricos, y ψ satisface la ecuación radial:

$$f\frac{d}{dr}\left(f\frac{d}{dr}\psi\right) + V_{\rm ef}(r)\psi = 0 , \qquad (8)$$

con el potencial efectivo:

$$V_{\rm ef}(r) = f\left(\mu^2 + \frac{l(l+1)}{r^2} + \frac{f'}{r}\right) - \left(\omega - \frac{qQ}{r}\right)^2 \ . \tag{9}$$

Si utilizamos la coordenada tortuga r_* , definida por la relación $dr_*/dr = 1/f$ en la ecuación (8) podemos obtener una ecuación tipo Schrödinger:

$$-\frac{d^2}{dr^2_*}\psi + V_{\rm ef}(r)\psi = 0.$$
 (10)

Tomando el límite asintótico en la ecuación (10) encontramos soluciones de la forma:

$$\psi(r) = \begin{cases} T e^{-i\xi r_*} & \text{si} & r \to r_+ \\ e^{-i\rho r_*} + R e^{i\rho r_*} & \text{si} & r \to r_\infty \end{cases},$$
(11)

en donde $\xi = \omega - qQ/r_+$ y $\rho = \sqrt{\omega^2 - \mu^2}$. T y R están relacionados con las partes entrantes y salientes de la onda. Imponiendo la conservación del flujo en el horizonte se obtiene:

$$|R|^{2} + \frac{1}{\rho} \left(\omega - \frac{qQ}{r_{+}} \right) |T|^{2} = 1 , \qquad (12)$$

 $|{\cal T}|^2$ y $|{\cal R}|^2$ son los coeficientes de reflexión y transmisión de la onda.

De la ecuación (12) notamos que si $\omega < \omega_c := |qQ|/r_+$ el coeficiente de reflexión debe ser mayor que el de transmisión y por lo tanto las ondas con frecuencia menor que ω_c serán superradiantes. La descripción anterior funciona muy bien para ondas con dependencia $e^{-i\omega t}$ (ondas planas) pero en la siguiente sección exploraremos casos más generales, permitiendo que el campo tenga una dependencia temporal arbitraria $\Phi(t, r, \theta, \varphi) = \Psi(t, r)Y(\theta, \varphi)$. Esto nos permitirá estudiar situaciones con una dependencia temporal más rica.

5 Soluciones dinámicas

Como mostramos en la sección anterior, un campo escalar cargado alrededor de un agujero negro cargado puede ser superradiante y extraer energía de un agujero negro si oscila con una frecuencia menor a una frecuencia crítica ω_c . El siguiente paso consiste en determinar que ocurre si encerramos al campo y al agujero para inducir una inestabilidad, es decir determinar cual es el desarrollo final del *black hole bomb* sin restringirnos a un modelo con una sola frecuencia.

Como primer paso, ilustraremos como es el mecanismo de crecimiento exponencial del campo en el tiempo. Esto lo hacemos resolviendo la ecuación (6) manteniendo el espacio tiempo fijo dado por la métrica (4).

²Esto quiere decir que $\nabla_{\nu}g^{\mu\nu} = 0.$

Para resolver la ecuación numéricamente, asumimos que el campo escalar tiene inicialmente una distribución gaussiana de la forma $\psi = \psi_0 \exp[-(r-r_o)^2/\sigma^2]$. Con este dato inicial, el campo escalar contiene inicialmente muchas frecuencias pero después de algún tiempo las que dominarán son las superradiantes. Esto es fácil de entender pues los modos superradiantes dominarán en amplitud a los que no lo son y la amplitud del campo escalar crecerá extrayendo energía del agujero negro.

Este proceso se ilustra en la figuras 1 y 2. En el panel superior de la figura 1 se muestra el comportamiento oscilatorio del campo y en el panel inferior se muestra su energía asociada definida como:

$$E(t) = \int_{r_{+}}^{\infty} \rho(t, r) r^2 \, dr \,, \tag{13}$$

en donde

$$\rho(t,r) = \frac{1}{2} \left[\frac{1}{\alpha^2} \left(|\partial_t \Psi|^2 - 2qA_t \operatorname{Im}(\Psi^* \partial_t \Psi) \right) + (1-\beta_r) |\partial_r \Psi|^2 + \left(\frac{q^2 A_t^2}{\alpha^2} + \frac{l(l+1)}{r^2} + \mu^2 \right) |\Psi|^2 \right].$$

es la densidad de energía y

$$\alpha = \left(1 + \frac{2M}{r} - \frac{Q^2}{r^2}\right)^{-1/2} , \quad \beta_r = \frac{2M}{r} - \frac{Q^2}{r^2} . \tag{14}$$

La línea discontinua representa una función exponencial y se ha colocado para mostrar que a tiempos tardíos la energía del campo crece exponencialmente.



Fig. 1: (Panel superior)Evolución temporal del campo escalar. (Panel inferior) Energía asociada al campo escalar. La escala logarítmica muestra que el crecimiento de la energía y por lo tanto del campo, es exponencial.

En la figura 2 se muestra el espectro de potencias obtenido mediante una transformada de Fourier para la amplitud del campo a diferentes tiempos. En las gráficas de la columna izquierda está graficada la amplitud del campo y en la columna derecha su espectro para $t = 10^3 M$ y $t = 10^5 M$. La línea vertical a trazos en el espectro, muestra la frecuencia crítica $\omega_c = qQ/r_+$. A tiempos relativamente cortos, las frecuencias dominantes no son las superradiantes, pero conforme avanza el tiempo, el comportamiento se invierte.

El siguiente paso para entender lo que ocurre *después* del crecimiento exponencial es resolver las ecuaciones de Einstein para la geometría, las ecuaciones de Maxwell para el campo electromagnético y la ecuación de Klein-Gordon para un campo cargado. Con la hipótesis de simetría esférica es posible hacerlo utilizando códigos computacionales que hemos desarrollado para tal efecto. Las ecuaciones de Einstein se pueden escribir como un conjunto de ecuaciones diferenciales parciales para las componentes de la métrica. Para hacerlo es necesario descomponer todos los elementos tensoriales en términos de sus proyecciones en superficies en tres dimensiones y luego escribir las ecuaciones de Einstein como un problema de valores iniciales. Dados los datos en una superficie inicial queremos determinar como



Fig. 2: (Columna izquierda) Amplitud del campo a diferentes tiempos.(Columna derecha) Espectro de potencias obtenido mediante la transformada de Fourier. Se observa que el campo oscila con diferentes frecuencias, pero a tiempos tardíos la frecuencia dominante es la superradiante.

evolucionan en el tiempo. Este formalismo es actualmente muy utilizado para modelar la dinámica de sistemas astrofísicos compactos y se le conoce como formalismo Baumgarte-Shapiro-Shibata-Nakamura. Para las soluciones encontradas en este trabajo utilizamos la siguiente descomposición de la métrica

$$ds^2 = -(\alpha^2 - \beta^r \beta_r)dt^2 + 2\beta_r dt dr + e^{4\chi}[adr^2 + br^2 d\Omega^2],$$
 (15)

en donde α es la función de lapso, β^r es la componente radial del vector de corrimiento y las funciones métricas χ , $a \neq b$ dependen de $t \neq r$. El campo eléctrico tiene solo una componente radial y el campo magnetico se anula. La simetría esférica implica que sólo debemos considerar las ecuaciones para el potencial eléctrico ⁽³⁾ ϕ . Con estas herramientas, hemos podido describir la dinámica de campos escalares en agujeros negros en diferentes contextos [2, 5].

Con la solución numérica del conjunto de ecuaciones, es posible seguir el desarrollo del sistema agujero negro cargado con campo escalar. En nuestros experimentos computacionales hemos observado que después del crecimiento exponencial del campo, el sistema alcanza un estado estacionario en el que el espacio tiempo no cambia y la amplitud del campo no crece más tal como lo ilustra la figura 3.

Mediante estas simulaciones, es posible mostrar que existe una solución a las ecuaciones de Einstein que representa a un agujero negro cargado rodeado de una nube de campo escalar cargado jy todo está en equilibrio!. Esta nueva configuración es interesante desde el punto de vista teórico porque muestra que existen soluciones a las ecuaciones de Einstein que representan agujeros negros que no están aislados, que están rodeados de materia. Este es un ejemplo de un agujero negro que coexiste con la materia a su alrededor, algo que no se había anticipado en el modelo de extracción energética de Penrose. A este tipo de configuraciones se les ha llamado *agujeros negros con pelo* y no era claro si podían formarse dinámicamente. Con el experimento descrito anteriormente mostramos que los agujers negros con pelo pueden obtenerse como la fase final del *black hole bomb*.

6 Conclusiones

Resolver las ecuaciones de Einstein numéricamente se ha convertido en una tarea común hoy en día. Varios grupos de investigación en el mundo poseen códigos capaces de modelar sistemas tan complicados como la colisión de dos agujeros negros y su respectva emisión de ondas gravitacionales. Se ha llegado a un punto de madurez de la relatividad numérica tal que podemos descubrir nueva física mediante experimentos computacionales. Quedan diversos escenarios por explorar, tanto en astrofísica como en



Fig. 3: En un principio, la amplitud del campo crece exponencialmente extrayendo energía del agujero negro y luego las oscilaciones tienen una amplitud constante, señal de que se ha alcanzado un estado estacionario.

física teórica. La solución a las ecuaciones de Einstein que hemos descrito, es un ejemplo de que la física de los agujeros negros es más rica de lo que se suponía mediante técnicas analíticas y conjeturas sobre las simetrías en las ecuaciones de Einstein. Todavía quedan muchos aspectos que explorar como la dinámica de partículas de prueba a su alrededor o si pueden producirse por algún otro mecanismo. Estos son aspectos que se desarrollarán en un futuro cercano. Actualmente los métodos numéricos y el uso de códigos computacionales aplicados a la relatividad general nos muestran que podemos aprender mucho más sobre la física de agujeros negros en situaciones dinámicas. Con el desarrollo y puesta en práctica de estas herramientas podremos describir casos mas generales y extraer propiedades tanto de los agujeros negros como de las partículas que constituyen el ambiente alrededor de ellos tal como la solución descrita en este trabajo.

7 Agradecimientos

Agradezco al Dr. José Recamier la invitación para impartir una plática en la XXIV Escuela de Verano en Física. Agradezco también el apoyo por parte de la DGAPA-UNAM proyecto No. IA103616.

Referencias

- V. Cardoso, O. J. Dias, J. P. S. Lemos, and S. Yoshida. The Black hole bomb and superradiant instabilities. *Phys. Rev.*, D70:044039, 2004.
- [2] J. C. Degollado and C. A. R. Herdeiro. Time evolution of superradiant instabilities for charged black holes in a cavity. *Phys. Rev.*, D89:063005, 2014.
- [3] S. L. Detweiler. Klein-Gordon equation and rotating black holes. *Phys. Rev.*, D22:2323–2326, 1980.
- [4] S. R. Dolan. Superradiant instabilities of rotating black holes in the time domain. 2012.
- [5] C. A. R. Herdeiro, J. C. Degollado, and H. F. Rúnarsson. Rapid growth of superradiant instabilities for charged black holes in a cavity. *Phys. Rev.*, D88:063003, 2013.
- [6] C. W. Misner, K. S. Thorne, and J. A. Wheeler. Gravitation. W. H. Freeman and Company, 1973.
- [7] R. Penrose. Gravitational collapse: The role of general relativity. Riv. Nuovo Cim., 1:252–276, 1969.
- [8] W. H. Press and S. A. Teukolsky. Floating Orbits, Superradiant Scattering and the Black-hole Bomb. Nature, 238:211–212, 1972.
- [9] Y. Zeldovich. . Pisma Zh. Eksp. Teor. Fiz., 14:270, 1971.
- [10] T. Zouros and D. Eardley. Instabilities of massive scalar perturbations of a rotating black hole. Annals Phys., 118:139–155, 1979.

DINÁMICA DE LAS FRONTERAS DE GRANO EN LA CRISTALIZACIÓN COLOIDAL

Agustín E. González

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

RESUMEN

Se estudia la dinámica de las fronteras de grano que aparecen en el proceso de cristalización coloidal, las que se forman cuando dos granos cristalinos chocan en su crecimiento. El movimiento colectivo de las fronteras de grano es debido al intercambio de partículas de un cristal a otro, lo que ocurre en forma aleatoria. Ya se había propuesto anteriormente una dinámica de tipo caminata aleatoria unidimensional perpendicular a las fronteras, lo cual se probó para tiempos cortos. Con simulaciones de Monte carlo hemos demostrado que, para tiempos largos, ocurre no sólo esta caminata aleatoria unidimensional sino también una rotación y una deformación de dichas fornteras. Finalmente, se discute el porqué de este comportamiento.

El problema de la cristalización juega un papel muy importante en infinidad de procesos que ocurren, no sólo espontáneamente en la naturaleza, sino también en el laboratorio así como en la industria. Los procesos de cristalización son muchos y muy variados y aparecen no únicamente en la formación de cristales compuestos por átomos y moléculas pequeñas, del orden de algunos angstroms, sino también en el caso en que las partículas cristalizantes son de tamaños mesoscópicos (de decenas a varios cientos de nanómetros) como en la cristalización coloidal, la formación de cálculos renales, el procesamiento de comestibles, la cristalización de proteínas, etc. En fechas recientes ha habido un interés creciente en el estudio de los cristales formados por partículas coloidales [1, 2]. Una de las razones estriba en la posibilidad de ajustar el potencial de interacción entre las partículas modificando químicamente sus superficies, lo que nos permitiría estudiar las diferentes estructuras obtenidas por autoensamble. Otra razón importante está en la acapacidad de los investigadores para estudiar el proceso de cristalización con difracción laser y videomicroscopía óptica, dado que su tamaño y distancia entre partículas es usualmente comparable a la longitud de onda de la luz visible y a que su dinámica es lo suficientemente lenta para seguir las trayectorias de las partículas individuales. No se queda atrás el estudio de los llamados cristales fotónicos, que son las estructuras coloidales cristalinas que permanecen después de haber desecado la suspensión coloidal cristalizada. Nuevamente, debido a que el tamaño de estas partículas coloidales y su separación de primeros vecinos en el cristal es del orden de la longitud de onda de la luz, tienen comportamientos peculiares tales que ciertos rangos de longitudes de onda de la luz no los pueden atravesar, existiendo así las llamadas brechas fotónicas. Esto es similar a lo que ocurre con cristales ordinarios formados por moléculas pequeñas, en los que existen brechas electrónicas de energía que están vedadas a esos electrones [3, 4, 5, 6, 7]. Por último, pero no por ello menos importante, es la posibilidad de aplicar los resultados de la cristalización coloidal al auto-ensamble de proteínas en su estado globular nativo, un tema al cual se le ha prestado una atención considerable en los años recientes [8], dada su



Figure 1: La barrera de energía libre de un núcleo de cristalización. En general, los núcleos da tamaño subcrítico $(r < r_c)$ se empequeñecen y desaparecen, mientras que los núcleos que alcanzan un tamaño poscrítico $(r > r_c)$ continúa en su crecimiento, decreciendo de esta forma su energía

aplicabilidad en la bioquímica y en la industria farmacéutica.

En particular, las monocapas de partículas coloidales atrapadas por tensión superficial en la superficie libre de un líquido han sido propuestas como análogos bidimensionales (2D) de sistemas atómicos en fase condensada en 2D. P. Pieranski fue el primero en observar la formación de un cristal coloidal en 2D, hecho de esferas de poliestireno confinadas en el pozo de energía superficial en la interfaz aire-agua [9]. Como en este caso sólo una parte de la superficie de las esferas coloidales está sumergida en agua, permitiendo la disociación de los grupos de ácido sulfónico en esa parte de la superficie, se atribuyó el ordenamiento a la repulsión de dipolos eléctricos perpendiculares a la interfaz aire-agua. Una condición necesaria por lo tanto es que el rango de las interacciones repulsivas sea mayor a la distancia entre partículas vecinas cercanas. Para sistemas diluidos, se requiere un potencial de interacción atractivo para ordenar al sistema en un cristal. G. Y. Onoda [10] fue el primero en tener éxito en encontrar las partículas adecuadas en este otro caso. Consistían también en partículas de poliestireno de diferentes tamaños, estabilizadas con el tensoactivo dodecil sulfato de sodio injertado en sus superficies, para evitar la agregación de las partículas. El potencial de interacción atractivo consistía de las fuerzas de van der Waals experimentadas por dos partículas cercanas más la esfera dura al contacto. En sus experimentos encontró ordenamiento de las partículas del sistema y nucleación cristalina.

En la teoría clásica de la nucleación [11], a los núcleos de cristalización (llamados simplemente *cristalitos*) se les supone de forma esférica (circular en nuestro caso bidimensional). El crecimiento de estas regiones en un cristal depende de dos factores: (i) una disminución en la energía del bulto que favorece el crecimiento, y (ii) un aumento en la energía de la frontera que lo desfavorece. Estos factores se ven reflejados en el cambio de energía libre


Figure 2: El aspecto de una frontera de grano a un tiempo muy posterior al del tiempo de nucleación y de crecimiento de los cristales individuales. Nótese la desalineación de los ejes de simetría de los dos cristales que forman la frontera de grano

para la formación de un cristalito circular (en 2D) de radio r:

$$\Delta G = 2\pi r \,\gamma - \pi r^2 \Delta \mu \,n,\tag{1}$$

en donde γ es la tensión de línea de la frontera de los cristalitos con el fluido, $\Delta \mu = \mu_{fluido} - \mu_{cristal}$ 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 diferencia de energías libres tiene la forma de una barrera la que está bosquejada en la Fig. 1, en donde el radio del cristalito crítico viene dado por

$$r_c = \gamma / n \Delta \mu \quad . \tag{2}$$

Una vez que los cristalitos sobrepasan el tamaño crítico continúan creciendo, atrapando más particles del fluido que los rodea y que se pegan a sus fronteras. En este sentido, las fronteras cristalinas no son estáticas sino altamente dinámicas, colectando y soltando párticulas de y hacia los alrededores. Para un sistema con número de partículas y volumen fijos (área fija en nuestro caso 2D), llega un momento en que los cristales se empiezan a tocar en su crecimiento y forman las llamadas fronteras de grano. Nótese que los dos cristales que "chocaron" no se pueden fundir en uno solo, debido a las diferentes direcciones de los ejes de simetría de ambos cristales. A medida que transcurre el tiempo, dichas fronteras de grano aumentan en amplitud y se hacen más grandes. En la Fig. 2 mostramos el aspecto de una frontera de grano en una cristalización colloidal en 2D a un tiempo muy posterior al de la nucleación y crecimiento de los cristales, en la que notamos las diferentes direcciones



Figure 3: El potencial de interacción V utilizado en las simulaciones. Dicho potencial está normalizado por $k_B T$ mientras que la distancia r entre centros de partículas está normalizada por el diámetro del disco duro

de los ejes de simetría de los cristales que "chocaron".

Se ha observado tanto experimentalmente como con simulaciones que las fronteras de grano no son estáticas sino que describen un movimiento aleatorio, esencialmente perpendicular a la frontera de grano, en la forma de una caminata aleatoria unidimensional. Hay que hacer notar que el movimiento de las fronteras ocurre debido al intercambio de partículas entre los dos cristales, las cuales cruzan dicha frontera en una forma aleatoria. Ese intercambio produce el movimiento conjunto de la frontera de grano. Sólo que esta caminata aleatoria ha sido observado en intervalos de tiempo muy cortos, tanto en los experimentos como en las simulaciones. Para tratar de describir este hecho, hemos efectuado simulaciones de Monte Carlo cinético pero ahora para tiempos muy largos. Aunque el Monte Carlo cinético es una forma de resolver la ecuación maestra y debe ser complementada con las probabilidades de transición aproximadas de un estado al siguiente, Kikuchi et al [12, 13, 14] han demostrado que usando una dinámica de Metropolis [15] pueden recuperar la ecuación de difusión de Fokker-Planck. La única restricción impuesta en su derivación es que el desplazamiento máximo de una partícula debe ser suficientemente pequeño comparado con el tamaño del sistema y con las irregularidades del potencial entre las partículas [14]. Por lo tanto, una simulación de Monte Carlo cinético con dinámica de Metropolis es equivalente a una simulación de dinámica Browniana, que es la que se supone que obedecen las partículas coloidales. El potencial de interacción utilizado es el que se muestra en la Fig. 3, cuya expresión matemática es como sigue: $V/k_BT = (75/r^3) \exp(-2.5 (r-1)) \cos(10 (r-1)^4)$, en donde r es la distancia entre las partículas normalizada por el diámetro del disco duro. Después de que el coseno llega a cero por segunda ocasión, alrededor de $r_0 \simeq 1.8285$, el potencial se hace cero. Dicho potencial se tabuló para acelerar los cálculos.

Consideramos $N_{tot} = 53715$ discos monodispersos distribuidos al azar en una caja cuadrada con condiciones de frontera periódicas, cuidando de no traslapar los discos duros, que ocupan una fracción de área de $\phi = 0.15$. El algoritmo es como sigue: (i) se escoge una partícula al azar y se mueve por una distancia igual a un décimo del diámetro del disco duro en una dirección al 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



Figure 4: En las figuras de (a) a (d) se nuestra la caja de la simulación total a los tiempos de de Monte carlo (a) t = 14001, (b) t = 35001, (c) t = 56001 y (d) t = 98001. Las líneas rectas etiquetadas con las letras de A a D definen la posición promedio de las fronteras de grano al tiempo de Monte carlo 14001 para las líneas A a C, mientras que la línea D se define para el tiempo de Monte carlo 35001. Una vez que dichas líneas estan definidas, permanecen fijas durante todo el tiempo de la simulación.

 $\Delta U > 0$ sólo se acepta el movimiento si $exp(-\Delta U/k_BT) > R$, en donde R es un número aleatorio uniformemente distribuido en [0,1). (v) Después de cada intento de movimiento, sea aceptado o no, 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 intentado moverse una vez *en promedio*. De ahora en adelante el tiempo de Monte Carlo tsignificará el tiempo en barridas de Monte Carlo.

En la Fig. 4 se ilustra nuestra descripción del fenómeno de la dinámica de las fronteras de grano. Hay que recalcar que el movimiento de las fronteras ocurre debido al intercambio de partículas entre los dos cristales, las cuales cruzan dicha frontera en una forma aleatoria. Ese intercambio produce el movimiento conjunto de la frontera de grano. Entonces, las fronteras se pueden considerar como interfaces fluctuantes. Como ya se mencionó, se ha encontrado con simulaciones en 3D [16] y en experimentos en 2D [17] el movimiento colectivo de dichas fronteras de grano, en la forma de una caminata aleatoria unidimensional.

Esas observaciones fueron hechas para tiempos muy cortos, debido primordialmente a que las simulaciones fueron hechas con Dinámica Molecular, la que no es factible de avanzar a tiempos largos. Nosotros con nuestras simulaciones de Monte Carlo cinético podemos avanzar a tiempos extremadamente largos. En la Fig. 4 se definieron las fronteras de grano A, B y C, al tiempo de Monte Carlo de t = 14001, descritas por medio de segmentos de rectas que pasaban precisamente por dichas fronteras al tiempo indicado. Esto lo puede comprobar el lector haciéndole un zoom a la Fig. 4 con ayuda del visor de su computadora. La frontera etiquetada con la letra D se definió al tiempo t = 35001, una vez que se formó dicha frontera. Como podemos observar a los diferentes tiempos, las fronteras efectúan no únicamente una difusión unidimensional como había sido propuesto [16, 17], en donde dicha frontera se movería esencialmente paralela a ella misma, sino que también efectúan un movimiento de rotación así como una deformación. Sólamente para el caso de la frontera A el movimiento es *aproximadamente* perpendicular a dicha frontera. Uno se pregunta el porqué de esta dinámica tan extraña. Esto no nos debe sorprender dado que estamos tratando el movimiento de las fronteras de grano proviniendo de un intercambio de partículas entre los dos granos, que cruzan dicha frontera. No podemos esperar que el cruzamiento neto de estas partículas en un extremo de la frontera sea igual al del otro extremo, o a cualquier punto a lo largo de la frontera, dado que estamos tratando con eventos aleatorios. Entonces, una nueva descripción del movimiento colectivo de las fronteras de grano se vuelve necesaria, con lo que arribamos a nuestra intención de mostrar que en el proceso de cristalización no se ha dicho la última palabra en muchos de los temas interesantes que aparecen en los diferentes subprocesos del proceso total [18]. Dada la importancia del proceso de cristalización ya mencionada en la introducción, notamos que es imperativo continuar estudiando la multitud de subprocesos que ocurren en el proceso combinado de la cristalización, para tener una descripción lo más cercana posible al proceso real. Esto ayudaría enormemente al emplear nuestras simulaciones en otros procesos de cristalización, como los que ocurren en las proteínas, las cuales tienen una aplicación muy importante en nuestros días.

Experimentalmente, las partículas coloidales han demostrado ser un modelo útil con el que podemos probar fenómenos atómicos [3, 4, 6]. Su tamaño grande (~ μm) y dinámica lenta (~ s) nos permite usar técnicas ópticas simples, (como videomicroscopía), para estudiar el fenómeno de cristalización. La contraparte simulacional del estudio coloidal es el método de la dinámica Browniana la que, como hemos visto, es equivalente a una simulación de Monte Carlo cinético con dinámica de Metropolis. Por lo tanto, en el futuro cercano, esperamos ver un interés creciente en el uso de partículas coloidales para probar fenómenos atómicos. No obstante, hay muchos retos quer superar en este estudio, que merecen el esfuerzo intelectual debido a su aplicación práctica como sistemas modelos.

El autor agradece al Comité de Supercómputo de la UNAM por los recursos computacionales asignados para este proyecto (SC16-1-IR-52) a través de DGTIC-UNAM.

References

- [1] Pieranski, P. Colloidal crystals. Contemp. Phys. 1983, 24, 25-73.
- [2] Gasser, U. Crystallization in three- and two-dimensional colloidal suspensions. J. Phys.: Condens. Matter 2009, 21, 203101:1-203101:24.

- [3] Schall, P.; Cohen, I.; Weitz, D.A.; Spaepen, F. Visualization of dislocation dynamics in colloidal crystals. *Science* 2004, 305, 1944-1948.
- [4] Schall, P.; Cohen, I.; Weitz, D.A.; Spaepen, F. Visualizing dislocation nucleation by indenting colloidal crystals. *Nature* 2006, 440, 319-323.
- [5] Suresh, S. Colloid model for atoms *Nature Mater.* **2006**, *5*, 253-254.
- [6] Schall, P. Laser diffraction microscopy Rep. Prog. Phys. 2009, 72, 076601:1-076601:34.
- [7] Wang, Z.; Wang, F.; Peng, Y.; Zheng, Z.; Han, Y. Imaging the homogeneous nucleation during the melting of superheated colloidal crystals. *Science* **2012**, *338*, 87-90.
- [8] Véase por ejemplo Crystallization of nucleic acids and proteins: A practical approach, 2nd Edition; Ducruix, A., Giegé, R., Eds.; Oxford University Press: Oxford, U. K., 2000.
- [9] P. Pieranski, Phys. Rev. Lett. **1980**, *45*, 569.
- [10] G. Y. Onoda, Phys. Rev. Lett. **1985**, 55, 226.
- [11] Kelton, K.F. Crystal nucleation in liquids and glasses. In Solid State Physics, Vol. 45; Ehrenreich, H., Turnbull, D., Eds.; Academic Press: San Diego, U.S.A., 1991; pp 75-177
- [12] Kikuchi, K.; Yoshida, M.; Maekawa, T.; Watanabe, H. Metropolis Monte Carlo method as a numerical technique to solve the Fokker-Planck equation. *Chem. Phys. Lett.* **1991**, *185*, 335-338.
- [13] Kikuchi, K.; Yoshida, M.; Maekawa, T.; Watanabe, H. Metropolis Monte Carlo method for Brownian dynamics simulation generalized to include hydrodynamics interactions. *Chem. Phys. Lett.* **1992**, *196*, 57-61.
- [14] Yoshida, M.; Kikuchi, K. Metropolis Monte Carlo Brownian dynamics simulation of the ion atmosphere polarization around a rodlike polyion. J. Phys. Chem. 1994, 98, 10303.
- [15] Metropolis, N; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. J. Chem. Phys. 1953, 21, 1087-1092.
- [16] Trautt, Z. T.; Upmanyu, M; Karma, A. Interface mobility from interface random walk. Science 2006, 314, 632.
- [17] Skinner, T.O.E.; Aarts, D.G.A.L.; Dullens, R.P.A. Grain-boundary fluctuations in two-dimensional colloidal crystals. *Phys. Rev. Lett.* **2010**, *105*, 168301.
- [18] González, A. E. Colloidal Crystallization in 2D for Short-Ranged Attractions: A Descriptive Overview. Crystals 2016, 6, 46.

Formación de estructura con un campo escalar.

Juan Carlos Hidalgo^{1,*}

¹Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, 62210, Mexico (Dated: May 28, 2017)

En esta ponencia se muestra el modelo de colapso esférico, con énfasis en la caracterización de inhomogeneidades colapsantes. Se muestra que en un ambiente de fluido perfecto, donde se manifiesta la bien conocida inestabilidad de Jeans, se puede inferir un valor crítico para el colapso gravitacional en términos del contraste de densidad. Este mismo criterio se extrapola a un universo dominado por campo escalar. SE muestra que este criterio predice aceptablemente el corte en el espectro de potencias del modelo de campo escalar como materia oscura

I. EL UNIVERSO EN EXPANSIÓN

Hasta los a nos 1990s la cosmología fue una ciencia más especulativa que observacional. Los principios físicos en que se sustentan las ecuaciones que describen al universo como un todo tuvieron su origen en un debate filosófico más que una motivación descriptiva de las observaciones astronómicas. De hecho, la idea de que nuestro Universo se encuentre en un estado de expansión fue primeramente una elucubración literaria atribuida a Edgar Alan Poe [1], y el hecho de que el Universo haya tenido un origen temporal, el Big Bang, no fue confirmado sino hasta la observación de cuasares en ambientes exclusivos de alto corrimiento al rojo, observación reportada por primera vez en 1961 [2].

El principio cosmológico argumenta que como observadores no ocupamos un lugar especial en el Universo. De hecho, ningún observado ocupa un lugar especial en cuanto a que todos observan las mismas propiedades del universo a gran escala. Esta premisa extiende a nivel cosmológico el principio Copernicano heliocéntrico y matemáticamente consiste en imponer la característica de isotropía en el espaciotiempo que describe al Universo. A gran escala, si medimos la velocidad relativa de una galaxia $\mathbf{v}(r)$, ésta puede descomponerse en una componente en recesión radial y una componente en rotación:

$$\mathbf{v} = \boldsymbol{\Theta} \cdot \mathbf{r} + \mathbf{W} \times \mathbf{r}. \tag{1}$$

en donde Θ y **W** son tensores de deformación y rotación. En un universo isotrópico las propiedades de objetos son independientes de la dirección, de modo que la rotación es nula (**W** = 0), además de que en el tensor de deformación sólo puede tener cabida la expansión isotrópica, de modo que los valores propios de la expansión Θ deben ser iguales y la velocidad resulta proporcional a la distancia, como una ley de Hubble!

La posibilidad de un universo en expansión isotrópica, que da lugar a un universo homogéneo, fue descrita como una solución a las ecuaciones de Einstein por Georges Lemaitre en 1922. Friedmann fue el primero en demostrar que un universo homogéneo e isotrópico debe estar descrito por la métrica llamada de Friedmann-Lemaitre-Robertson-Walker (FLRW):

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

 $[\]ast$ Corresponding author: hidalgo@fis.unam.mx

donde a(t) es el factor de escala que modula el radio físico dR = a(t)dr y donde la curvatura espacial puede ser esférica (k > 0), plana (k = 0) o hiperbólica (k < 0). Una consecuencia inmediata de esta descripción del universo es la ley de Hubble: La velocidad de expansión de los distintos observadores es pues proporcional a la coordenada radial

$$\mathbf{v} = \mathbf{\Theta} \cdot \mathbf{r} = H\mathbf{r}.\tag{3}$$

Esta ley fue confirmada observacionalmente por Edwin Hubble en 1929. El factor de Hubble $H = d \ln(a)/dt$ es el único valor propio del tensor de deformación en Eq. (1) y dicha ley continúa siendo empleada para describir la expansión de galaxias en nuesto Universo.

El parámetro de Hubble sin embargo, no es en general constante en el tiempo. La evolución de la métrica está dictada por las ecuaciones de Einstein: Seis ecuaciones de campo que se reducen a dos ecuaciones en el caso de un universo isotrópico y homogéneo. De ellas, la primera se conoce como ecuación de Friedmann

$$H^{2} + \frac{kc^{2}}{a(t)^{2}} = \frac{8\pi G}{3}\bar{\rho} + \frac{1}{3}\Lambda.$$
(4)

Aquí el signo de la curvatura k dicta la evolución y el destino final del factor de escala a(t) en el espaciotiempo FLRW incluso para condiciones iniciales de $a(t_i)$ y $da/dt(t_i)$ comunes. En un universo con geometría esférica (k > 0) la curvatura cierra al universo que recolapsa en un tiempo finito. En el universo plano (k = 0)una expansión inicial garantiza que el espaciotiempo se expande eternamente alcanzando un factor de escala $a \to \infty$ sólo al tiempo infinito. Finalmente, el universo con geometría hiperbólica (k < 0) presenta un factor de escala que diverge a tiempos finitos. Estos tres casos se ilustran en la figura 1.

Una forma más compacta de la métrica FLRW es

$$ds^{2} = -dt^{2} + a(t)^{2} \left[d\chi^{2} + f_{k}(\chi)^{2} \left(d\theta^{2} + \sin(\theta)^{2} d\phi^{2} \right) \right]$$
(5)

Aquí se toma al radio comovil χ como una coordenada que define al radio de área dependiendo del signo de la curvatura, es decir:

$$f_k(\chi) = \begin{cases} \sin \chi & k = +1 \\ \chi & k = 0 \\ \sinh \chi & k = -1 \end{cases}$$

Dos casos son de interés particular para esta ponencia. El primero es el caso del Universo plano, donde las ecuaciones de Einstein y la Ec. de continuidad nos indican que un fluido barotrópico de presión $p = \omega \rho$ tendrá una evolución cosmológica

$$\bar{\rho}(t) = \left(\frac{a}{a_0}\right)^{-3(1+\omega)} \qquad \text{con} \quad a(t) = \left(\frac{t}{t_0}\right)^{2/3(1+\omega)} \tag{6}$$

En el caso del universo con curvatura positiva, el factor y el tiempo cósmico evolucionan de manera paramétrica con el tiempo conforme como parámetro. Si consideramos un universo con materia sin presión (polvo), la evolución es:

$$a(\tau) = \frac{a_{max}}{2} (1 - \cos \tau), \qquad t(\tau) = \frac{1}{\pi} t_{max} (\tau - \sin \tau).$$
(7)



FIG. 1. FIG. 1: Evolución del factor de escala en el tiempo cósmico (Credito: University of Northern Arizona)

Estos valores garantizan una etapa de expansión hasta un radio máximo a_{max} (en $\tau = \pi/2$) y un posterior recolapso del factor de escala cuando el factor de escala es nulo en $\tau = \pi$. Las dos soluciones presentadas nos servirán a continuación para hacer una descripción mínima de una fluctuación de materia que colapsa.

II. COLAPSO ESFÉRICO

La descripción del proceso de formación de estructura cósmica (galaxias y sus conglomerados) comprende dos etapas de evolución de inhomogeneidades. La primera es la etapa perturbativa, las inhomogeneidades de materia (y las correspondientes fluctuaciones de la métrica) se caracterizan por cantidades perturbativas. Por ejemplo, la inhomogeneidad de densidad de materia es una concentración mayor a la densidad promedio homogénea $\bar{\rho}(t)$ y está descrita por la perturbación de densidad $\delta = \rho(r, t)/\bar{\rho}(t) - 1$. La ecuación que dicta la evolución de esta cantidad es

$$\ddot{\delta} + 2H\dot{\delta} - \frac{4}{3}\frac{\dot{\delta}^2}{1+\delta} = 4\pi G\bar{\rho}\delta(1+\delta),\tag{8}$$

que en el límite lineal (cuando $\delta \ll 1$) presenta la forma:

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\bar{\rho}\delta = 0. \tag{9}$$

La solución de esta última ecuación para un universo dominado por povlo es:

$$\delta \propto \frac{a}{a_0} \propto \left(\frac{t}{t_0}\right)^{2/3} \tag{10}$$

El régimen perturbativo, y en particular la aproximación lineal, son válidas para valores pequeños de δ . En nuestro universo, las observaciones de inhomogeneidades en la radiación cósmica de fondo indican que las fluctuaciones comenzaron su evolución desde su entrada al horizonte causal (horizonte de Hubble), donde presentan amplitudes primordiales de orden perturbativo¹:

$$\delta_{\rm prim}^2 \approx 10^{-10} \left(\frac{k}{aH}\right)^4 \left(\frac{k}{k_\star}\right)^{(n_s-1)}.$$
(11)

La ecuación lineal para δ en la Ec. (8) muestra que la evolución es independiente de la escala (no presenta gradientes espaciales). La solución en el universo dominado por polvo resulta en la misma dependencia temporal que el factor de escala:

$$\delta = C(\mathbf{x})D(t) = C(\mathbf{x})(t/t_0)^{2/3}.$$
(12)

En el régimen no-lineal, las inhomogeneidades ya no pueden ser descritas por perturbaciones. Esto sucede, para escalas de interés astrofísico, a edades del universo anteriores a los cuasares más lejanos (redshift cósmico z = 5), es decir que todas las observaciones de conglomerados galácticos serán en escalas donde las inhomogeneidades se encuentran en el régimen no-lineal. Se hace entonces necesario, para interpretar observaciones actuales, que las fluctuaciones sean descritas por simulaciones numéricas de N-cuerpos o por soluciones exactas no-lineales. La solución exacta de colapso gravitacional más sencilla es el modelo de colapso esférico tipo Top-Hat que se describe a continuación.

Supongamos una región geométricamente equivalente a una 3-esfera, con densidad $\rho_a = \rho_b + \delta \rho$. La inhomogeneidad de materia se modela por esta región embebida en un espacio-tiempo FLRW plano y con densidad equivalente a la densidad de "fondo" ρ_b . Veamos cómo la región interna queda entonces descrita por un modelo FLRW con curvatura positiva. Si imponemos, para un tiempo inicial, una continuidad suave del espaciotiempo, se requiere que en la frontera entre ambas regiones, el factor de escala y su derivada temporal (el factor de hubble) sean identificados. Entonces $H_a = H_b$ y

$$H_a^2 = H_b^2 = \frac{8\pi G}{3}\rho_b = \frac{8\pi G}{3}\left[\rho_b + \delta\rho - \delta\rho\right] = \frac{8\pi G}{3}\left[\rho_a - \delta\rho\right] = \frac{8\pi G}{3}\rho_a - \frac{kc^2}{a^2(t_i)}.$$
 (13)

En consecuencia el modelo de colapso esférico Top-Hat compensa la sobredensidad central con una curvatura positiva de la misma región (es importante notar que la región de sobredensidad tiene una ρ_a homogénea y una curvatura positiva constante) de modo que en la región sobredensa queda modelada por un universo cerrado con curvatura:

$$k = \frac{8\pi G}{3c^2} a^2(t_i)(\rho_a - \rho_b).$$
 (14)

Esto nos genera una relación especial para la perturbación de densidad. En la norma donde la expnasión de

¹Es lógico inferir que las inhomogeneidades debieron permanecer sin evolución ó "congeladas" fuera del horizonte causal y que por tanto las condiciones iniciales de evolución se definen al tiempo del cruce de horizonte donde el número de onda que caracteriza el tamaño de la inhomogeneidad cumple k = a(t)H(t).

ambas regiones es idéntica para todo tiempo (durante el régimen perturbativo) y tenemos que la el contraste de densidad δ_{UH} cumple:

$$\frac{8\pi G}{3}\delta\rho_{UH} = \left(\frac{c}{a(t)}\right)^2 = H_a^2\delta_{UH},\tag{15}$$

donde $_{UH}$ denota la norma de expansión uniforme. El radio de área en el universo cerrado R_a y el radio de Hubble $R_H = c/H$ entonces se relacionan a partir de la ecuación anterior. En virtud de la ecuación de Friedmann, se infiere que

$$\delta_{UH} = \left(\frac{R_H}{R_a}\right) \sin^2 \chi. \tag{16}$$

Finalmente, si identificamos estas dos escalas, es decir, evaluamos la perturbación de curvatura cuando cruza el horizonte cosmológico, tenemos que

$$\delta_{\rm UH}^H = \sin^2 \chi. \tag{17}$$

Esta relación muestra que la perturbación de densidad en la norma de expansión uniforme, y al momento del cruce de horizonte, está limitada a valores en el rango $(0,1)^2$. Más allá de esta restricción, cualquier inhomogeneidad en esta norma debe evolucionar hasta el colapso gravitacional.

Vale la pena estudiar esta evolución más a fondo ya que es uno de los pilares de la teoría de formación de estructura. Las configuraciones en el colapso esférico comienzan en etapas perturabativas, donde la solución en Eq. (6) puede expandirse para valores pequeños del parámetro τ :

$$\frac{a}{a_{\max}} \simeq \frac{\tau^2}{4} - \frac{\tau^4}{48}, \qquad \frac{t}{t_{\max}} \simeq \frac{1}{\pi} \left(\frac{\tau^3}{6} - \frac{\tau^5}{120}\right).$$
 (18)

Las anteriores expresiones se combinan para escribir, a tiempos tempranos de evolución (amplitudes pequeñas de δ_{UH}),

$$\frac{a}{a_{\max}} = \frac{1}{4} \left(6\pi \frac{t}{t_{max}} \right)^{2/3} \left[1 - \frac{1}{20} \left(6\pi \frac{t}{t_{max}} \right)^{2/3} \right].$$
 (19)

Esta solución consiste de un factor $t^{3/2}$ que coincide con la evolución de un universo homogéneo de polvo (cf. Eq. (6)), y un factor en paréntesis cuadrados menor a 1. Esto reduce la expansión del factor de escala como se espera en el universo cerrado ilustrado en la Figura 1. Más aún, notamos que la perturbación de densidad, entendida como el inverso de la perturbación de volumen, puede escribirse como:

$$1 + \delta = \frac{\rho_a}{\rho_b} = \frac{M/\frac{4}{3}\pi(ar_{com})^3}{M/\frac{4}{3}(a_b r_{com})^3} = \frac{a_b^3}{a^3},$$
(20)

²Un universo con radio comovil $\chi = \pi$ puede presentar a distancias finitas un radio de área $R_a = a(t)\sin(\chi)$ nulo. Estas configuraciones se cierran en símismas y generan las Estos valores son descartados en el rango de amplitudes al cruce del horizonte.



FIG. 2. FIG. 2: Evolución del contraste de densidad δ lineal (azúl) y no-lineal (verde) en el modelo Top-Hat para un universo dominado por polvo. CREDITO: Tesis de Licenciatura. Waldemar Ruiz, UNAM, 2017

de modo que, sustituyendo los valores de las Eqs. (6) y (18) se obtiene

$$\delta_{\rm lin} = \frac{3}{20} \left(6\pi \frac{t}{t_{max}} \right)^{2/3},\tag{21}$$

lo cual reproduce la evolución lineal de las perturbaciones (cf. Eq. (9)).

En un régimen no-lineal, la fluctuación crece sin límite hasta que el factor de escala es nulo $(a(\tau = \pi) \rightarrow 0)$, lo que anula al volumen de la región sobredensa y provoca que la densidad central diverja en un tiempo finito. En la práctica, el tiempo en que la sobredensidad diverge es parametrizado por la perturbación lineal. Es decir, la amplitud lineal de la perturbación (20) es un nuevo reloj con el que podemos marcar los tiempos de máxima expansión (cuando $a(t_{max}) = a_{max}$) o los tiempos de colapso $(a(t_{col}) = 0)$. De hecho, para un universo dominado por polvo se tienen amplitudes:

$$\delta_{lin}(t_{\rm max}) = 1.06, \qquad \delta_{lin}(t_{\rm col}) = 1.686.$$
 (22)

Estos comportamientos son il
ustrados en la Figura2

El valor crítico de colapso δ_{col} es utilizado en el formalismo Press-Schechter [3] para cuantificar el número

de halos que se foman o virializan desde un espectro de potencias evolucionado linealmente. Este formalismo es una aproximación analítica (sorprendentemente eficiente!) a la formación de estructura modelada con la evolución newtoniana de N-cuerpos. Si cambiamos el contenido de materia o la fracción de masa que colapsa en halos de materia, obtendremos valores distintos para δ_{col} y el número de halos formados en este nuevo modelo cosmológico. Esta práctica es usual para caracterizar la formación de estructura con diversos modelos de materia y energía oscuras.

III. FORMACIÓN DE AGUJEROS NEGROS PRIMORDIALES

Los agujeros negros primordiales son típicamente de masas menores a 1 M_{\odot} . Las concentraciones de materia requerida para su formación son del orden de la densidad nuclear o mayores, y tales densidades de energía son sólo encontradas en el universo temprano, en épocas en que la materia dominante corresponde a partículas relativistas con una ecuación de estado tipo radiación ($\omega = 1/3$).

En estos ambientes, se requieren sobredensidades bastante pronunciadas, en el límite del régimen perturbativo, para vencer las fuerzas de presión y colapsar una inhomogeneidad hasta la formacin de un agujero negro. El modelo Top-Hat de colapso esférico es de utilidad en el cálculo ya que proporciona un valor mínimo de δ requerido para su formación. Una primera aproximación, propuesta por Carr en 1975 [4] muestra las bases físicas de este proceso. La ecuación de evolución de perturbaciones de materia en el espacio de Fourier y en un ambiente dominado por radiación es

$$\ddot{\delta} + 2H\dot{\delta} - (4\pi G\bar{\rho} - k^2 c_s^2)\delta = 0, \qquad (23)$$

donde c_s es la velocidad del sonido en el medio relativista ($c_s^2 = 1/3$ para un fluido de radiación o gas de fotones). Es crucial notar la diferencia de evolución respecto a la ecuación equivalente en un ambiente sin presión, Ec. (8). La presión del fluido barotrpico da lugar a la inestabilidad de Jeans y a su escala asociada

$$k_{\rm J}^2 = \frac{4\pi G\bar{\rho}}{c_s^2}$$
 ó $R_{\rm J} = c_s \frac{\pi}{G\bar{\rho}} \approx \sqrt{\omega} a_{\rm max}.$ (24)

Siguiendo el criterio de inestabilidad de Jeans, se concluye que el radio de área de una sobredensidad colapsando a un agujero negro debe cumplir que

$$\sqrt{\omega}a_{\max} < \sin(\chi)a_{\max} \le a_{\max}.$$
(25)

Más aún, en términos de la perturbación de materia en la norma de expansión uniforme, la desigualdad anterior implica que

$$\omega < \delta^H_{UH} \le 1. \tag{26}$$

Esto determina un valor crítico para el colapso de perturbaciones que pueden formar agujeros negros primordiales en la época de radiación, $\delta_{\text{crit}} = \omega$. Un artículo reciente ha revisitado este problema imponiendo un valor umbral más adecuado para el caso cosmolgico [5]. La primer ventaja de este nuevo análisis es tomar sobredensidades colapsantes que constituyen configuraciones compensadas en densidad (de modo que la masa total contenida en una inhomogeneidad no es distinta a la suma de la densidad de fondo $\bar{\rho}$ sumada dentro del mismo volumen). La segunda ventaja es definir un radio de inestabilidad distinto al radio de Jeans, ya que este debe corregirse para sobredensidades que se encuentran en un fondo en expansión. El valor umbral resultante para confguraciones de materia es

$$\delta_{\rm crit} = \sin^2 \left(\frac{\pi \sqrt{\omega}}{1 + 3\omega} \right). \tag{27}$$

Es claro que para un fluido de radiación ($\omega = 1/3$) el valor umbral $\delta_{\text{crit}} \approx .41$ es mucho mayor a la amplitud de perturbaciones primordiale en Ec. (10), con lo cual no se espera la producción abundante de ANPs en la época de radiación a menos de que se trate de un modelo de inflación exótico.

IV. COLAPSO DE UN CAMPO ESCALAR

Un caso de importancia en la formación de estructura cósmica es el colapso de configuraciones cosmológicas de campo escalar. El campo escalar libre, o campo Klein-Gordon (K-G), es el tipo de materia más simple y que cumple con la invarianza de Lorentz. El campo de K-G en la etapa de oscilación al fondo del potencial (donde $V = m^2 \varphi^2/2$) ha sido de gran utilidad en cosmología para modelar dos escenarios distintos. El primero es el escenario de materia oscura ya que en promedio, un campo de K-G oscilando en el fondo de su potencial presenta una presión casi nula y puede presentar conglomerados como veremos a continuación. Por otro lado, el campo K-G puede representar las etapas finales de inflación, llamada de recalentamiento. Dicha etapa es necesaria para transformar toda la energía contenida en el campo inflacionario al resto de partículas del modelo estándar.

A continuación se presenta el criterio de inestabilidad de fluctuaciones de un campo K-G. En una sección complementaria se mostrará cómo al empatar la amplitud crítica de colapso del campo escalar con la amplitud de perturbaciones de materia primordiales, esto resulta en un corte en el espectro de potencias que reproduce aceptablemente el corte deducido con la evolución de perturbaciones, pero sin necesidad de resolver toda la evolución.

A. Valor crítico de colapso para K-G

La ecuación de evolución de un campo escalar con potencial sin autointeracción (campo de Klein-Gordon) evolucionando en una métrica FLRW es

$$\ddot{\varphi} + 3H\dot{\varphi} + m^2\varphi = 0. \tag{28}$$

La solución a esta ecuación es de la forma

$$\varphi(t) = \varphi_0 \left(\frac{a_0}{a}\right)^{3/2} \left[\sin(mt) + \mathcal{O}\left\{(H/m)^2\right\}\right],\tag{29}$$

donde el resto de los términos es subdominante cuando se estudia el régimen $m \gg H$, lo cual es válido en los casos que nos conciernen.

Es posible interpretar al campo escalar en el fondo como un fluido perfecto. Esto resulta de identificar términos del tensor de energía-momento en el sistema propio de referencia, donde $T^{\mu\nu} = \text{Diag}(\rho, p, p, p)$, de

modo que

$$\rho_{\varphi} = -T_0^0 = 1/2\dot{\varphi}^2 + 1/2m^2\varphi^2, \tag{30}$$

$$p_{\varphi} = 1/3T_j^j = 1/2\dot{\varphi}^2 - 1/2m^2\varphi^2 \tag{31}$$

Más aún, si se promedian sobre el periodo $\Delta t = 1/m$ se tiene que estas cantidades se comportan como función del factor de escala

$$\begin{aligned} \langle \rho \rangle &= \frac{1}{2} \langle \dot{\varphi}^2 \rangle + \frac{1}{2} m^2 \langle \varphi^2 \rangle \approx m^2 \langle \varphi^2 \rangle, \\ \Rightarrow \quad \langle \rho \rangle &= \frac{C^2}{2} m^2 a^{-3} + \mathcal{O} \left\{ (H/m)^2 \right\}. \end{aligned} \tag{32} \\ \langle p \rangle &= \frac{1}{2} \langle \dot{\varphi}^2 \rangle - \frac{1}{2} m^2 \langle \varphi^2 \rangle \approx \frac{9C^2 H^2}{16 a^3}, \end{aligned}$$

$$\Rightarrow \quad \langle p \rangle = \mathcal{O}\left\{a^{-6}\right\}. \tag{33}$$

Este comportamiento promedio se hace dominante cuando estudiamos la dinámica en tiempos cosmolgógicos. Es notable que la densidad de materia presenta el comportamiento de una componente de polvo mientras que la presión decae rápidamente de modo que puede ser ignorada cuando se toman tiempos mayores a 1/H.

Habiendo caracterizado el comportamiento del campo K-G en un fondo cosmológico da pie al estudio de las inhomogeneidades del mismo. Tal y como sucede en el caso de un fluido perfecto, las ecuaciones de evolución de perturbaciones muestran una escala de corte. Esto es evidente en la ecuación para la perturbación de la variable (modificada) de Mukhanov-Sasaki $\hat{\mu}$ [6, 7]. Esta variable se relaciona con la perturbación de curvatura en la norma comovil ζ , como

$$\hat{\mu} \equiv -2\sqrt{a^3\gamma}\zeta \,, \tag{34}$$

donde $\gamma = 1 - \frac{\ddot{a}}{aH^2}$. A su vez esta perturbación de curvatura se relaciona con la perturbación de densidad como

$$\delta_{UH} = \frac{2(1+\omega)}{5+3\omega} \left(\frac{k}{aH}\right)^2 \zeta,\tag{35}$$

de modo que siempre podemos escribir la variable de Mukhanov-Sasaki en términos del contraste de densidad. Más aún, las relaciones anteriores hacen evidente el hecho de que no hay una mezcla de modos perturbativos (modos-k) a nivel lineal. En consecuencia, la escala de estabilidad para una variable es la misma para todas las cantidades perturbativas.

La ecuación para $\hat{\mu}$ es [8, 9], hasta orden $\mathcal{O}\left\{(H/m)^2\right\}$

$$\left[\frac{k^2c^2}{a^2} + m^2 + 2\sqrt{6\kappa}m^2Ca^{-3/2}\sin(mt)\cos(mt)\right]\tilde{\mu}_s + \ddot{\tilde{\mu}}_s = 0.$$
(36)

Con ayuda de identidades trigonométricas y con un cambio de variable independiente a $z = mt + \pi/4$, llegamos a escribir lo anterior como una ecuación de Mathieu:

$$\frac{d^2 \tilde{\mu}_k}{dz^2} + [A_k - 2q\cos(2z)] \,\tilde{\mu}_k = 0,$$
(37)

donde

$$q = \frac{1}{2}\sqrt{6\kappa}\varphi_0 \left(\frac{a_0}{a}\right)^{3/2} = 3H/m + \mathcal{O}\left\{(H/m)^2\right\} \qquad \text{y} \qquad A_k = 1 + \frac{k^2 c^2}{m^2 a^2}.$$
 (38)

Estas últimas cantidades nos determinan la escala de inestabilidad. La banda de inestabilidad en la ecuación de Mathieu está en el rango:

$$1 - q < A_k < 1 + q, \leftrightarrow -q < \frac{k^2 c^2}{m^2 a^2} < q.$$
 (39)

De modo que en vista de la Ec. (37), se tiene la existencia de la escala de inestabilidad

$$k_{\rm inst} = a\sqrt{3mH}/c^2. \tag{40}$$

Para el caso del camplo K-G a dicha escala se le identifica con el radio de De Broglie del campo.

$$R_{\rm dB} = \frac{a}{k_{\rm inst}} = \frac{c^2}{\sqrt{3mH}}.$$
(41)

La naturaleza de la inestabilidad es entonces distinta al caso del fluido perfecto. Sin embargo, dado que la inestabilidad se presenta para todo un rango de valores arriba del radio de De Broglie, se puede definir un valor umbral para el colapso de perturbaciones, que se encuentra siguiendo los pasos de la sección anterior. La desigualdad

$$R_{\rm max}/a_{\rm max} > R_{\rm dB}/a_{\rm max} = \sqrt{\frac{H}{3m}} \equiv f_{\rm dB}, \tag{42}$$

implica la existencia de un valor umbral para el contraste de densidad al momento del cruce del horizonte δ_{UH}^{H} , es decir,

$$f_{\rm dB}^2 \equiv \delta_{\rm crit,\varphi} < \delta_H^{UH} < 1.$$
(43)

B. Corte en el espectro de SFDM

En el caso de un campo escalar como materia oscura (SFDM) el modelo ha probado ser útil en la formación de halos de materia oscura alrededor de las galaxias observadas. Los valores de la masa del campo SFDM son generalmente del orden $m \sim 10^{-20}$ eV. En el universo actual, donde $H_0 \approx 10^{-33}$ se puede verificar que al día de hoy, y desde tiempo en el cruce de horizonte las fluctuaciones primordiales cumplen que $m \gg H$. Esto da validez al análisis anterior y permite utilizar el resultado anterior, el cual muestra cual debe ser el valor umbral para la formación de estructura en un ambiente donde gobierna un campo escalar K-G.

Tomando la desigualdad en la Ec. (42), se puede evaluar si el espectro primordial de la Ec. (10) es mayor o menor a $\delta_{\operatorname{crit},\varphi}$ al momento del cruce del horizonte, de modo que un halo de materia oscura ser o no formado a partir de fluctuaciones del campo K-G. Al comparar Ecs. (10) y (42) se obtiene un valor máximo para el número de onda (un valor mínimo para el radio de colapso) que pueda llegar a formar un halo de materia oscura. Esto se puede tabular para distintos valores de la masa del campo K-G y en la Tabla I se muestra el valor crítico $k_{c,m}$ para la formación de estructura.

Finalmente, se puede comparar el valor numérico presentado en la tabla I con el valor del corte calculado a

Field Mass m	$H_{c,m}$	$z_{c,m}$	$Mpc \cdot k_{c,m}$
$10^{-20} \mathrm{eV}$	$4.32 \times 10^{-21} \text{ eV}$	$1.76 imes 10^7$	38.34
$10^{-22} \mathrm{eV}$	$4.32\times 10^{-23}~{\rm eV}$	$1.76 imes 10^6$	3.837
$10^{-24} \mathrm{eV}$	$4.32\times 10^{-25}~{\rm eV}$	1.75×10^5	0.3854
$10^{-25} \mathrm{eV}$	$4.32\times 10^{-26}~{\rm eV}$	$5.50 imes 10^4$	0.123
$10^{-26} \mathrm{eV}$	$4.32\times 10^{-27}~{\rm eV}$	1.69×10^4	0.040

TABLE I. Modos de Fourier y redshift de entrada al horizonte $(k_{c,m} = H(z)/1 + z \text{ para el corte en el espectro de potencias para la materia oscura en función de la masa de distintas realizaciones del modelo ASFDM.$



FIG. 3. Espectro de potencias para el modelo ΛSFDM para un rango de masas del camp Klein-Gordon. Las curvas de colores siguen la prescripción numérica reportada por Hu *et al.* en [?] (El espectro del modelo de concordancia ΛCDM semuestra con una línea negra). El corte predicho analíticamente, y cuyos valores numéricos se despliegan en la Tabla IV B, se presenta con líneas negras verticales.

partir de la evolución de las perturbaciones desde la entrada al horizonte [?] como se hacer en la Figura 3. Las coincidencias con los valores encontrados a partir de nuestra simple fórmula son alentadores y proveen un nuevo método para encontrar cortes en el espectro de otros modelos de materia oscura que presenten bandas de inestabilidad.

V. SÍNTESIS Y PERSPECTIVAS

En esta exposición se ha presentado un criterio para discriminar los valores del contraste de densidad que pudieran colapsar y formar estructuras. Este criterio fue derivado por Carr [4] en el contexto de un fluido perfecto y depurado por Harada *et al.* [5]. Aquí hemos presentado dicho criterio para el colapso gravitacional en el contexto de un campo escalar. Hemos aplicado tal criterio para determinar el corte en el espectro de potencias del campo escalar como materia oscura, con resultados coincidentes con aquellos calculados a partir de evoluciones numéricas de perturbaciones primordiales.

Es inmediato inferir que el corte del espectro de potencias también es predecible con este método para cualquier modelo de materia oscura que presente un régimen inestable marcado por un valor crítico de estabilidad. Por otro lado, este tipo de valores umbral para el colapso gravitacional son de utilidad para determinar la función de masa, que representa la abundancia de halos de una masa dada a un redshift dado.

Finalmente es importante mencionar que la formación de agujeros negros primordiales podría ser cuantificada dado este umbral de formación de estructura con campo escalar tipo Klein-Gordon. Las abundancias de estos objetos para modelos realistas de recalentamiento, y su consecuencia en la saturación de cotas observacionales se discuten en un artículo reciente [11]

- P. Molaro and A. Cappi, "Edgar Allan Poe: the first man to conceive a Newtonian evolving Universe," CULTURE and COSMOS Volume 16 no1 and 2 (2012), pg 225-239, Nicholas Campion and Ralf Sinclair eds [arXiv:1506.05218 [physics.hist-ph]].
- [2] Ryle, M., & Clarke, R. W. "An examination of the steady-state model in the light of some recent observations of radio sources" 1961, Mon. Not. R. A. S., 122, 349
- W. H. Press and P. Schechter, "Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation," Astrophys. J. 187 (1974) 425. doi:10.1086/152650
- [4] B. J. Carr, "The Primordial black hole mass spectrum," Astrophys. J. 201 (1975) 1. doi:10.1086/153853
- [5] T. Harada, C. M. Yoo and K. Kohri, "Threshold of primordial black hole formation," Phys. Rev. D 88 (2013) no.8, 084051 Erratum: [Phys. Rev. D 89 (2014) no.2, 029903] doi:10.1103/PhysRevD.88.084051, 10.1103/PhysRevD.89.029903 [arXiv:1309.4201 [astro-ph.CO]].
- [6] V. F. Mukhanov, Sov. Phys. JETP 67 (1988) 1297 [Zh. Eksp. Teor. Fiz. 94N7 (1988) 1].
- [7] M. Sasaki, Prog. Theor. Phys. 76 (1986) 1036. doi:10.1143/PTP.76.1036
- [8] L. Kofman, A. D. Linde and A. A. Starobinsky, Phys. Rev. D 56 (1997) 3258 doi:10.1103/PhysRevD.56.3258
 [hep-ph/9704452].
- [9] K. Jedamzik, M. Lemoine and J. Martin, JCAP 1009 (2010) 034 doi:10.1088/1475-7516/2010/09/034
 [arXiv:1002.3039 [astro-ph.CO]].
- [10] W. Hu, Astrophys. J. 506 (1998) 485 doi:10.1086/306274 [astro-ph/9801234].
- [11] J. C. Hidalgo, J. De-Santiago, G. German, N. Barbosa-Cendejas and W. Ruiz-Luna, "Collapse threshold for a cosmological Klein Gordon field," arXiv:1705.02308 [astro-ph.CO].

La relación entre el conocimiento fundamental, generado por la Ciencia y su impacto en la economía de un país

Autor: Dr. Antonio M. Juárez Reyes Instituto de Ciencias Físicas, UNAM

En la presentación que tuve el gusto de presentar, en la XXIV edición de la Escuela de Verano en Física, llevada a cabo en el Instituto de Ciencias Físicas, se presentaron ejemplos explícitos de cómo la Física básica puede tener impacto en problemas tan importantes como la salud y el cuidado de los recursos naturales. Concretamente se presentó el ejemplo, por un lado, de cómo las técnicas de espectroscopia de absorción, en particular en el rango del infrarrojo medio, puede usarse para realizar diagnósticos médicos, por medio del análisis del aliento exhalado. Por otra parte, presenté ejemplos de cómo es posible emplear microdescargas atmosféricas para determinar la presencia de metales pesados en el agua. Estos ejemplos son solamente un conjunto pequeño de las grandes oportunidades que hay de emplear la ciencia fundamental en aplicaciones directas. Por otra parte, en un círculo virtuoso, las aplicaciones y problemas prácticos pueden generar temas de investigación fundamental. Aunque en México este tipo de sinergia virtuosa todavía es tema de debate, en los países desarrollados es un tema ya madurado. A los interesados en explorar a uno de los más claros exponentes de que la el estudio de Ciencia Fundamental (con Mayúsculas) no se pelea con las aplicaciones, le invito a leer el clásico documento de Vannevar Bush (https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm) que ya en 1945 planteaba esta sinergia como algo fundamental para el desarrollo de los Estados Unidos. Este documento, ahora clásico dio como lugar, entre otras cosas, a la creación de la National Science Foundation.

Motivado por el esbozo de la fecunda interacción entre el estudio fundamental de la Física y de su potencial impacto directo en la solución de problemas inmediatos e importantes para nuestro país, quiero presentar aquí un escrito en el que documento esta interacción de manera más cuantitativa. Este texto, que presenta la definición de Economía del Conocimiento, y otras definiciones importantes, muestra de manera cuantitativa la importancia de integrar Ciencia y Aplicaciones de la Ciencia, como un binomio útil y productivo, en más de un sentido, para los países que logran armonizar ambos.

La prioridad fundamental para el gobierno de un país es asegurar un nivel de vida digno para sus habitantes. Para lograr este objetivo, una nación debe asegurar estructuras organizacionales, iniciativas de fomento y políticas de crecimiento eficientes, productivas y sustentables a largo plazo. Actualmente el conocimiento, entre otros muchos factores, es uno de los pilares más importantes para generar el bienestar de las personas. Desafortunadamente, los beneficios económicos derivados del desarrollo científico y tecnológico sólo se concentran en un puñado de países, los generadores de los productos de alto valor agregado. A estos países, que vinculan su desarrollo económico en base a su conocimiento se les conoce como Economías o Sociedades del Conocimiento.

La Economía del Conocimiento (EC) y su relación con el bienestar de un país.

En años recientes han ocurrido auténticas revoluciones en las áreas de la biotecnología, la física, la instrumentación y las tecnologías de información. Nunca antes en la historia de la humanidad se había tenido un acceso tan eficiente y masivo al conocimiento, ni éste se había generado tan rápidamente como actualmente ocurre.

Aunque se hace ciencia en todas partes del mundo, incluyendo nuestro país, sólo los países que han invertido esfuerzos explícitos en la educación de sus ciudadanos, en el fomento de la ciencia y, más importante aún, en la vinculación entre sus sectores productivos y científicos, son cuantitativamente mucho más ricos y competitivos que los que no lo han hecho. Esta relación virtuosa entre el uso del conocimiento y su vinculación explícita en la generación de rigueza y en mejorar la calidad de vida de sus habitantes puede cuantificarse con el concepto de economía del conocimiento (EC). En el año 1996, la Organización para la Cooperación y el Desarrollo Económicos (OCDE, por sus siglas en Inglés) acuñó el término "Economía del conocimiento" para distinguir con esta definición al conjunto de países industrializados que basan su desarrollo y crecimiento en el uso del conocimiento generado por sus élites científicas y tecnológicas, de aquellos que no lo hacen. Es importante aclarar en este punto que el concepto de Economía del Conocimiento transciende el simple uso o consumo masivo de instrumentos de alta tecnología o de las tecnologías de la información. Los fundamentos de la economía del conocimiento se basan en la creación, difusión, uso y apropiamiento del conocimiento en su concepto más amplio. Los países y sociedades que se

benefician de la economía del conocimiento tienen como activo más valioso el uso explícito, en sus negocios y sistemas de producción, de la ciencia y la tecnología que desarrollan. En concreto, estos países basan la satisfacción de sus necesidades de alimento, vestido, salud y entretenimiento, fundamentalmente por medio de productos generados por la acción inventiva, o por la vinculación del conocimiento científico y los procesos productivos. En contraste con esto, las economías tradicionales como la mexicana, basan su desarrollo en la explotación de la mano obra, la creación de clústeres enfocados a la manufactura o la maquila, en la producción de bienes primarios (como el petróleo) o bien en la especulación financiera. Es evidente que, para aquellos que deseamos que nuestro país y sus habitantes tengamos un mejor nivel de vida, México debe transitar de manera acelerada y activa del estatus actual a una economía del conocimiento.

Es importante señalar aquí, que la EC se basa en la existencia de 4 pilares fundamentales:

- a) Un sistema educativo robusto y bien estructurado.
- b) El desarrollo y uso de una infraestructura de acceso a la información y telecomunicaciones.
- c) Un sistema de innovación y vinculación ágil entre academia y empresas.
- d) Un marco institucional de gobierno que fomente activamente el emprendimiento de base científica y que gestione de manera eficiente y transparente incentivos económicos para la innovación.

El índice de la economía del conocimiento (IEC) se mide en una escala del 1

al 10, mientras más bajo es este número, menor la capacidad de basar la actividad económica en el conocimiento. Un aspecto muy interesante del IEC es el hecho de que existe una relación directa entre el valor de este índice y, por ejemplo, el valor del ingreso per cápita de los países. La **figura 1** indica de manera gráfica esta

correlación directa. Es interesante apreciar que el crecimiento del ingreso per cápita es casi exponencial como función del IEC. Esto significa que pequeños incrementos en el valor del índice, en el sentido creciente, implican crecimientos medibles y cuantificables en el ingreso per cápita de los habitantes.



Figura 1.- Correlación entre el índice del conocimiento y el nivel de ingreso per-cápita de los respectivos países (IEC Banco Mundial, 2012).

¿Cuál es el IEC en México? En la evaluación del IEC internacional del año 2012, México se encontraba en el lugar 72 en relación al resto de los países, con un valor de IEC de 5.07. Esto se correlaciona de manera directa con un ingreso per cápita de entre 5,000 y 10,000 dólares anuales. Como todo promedio, el IEC en México oculta disparidades muy grandes. Esencialmente, el centro y norte del país acusan valores de IEC del orden de 6, mientras que estados del sureste, como Guerrero, Chiapas y Oaxaca no superan un valor de 2.6. El estado de Morelos, con una densidad de científicos por habitante muy elevada, incluso equiparable a la de varios países desarrollados, sólo alcanza un valor de 3.8 en el IEC (Banco Mundial, IEC index, 2012). Este último dato es muy interesante porque, en Morelos, hay una gran cantidad de institutos de investigación del más alto nivel que, en conjunto proporcionan una de las densidades más altas de investigadores por habitante del país. A pesar de lo anterior, el IEC para el estado de Morelos es muy pobre. Esto indica que la densidad y calidad de investigadores por habitante es un factor necesario, pero no suficiente, para lograr valores del IEC altos. Para impactar de manera directa el valor del IEC de una región, y por ende el nivel de bienestar de la población, es esencial que esta comunidad vincule sus esfuerzos a problemas y retos locales. Es esencial que esta comunidad explícitamente agregue esfuerzos con la cadena productiva y económica. Lo anterior, aunque hay contadas y notables excepciones, no ocurre en Morelos ni en México de manera sistemática.

México y sus retos para acceder a la Economía del Conocimiento

México cuenta con las condiciones suficientes para convertirse en una economía basada en el conocimiento ya que, al menos en las estadísticas oficiales, cuenta con universidades de alto nivel, así como una infraestructura de telecomunicaciones moderna. Asimismo, México cuenta con innumerables oficinas de transferencia tecnológica, programas de fomento para la vinculación entre empresas y academia (como el programa FINNOVA del CONACyT, entre varios otros). Finalmente, en nuestro país existe un marco regulatorio cada vez más flexible en reglas para la innovación, como lo prueba la reciente modificación a la ley de Ciencia y Tecnología. Uno de los aspectos torales de esta reforma, entre otras cosas, es que de manera explícita faculta a los investigadores a fundar empresas de manera legal y transparente, sin incurrir en conflictos de interés. Sin embargo, los elevados y dolorosos índices de pobreza, insuficiencia alimentaria, dependencia tecnológica del exterior y el deficiente acceso a un nivel digno de bienestar para nuestra sociedad, son una realidad tangible y cotidiana para un número elevado de mexicanos. Dado que tenemos una buena cantidad de factores que permitan acceder a otro estatus como nación, es importante preguntarse ¿Qué falta? ¿Qué estamos haciendo mal? ¿Qué podemos hacer en particular como comunidad académica para contribuir a remediar, en la parte que nos corresponde, el muy bajo nivel en el índice IEC en México? El problema es evidentemente complejo y pasa por la revisión de una política de estado en Ciencia, Tecnologia y normatividades a nivel federal. Como esa esfera es muy compleja, es importante saber, desde la UNAM y en general desde las Instituciones de Educación superior, ¿qué hacer, qué mejorar? desde nuestra labor y quehacer dentro de instituciones dedicadas a la investigación y la enseñanza a nivel posgrado. Revisar en este artículo todos los aspectos susceptibles de mejora a mejorar, desde la trinchera académica, es ciertamente una tarea compleja. Sin embargo, es útil, como ejercicio que espero motive revisiones similares en el lector, revisar aspectos en los que es posible mejorar cosas localmente, en el ámbito en el cual podemos incidir. En particular, presentaré algunos puntos muy específicos y sugerencias de mejora dentro de la UNAM. Los puntos planteados ni son todos, ni las soluciones son las mejores, pero hay que empezar a proponer y hacer. Mi propósito es que en el lector se catalice, con los puntos que refiero líneas abajo la meditación del quehacer de la UNAM como generadora de riqueza para los Mexicanos. En un mundo ideal, además de meditar y proponer, sería extraordinariamente útil para el país que llevemos a cabo estas mejoras y que las comuniquemos con los responsables de implementarlas de manera práctica, incluyéndonos a nosotros mismos, como académicos.

1. Aspecto educativo. En México sólo cuatro de 100 estudiantes estudian maestría y una proporción aún menor un doctorado. Estar en un posgrado en la UNAM o en cualquier IES es un privilegio. En general la formación de los estudiantes en nuestra Universidad a nivel Posgrado es elevada desde el punto de vista académico. Esto se puede constatar con la participación de la UNAM en el Programa Nacional de Posgrados de Calidad (PNPC) del CONACyT. Sin embargo, ¿para qué y con qué capacidades se preparan estos estudiantes? En las áreas físicas y biológicas, por ejemplo, el modelo de educación y formación de los estudiantes de doctorado es para que, a su vez, ellos se vuelvan académicos y, entre otras cosas sean mentores de estudiantes de doctorado con su mismo perfil. Esto es, preparamos a la élite educativa de nuestro país para ser académicos como nosotros, publicar y ser citados. En general no existen, en nuestro modelo de formación de estudiantes, las herramientas para que ellos puedan aspirar a otros destinos fuera de la academia, que además no está generando suficientes empleos

para estos doctores. Una alternativa sería que parte de su formación les faculte a generar empresas de base tecnológica basadas en el alto conocimiento que adquieren. En este aspecto, tenemos gente educada al más alto nivel y es muy necesario ahora que adquieran experiencia y conocimientos en actividades propias del sector productivo.

¿Qué tenemos que mejorar aquí? Evidentemente, entre otras cosas, vincular la educación de nuestros estudiantes con el uso práctico e inmediato de su conocimiento en problemas de alta relevancia. Desde luego, generar ciencia básica debe ser una parte fundamental de sus objetivos. Sin embargo, deben contemplarse, también, formaciones orientadas al emprendimiento de base científica, una cultura sana de protección intelectual y una orientación parcial del quehacer científico a problemas urgentes del país en agua, sustentabilidad alimentaria y salud, por ejemplo.

2. Los sistemas de Innovación. La evidencia muestra que la UNAM ha generado muy pocas empresas de base tecnológica de alto ingreso y alto impacto. Nuestro nivel de patentamiento es muy bajo, del orden de decenas de patentes al año, contra el orden de cientos o miles en paises de alto IEC, a pesar de ser la Universidad Nacional. Finalmente, los ingresos por licenciamiento de patentes de base tecnológica son pobres. Nuestra Universidad cuenta con capacidad científica y tecnologica equiparable a varias universidades europeas y norteamericanas. Sin embargo, nuestra generación de productos de alto valor, vinculación con los problemas nacionales y generación de licenciamientos y empresas de base

tecnológica es pobre. Instituciones como la Universidad de Oxford, por medio de su agencia de innovacion, llamada ISIS, genera regalías por protección intelectual, empresas de base tecnologica y "spin offs" del orden de mil millones de pesos anuales. MIT, sin considerar sus ingresos por colegiaturas genera, por vinculación, cerca de 27 mil millones de pesos anuales. Esa es la escala económica correcta que las vinculaciones de nuestra Universidad deberían estar generando. La UNAM podría, con políticas de vinculación más ágiles, generar recursos propios cada vez mayores y disminuir su dependencia de fondos federales.

¿Qué se puede mejorar aquí? El problema de la vinculación universitaria es complejo pero, en primer lugar, es importante que las personas encargadas de dirigir las políticas de vinculación tengan experiencia probada, ellas mismas en vinculación. Esto es, que los encargados a nivel directivo de vinculación y emprendimiento en nuestras universidades cuenten ellos mismos con experiencia empresarial, de patentamiento y exposición práctica a la vinculación a un alto nivel. En segundo lugar, es importante que las regulaciones de licenciamiento y patentamiento sean diseñadas en base a esquemas modernos en este tema a nivel internacional.

3. En el aspecto de acceso a tecnologías de la información, la UNAM es una de las instituciones pioneras y líderes de la conectividad y el acceso a redes. Aunque probablemente hay aspectos que mejorar, la conectividad de la UNAM es de de un nivel muy razonable y, en relación al IEC hay poco que comentar en este rubro. 4.-Regulaciones, transparencia y las políticas de fomento. A nivel federal, en diciembre de 2015 se aprobó la nueva Ley Federal de Ciencia y Tecnología. Los aspectos más relevantes de ésta se presentan en el recuadro en este artículo. Esta ley faculta y favorece, de manera liberal y explícita, la formación de empresas de base tecnológica, el fomento a la generación de regalías y recursos financieros por parte de académicos e investigadores vinculados a nivel federal. En contraste con esta regulación importante, la UNAM rige actualmente sus criterios en base a la normatividad universitaria, la cual no contempla explícitamente este tipo de asuntos. Es claro que en la UNAM existen vacíos normativos referentes a nuevos esquemas de transferencia de los productos del conocimiento a la sociedad, tales como el establecimiento de empresas "spin-off". Esto podría obedecer a que históricamente los investigadores de la UNAM han estado alejados de aplicaciones prácticas que impacten el mercado. Sin embargo, en la medida que más investigaciones puedan resolver demandas de la sociedad, es necesario atender a la normatividad faltante. Es fundamental generar, de manera ágil, un marco específico para el licenciamiento y formación de empresas en la UNAM, así como formar departamentos jurídicos modernos, especializados en el tema de

Transferencia. El personal adscrito a estos departamentos jurídicos y de vinculación debería contar, por razones prácticas, con experiencia fuera de la UNAM en aspectos corporativos, de negocios, financieros y de vinculación real en el mundo fuera de la academia, para ser efectivos. Aunque esto es complejo, la simple homologación de las reglas de la UNAM en términos de emprendimiento y

licenciamiento, con las actuales leyes federales, y en particular con la reforma del 2015 a la Ley de Ciencia y Tecnología sería un avance importante.

El tema del incremento del IEC en México es muy complejo. Sin embargo, es importante empezar en casa y preguntarse, como comunidad académica ¿Qué estamos haciendo, qué podemos hacer para contribuir a que nuestro país transite a una Economía Basada en el Conocimiento? ¿Qué podemos hacer para que esto ocurra rápidamente? Limitar nuestro quehacer académico a publicar, ser citados y formar doctores con pocas posibilidades de contratación fuera de la academia no está resultando, por sí solo, una opción que genere riqueza o bienestar tangible para nuestro país. Incrementar el nivel de bienestar y el ingreso de la población mexicana que confía en la UNAM, y que la sostiene con sus impuestos, depende de qué respuesta demos a algunas de las preguntas planteadas arriba, con la honestidad intelectual y autoevaluación critica que debe caracterizarnos como comunidad dedicada a la ciencia. Los estudiantes de ciencia, y los jóvenes que se incorporan a la ciencia tendrán la última palabra sobre la implementación de estas ideas en el futuro, donde el trabajo académico de excelente calidad que se hace en México se complemente con vinculación explícita del conocimiento que se genera en México con las necesidades más urgentes de nuestro país.

Respuesta magnética de metamateriales

Lucila Juárez y W. Luis Mochán Instituto de Ciencias Físicas, UNAM Cuernavaca, Morelos

Resumen

Discutimos la respuesta magnética a bajas frecuencias de un metamaterial compuesto por cilindros metálicos no magnéticos embebidos en un aislante. Obtenemos expresiones para dicha respuesta basada en la dispersión espacial de la respuesta dieléctrica macroscópica del sistema. La respuesta magnética proviene de los cambios en la respuesta dieléctrica asociados a la no uniformidad del campo eléctrico. Desarrollamos un modelo sencillo para el caso de cilindros concéntricos con cortes laterales que permite calibrar cálculos numéricos que posteriormente emplearemos para geometrías arbitrarias.

1. Introducción

Los metamateriales son sistemas artificiales compuestos de arreglos periódicos de microestructuras cuya configuración, geometría y composición determinan su respuesta óptica macroscópica. Controlando la microestructura es posible diseñar metamateriales con propiedades ópticas exóticas, que no se encuentran de forma natural en ningún material. Un ejemplo interesante son los metamateriales conocidos como *izquierdos*. [1, 2]. En éstos existen rangos de frecuencia ω en los cuales tanto la permitividad eléctrica $\epsilon(\omega)$ como la permeabilidad magnética $\mu(\omega)$ son negativas[3]. La ley de inducción de Faraday implica que para una onda plana el campo eléctrico E, la inducción magnética B y el vector de onda k forman una triada ordenada derecha. De igual forma, la ley de Ampere-Maxwell implica que *E*, el campo magnético *H* y el flujo de energía, descrito por el vector de Poynting S, también forman una triada derecha. En los metamateriales izquierdos $B = \mu H$ tiene la dirección opuesta a H, la triada ordenada E, H y k es izquierda y S apunta en la dirección opuesta a k [4, 3, 5]. La conservación del ímpetu a lo largo de una superficie plana de uno de estos metamateriales implica entonces que al refractarse en ella una onda se obedece la ley de Snell pero con un índice de refracción negativo (IRN). Así, los rayos que divergen de una fuente luminosa pueden converger sobre un punto focal después de ser refractados por superficies planas. Aprovechando este efecto es posible, por ejemplo, formar imágenes que alcancen resoluciones por debajo del límite de difracción [6, 7].

Es posible producir un metamaterial izquierdo a partir de incrustaciones metálicas en un medio dieléctrico, donde la geometría de las microestructuras dé origen a una respuesta magnética aún cuando los componentes no sean magnéticos[8]. Un sistema sencillo que ilustra este efecto es un arreglo periódico de tubos cilíndricos metálicos de conductividad muy alta. Un campo magnético oscilando en la dirección del eje del cilindro induciría un campo eléctrico, produciendo una corriente que circularía sobre la superficie de cada cilindro. Esta corriente generaría a su vez un campo magnético en la dirección opuesta al campo inicial, como si el metamaterial fuese diamagnético. El promedio de ésta respuesta dentro de todas las celdas que componen el metamaterial determina la permeabilidad magnética macroscópica del sistema. Un cálculo muy simple muestra que la permeabilidad efectiva de este sistema sería simplemente $\mu = 1 - f$ donde f es la fracción de llenado de los cilindros. Si bien este sistema simple ilustra cómo obtener una respuesta magnética en un metamaterial fabricado con componentes no magnéticas, esta geometría no da lugar a una permeabilidad negativa.

Un sistema un poco más complejo, con el cual sí es posible obtener una permeabilidad $\mu < 0$ es un arreglo de tubos cilíndricos metálicos con una apertura lateral a través de la cual la corriente no puede fluir, causando así una acumulación de carga. El sistema se comporta entonces como un circuito LC (inductorcapacitor) capaz de resonar a una frecuencia característica determinada por la geometría del sistema [9, 10]. Para frecuencias ligeramente superiores a la frecuencia de resonancia es posible obtener valores negativos de la permeabilidad magnética.

Diversos metamateriales con $\mu < 0$ han sido fabricados con éxito, principalmete para frecuencias en el rango de los GHz y THz [11, 12]. Para estas frecuencias, la estructura utilizada mas frecuentemente se conoce en inglés como resonador de anillos partidos (o *split-ring resonator* o SSR en inglés), y consiste en un par de espiras metálicas concéntricas con cortes laterales en direcciones opuestas. Un arreglo periódico tridimensional de SSRs, en combinación con una malla de alambres delgados forman un sistema que puede presentar tanto una $\mu < 0$ como una $\epsilon < 0$, y por tanto un índice de refracción negativo, en un rango de frecuencias que puede ser controlado variando ciertos parámetros de la estructura, como el tamaño, ancho y separación entre las espiras. Sin embargo, como hemos mencionado, este tipo de sistemas opera generalmente en un rango de GHz-THz. La posibilidad de manipular los parámetros de estos sistemas para lograr IRNs a frecuencias mas altas está limitada por el tamaño mínimo requerido para la fabricación de los anillos. Sería por lo tanto importante contar con un método que nos permita explorar geometrías más complejas y estructuras hechas de materiales diversos, con las cuales podamos generar resonancias en otras frecuencias.

El propósito de este artículo es presentar y validar un método de cálculo de la permeabilidad de un metamaterial que en principio permitiría estudiar metamateriales arbitrarios, con geometría arbitraria y con materiales que puedan ser tanto aislantes como conductores, dispersivos y disipativos o no. La estructura del artículo es la siguiente: En la sección (2) calculamos la permeabilidad magnética efectiva de un sistema sencillo compuesto de una red de parejas de cilindros concéntricos con aperturas laterales, en el límite de longitud de onda larga, asumiendo que las paredes de los cilindros son perfectamente conductoras e infinitamente delgadas. Este sistema es una referencia que permite hacer predicciones simples que pueden ser comparadas con modelos de mayor generalidad. En la sección 3 mostramos cómo se puede obtener la respuesta magnética de un metamaterial a partir de la *no localidad* de su respuesta dieléctrica. En la sección 4 presentamos un método recursivo muy eficiente que nos permite calcular la respuesta dieléctrica *no local* de metamateriales de geometría y composición arbitraria, y por ende de su respuesta magnética. En la sección (5) obtenemos algunos resultados para sistemas simples de los cuales obtenemos la respuesta dieléctrica y comparamos su no localidad con la respuesta magnética. Finalmente, en la sección 6 presentamos nuestras conclusiones.

2. Modelo simple

En esta sección estudiaremos la permeabilidad magnética efectiva de un metamaterial formado por una red cuadrada de cilindros metálicos en el límite de bajas frecuencias y longitud de onda larga. Cada celda unitaria de la red contiene un par de cilindros concéntricos que tienen una apertura lateral a través de la cual se interrumpe el flujo de corriente eléctrica. La apertura de los cilindros se encuentra en direcciones opuestas como se muestra en la figura 1.

Consideremos un campo externo de inducción magnética uniforme $B^{\text{ex}}(t) = \text{Re}B^{\text{ex}}e^{-i\omega t}\hat{z}$ que apunta en la misma dirección z que el eje de los cilindros y oscila a frecuencia ω , donde Re(...) denota que debemos tomar la parte real de la cantidad compleja (...), lo cual haremos implícitamente en adelante. El flujo de este campo induce un campo eléctrico que impulsa una corriente superficial a lo largo de la circunferencia de los cilindros conductores. Asímismo, esta corriente eléctrica produce un campo magnético inducido que a su vez genera una contribución adicional al campo eléctrico y a la corriente misma. Por otro lado, los cilindros tienen una apertura que impide el flujo de corriente, por lo cual la corriente eléctrica inducida no será homogénea, sino que variará alrededor de los cilindros y generará acumulaciones de carga que serán una fuente adicional de campo eléctrico. Haremos los cálculos en la aproximación de longitud de onda larga, es decir, supondremos que la frecuencia y el vector de onda son pequeños cuando los normalizamos con el tiempo que tarda la luz en recorrer la estructura y el tamaño de la estructura respectivamente.

Analicemos primero el caso de un solo cilindro de radio a. Como discutimos arriba, debido a presencia de la apertura, el campo magnético externo da lugar a una corriente superficial no uniforme $\mathbf{K}(\theta) = K(\theta)\hat{\theta}$ que oscila con la misma frecuencia ω , donde θ es el ángulo alrededor del eje z medido desde el eje xen la dirección de x a y, y $\hat{\theta}$ es un vector unitario tangencial a la superficie y ortogonal al eje z, como muestra la fig. 1. Note que el problema posee simetría translacional a lo largo del eje z y hemos colocado la apertura de los cilindros de tal forma que el sistema sea simétrico con respecto al eje x. De este modo,



Figura 1: Red cuadrada de cilindros conductores concéntricos con un corte lateral. Se muestran los radios de ambos ciliindros, los ejes coordenados, el ángulo θ y los vectores unitarios $\hat{\theta}$ y \hat{r} .

podemos expresar la corriente como una serie de Fourier en cosenos,

$$K(\theta) = \sum_{l} K_l \cos(l\theta).$$
(1)

De acuerdo a la ley de conservación de la carga, una corriente inhomogénea produciría una acumulación de carga que en dos dimensiones podemos describir mediante una densidad superficial de carga $\sigma(\theta)$. De la ecuación de continuidad en dos dimensiones,

$$\frac{\partial}{\partial t}\sigma + \nabla_{\parallel} \cdot \boldsymbol{K} = 0 \tag{2}$$

obtenemos

$$\sigma(\theta) = \sum_{l} \sigma_{l} \sin(l\theta), \qquad (3)$$

donde

$$\sigma_l = \frac{il}{\omega a} K_l. \tag{4}$$

La carga acumulada origina un potencial electrostático $\Phi(\mathbf{r}) = \Phi(r, \theta)$ que cumple con la ecuación de Laplace $\nabla^2 \Phi(r, \theta) = 0$ tanto afuera del cilindro r > acomo adentro del cilindro r < a. Empleando la solución general de la ecuación de Laplace en coordenadas polares, obtenemos

$$\Phi(r,\theta) = \sum_{l} \begin{cases} \phi_l^d r^l \sin(l\theta), & (r < a) \\ \phi_l^f r^{-l} \sin(l\theta), & (r > a) \end{cases}$$
(5)

donde empleamos la condición de que el potencial producido por el cilindro no debe divergir en r = 0 ni en $r \to \infty$ e introducimos las constantes a determinar $\phi_l^d \ge \phi_l^f$.

El potencial (5) produce un campo eléctrico longitudinal $E^L = -\nabla \Phi(\mathbf{r})$ cuya componente a lo largo de la dirección radial $\hat{\mathbf{r}}$ (fig. 1) escribimos como

$$E_r^L(r,\theta) = \sum_l E_{rl}^L(r)\sin(l\theta),\tag{6}$$

donde

$$E_{rl}^{L}(r) = \begin{cases} -\phi_{l}^{d} l r^{l-1}, & (r < a) \\ \phi_{l}^{f} l r^{-l-1}. & (r > a) \end{cases}$$
(7)

La continuidad $\Phi(a^-,\theta)=\Phi(a^+,\theta)$ del potencial electrostático (5) en r=a implica

$$\phi_l^d a^l - \phi_l^f a^{-l} = 0. (8)$$

Por otro lado, el campo eléctrico radial (7) tiene una discontinuidad $E_r(a^+, \theta) - E_r(a^-, \theta) = 4\pi\sigma(\theta)$ determinada por la carga superficial (ec. (4)), de donde

$$\phi_l^f l a^{-l-1} + \phi_l^d l a^{l-1} = 4\pi\sigma_l.$$
(9)

De las ecs. (4), (8) y 9 podemos despejar

$$\phi_l^f = \frac{2\pi i}{qc} K_l a^l \quad \text{y} \quad \phi_l^d = \frac{2\pi i}{qc} K_l a^{-l}, \tag{10}$$

donde definimos el número de onda en el vacío $q \equiv \omega/c$. Substituyendo en la ec. (7) obtenemos el campo radial

$$E_{rl}^{L}(r) = \frac{2\pi i}{qrc} lK_{l} \times \begin{cases} -(r/a)^{l}, & (r < a) \\ (a/r)^{l}, & (r > a) \end{cases}$$
(11)

Ahora podemos escribir la componente del campo eléctrico longitudinal E^L a lo largo de la dirección tangencial $\hat{\theta}$

$$E_{\theta}^{L}(r,\theta) = \sum_{l} E_{\theta l}^{L}(r) \cos(l\theta)$$
(12)

con

$$E_{\theta l}^{L}(r) = -\frac{2\pi i}{qrc} l K_{l} \times \begin{cases} (r/a)^{l}, & (r < a) \\ (a/r)^{l}, & (r > a) \end{cases}$$
(13)

De las ecuaciones (11) y (13) notamos que el término l = 0 del campo longitudinal es nulo. Por lo tanto, la contribución más importante de este término proviene del campo transversal, que calcularemos más adelante. Por otro lado, el campo eléctrico longitudinal $E^L(r,\theta)$ induce un campo de inducción magnética $B(r) = B(r,\theta)\hat{z}$ en la dirección z que cumple con la ley de Ampere-Maxwell,

$$(\nabla \times \boldsymbol{B}) = -iq\boldsymbol{E}^L,\tag{14}$$

tanto fuera como dentro del cilindro. Notamos que esta ecuación no es exacta, pues puede haber contribuciones adicionales al campo eléctrico y por tanto al campo magnético. Éstas serán discutidas más adelante. Escribiendo la inducción como

$$B(r,\theta) = \sum_{l} B_{l}(r) \cos(l\theta), \qquad (15)$$

tenemos que para $l\geq 1$

$$\left(\nabla \times \boldsymbol{B}\right)_{r} = \frac{1}{r} \partial_{\theta} B(r, \theta) = -\sum_{l} \frac{l}{r} B_{l}(r) \sin(l\theta), \qquad (16)$$

$$(\nabla \times \boldsymbol{B})_{\theta} = -\partial_r B(r,\theta) = -\sum_l \partial_r B_l(r) \cos(l\theta), \qquad (17)$$

donde introdujimos las abreviatura $\partial_r \equiv \partial/\partial r$ y $\partial_{\theta} \equiv \partial/\partial \theta$. Sustituyendo (11) y (16) en (14) obtenemos

$$B_l(r) = \frac{2\pi}{c} K_l \times \begin{cases} (r/a)^l, & (r < a) \\ -(a/r)^l, & (r > a) \end{cases}$$
(18)

que además cumple con la proyección tangencial de la ec. (14) y con la proporcionalidad entre la corriente superficial y la discontinuidad en la superficie $B(a^+, \theta) - B(a^-, \theta) = -4\pi K(\theta)/c$, $B_l(a^+) - B_l(a^-) = -4\pi K_l/c$, como podríamos obtener de integrar la ley de Ampere en un pequeño circuito que atraviese la superficie del cilindro.

Recordemos que el término l = 0 no contribuye al campo eléctrico longitudinal y por lo mismo no hay una corriente de desplazamiento longitudinal con l = 0 y el campo de inducción correspondiente obedece simplemente la ley de Ampere con una corriente superficial homogénea como fuente, i.e., su rotacional es nulo tanto dentro como fuera del cilindro, i.e., $\partial_r B_0(r) = 0$ para r > a y para r < a, con soluciones constantes $B_0(r) = B_0^f$ fuera y $B_0(r) = B_0^d$ dentro. La discontinuidad del campo en la superficie es $B_0^f - B_0^d = -4\pi K_0/c$.

De acuerdo a la ley de Faraday, el campo \boldsymbol{B} calculado arriba induce un campo eléctrico transversal \boldsymbol{E}^T que obedece $\nabla \cdot \boldsymbol{E}^T = 0$ y

$$(\nabla \times \boldsymbol{E}^T) = iq\boldsymbol{B},\tag{19}$$

para el cual podemos hacer una expansión de Fourier, análoga a las ecs. (6) y (12). Para el caso l = 0 la ec. (19) se convierte en

$$\frac{1}{r}\partial_r r E_{\theta 0}^{\gamma T} = iq B_0^{\gamma}, \tag{20}$$
con solución $E_{\theta 0}^{\gamma T}(r) = \frac{iqB_0^{\gamma}}{2}r + c^{\gamma}/r$, donde c^{γ} son constantes a determinar y γ indica si estamos dentro ($\gamma = d, r < a$) o fuera ($\gamma = f, r > a$) del cilindro. Exigiendo que \mathbf{E}^T sea finito para todo r y continuo en r = a, pues no tiene fuentes singulares, obtenemos $B_0^f = 0, c^d = 0$ y $c^f = \frac{iq}{2}B_0^d a^2$, de donde $B_0^d = 4\pi K_0/c$,

$$B_0(r) = \frac{4\pi}{c} K_0 \times \begin{cases} 1, & (r < a) \\ 0, & (r > a) \end{cases}$$
(21)

У

$$E_{\theta 0}^{T}(r) = \frac{2\pi i q a}{c} K_0 \times \begin{cases} r/a, & (r < a) \\ a/r. & (r > a) \end{cases}$$
(22)

Los términos $l \geq 1$ de la componente transversal del campo eléctrico son más pequeños que los correspondientes términos longitudinales por un factor $(qa)^2$ (ver apéndice A) y por tanto, en el límite de bajas frecuencias y longitud de onda larga podemos ignorar su contribución. Por lo tanto, en este límite, el campo eléctrico tangencial total E_{θ} está dominado por el término l = 0 (ec. (22)) del campo transversal y los términos $l \geq 1$ del campo longitudinal (ec. (13)).

Hemos encontrado arriba el campo electromagnético producido en la aproximación de longitud onda larga por una corriente que circula en la superficie de un cilindro. Sin embargo, dicha corriente depende a su vez del sistema físico específico y de la forma como el mismo es excitado. Consideremos primero un cilindro metálico con un apertura de ancho angular 2α y excitado por un campo magnético externo oscilante dirigido a lo largo de su eje. Podemos caracterizar la apertura a través de una función escalón

$$\Theta(\theta) = \begin{cases} 1, & (-\alpha < \theta < \alpha) \\ 0. & (\text{en otro caso}) \end{cases}$$
(23)

Podemos expandir este escalón en una serie de Fourier

$$\Theta(\theta) = \sum_{l=0}^{\infty} \Theta_l \cos(l\theta), \qquad (24)$$

con coeficientes

$$\Theta_l = \begin{cases} \alpha/\pi, & (l=0)\\ 2\sin(l\alpha)/l\pi. & (l>0) \end{cases}$$
(25)

La corriente eléctrica debe ser nula en el escalón, $K(\theta)\Theta(\theta) = 0$, condición que nos lleva a un sistema de ecuaciones acopladas para K_l (ver apéndice B)

$$\sum_{m} \Theta_{lm} K_m = 0 \tag{26}$$

donde Θ_{lm} se relaciona con Θ_l (ec. (25)) a través de

$$\Theta_{lm} = \begin{cases} \Theta_0 \delta_{0m} + \Theta_m, & (l=0) \\ \Theta_0 \delta_{lm} + \Theta_{l+m} + \Theta_{|l-m|}, & (l \ge 1) \end{cases}$$
(27)

con δ_{lm} la delta de Kronecker.

Por otro lado, el campo eléctrico tangencial total tiene tres contribuciones. Una viene del campo inducido por el campo de inducción externo, $E_{\theta}^{\text{ex}}(r) = iqrB^{\text{ex}}/2$. Otra es el campo transversal con l = 0 dado por la ec. (22) y la última es el campo longitudinal con $l \ge 1$ dado por la ec. (13). Evaluando estos tres campos en r = a obtenemos

$$E_{\theta}(a,\theta) = iqaB^{\text{ex}}/2 + \sum_{l=0}^{\infty} e_l K_l \cos(l\theta), \qquad (28)$$

donde

$$e_{l} = \frac{2\pi i q a}{c} \times \begin{cases} 1, & (l=0) \\ -l/(qa)^{2}. & (l \ge 1) \end{cases}$$
(29)

A bajas frecuencias la conductividad de un metal es muy alta, por lo cual supondremos que el cilindro está hecho de un conductor *perfecto* en cuyo interior el campo eléctrico debe anularse. La continuidad del campo eléctrico tangencial implica que éste debe ser nulo en la superficie del conductor, es decir, en todo el cilindro excepto en la apertura, condición que podemos escribir como $E_{\theta}(a, \theta)(1 - \Theta(\theta)) = 0$. Reescribimos ésta como otro sistema de ecuaciones para K_l ,

$$iqaB^{\text{ex}}(\delta_{l0} - \Theta_l) + 2e_lK_l - \sum_{m=0}^{\infty} \Theta_{lm}e_mK_m = 0.$$
 (30)

Aprovechando la formulación obtenida arriba para un solo cilindro, es muy sencillo generalizarla para el caso de un par de cilindros concéntricos de radios ay b > a, y aperturas $2\alpha^{\zeta}$ en lados opuestos del eje x, como ilustra la fig. 1, donde el índice ζ toma los valores a y b. En cada cilindro circula una corriente superficial $K^{\zeta}(\theta)$ caracterizada por sus coeficientes de Fourier K_l^{ζ} . Estas corrientes producen campos eléctricos análogos a los descritos por las ecs. (13) y (22) y los cuales podemos evaluar en r = a, b, lo cual nos lleva a una ecuación análoga a la ec. (28) y que escribimos como

$$E(r = \zeta, \theta) \equiv E_{\theta}^{\zeta}(\theta) = iq\zeta B^{\text{ex}}/2 + \sum_{l=0}^{\infty} \sum_{\eta=a,b} e_l^{\zeta\eta} K_l^{\eta} \cos(l\theta), \quad (\zeta = a, b)$$
(31)

donde $e_l^{\zeta\eta}$ denota la contribución al campo en el cilindro ζ debido a la corriente K_l^{η} que circula en el cilindro η , y toma los valores

$$\begin{aligned}
e_0^{aa} &= \frac{2\pi i q a}{c}, \quad e_0^{ba} &= \frac{2\pi i q a}{c} \frac{a}{b}, \quad e_0^{ab} &= \frac{2\pi i q a}{c}, \quad e_0^{bb} &= \frac{2\pi i q b}{c} \\
e_l^{aa} &= -\frac{2\pi i l}{q c a}, \quad e_l^{ba} &= -\frac{2\pi i l}{q c b} \frac{a^l}{b^l}, \quad e_l^{ab} &= -\frac{2\pi i l}{q c a} \frac{a^l}{b^l}, \quad e_l^{bb} &= -\frac{2\pi i l}{q c b}.
\end{aligned} \tag{32}$$

Caracterizamos las aperturas de ambos cilindros mediante dos funciones escalón $\Theta^{\zeta}(\theta)$ que toman el valor 1 si $\zeta = a$ y $-\alpha^{a} < \theta < \alpha^{a}$ o si $\zeta = b$ y $\pi - \alpha^{b} < \theta < \pi + \alpha^{b}$, y toman el valor 0 en otro caso. Estas funciones están caracterizadas por los coeficientes de Fourier

$$\Theta_l^{\zeta} = \begin{cases} \alpha^{\zeta}/\pi, & (l=0)\\ 2(s^{\zeta})^l \sin(l\alpha^{\zeta})/l\pi, & (l \ge 1) \end{cases}$$
(33)

donde el signo $s^a = +1$ y $s^b = -1$.

En cada apertura la corriente es nula, lo cual expresamos mediante

$$\sum_{m} \Theta_{lm}^{\zeta} K_{m}^{\zeta} = 0, \qquad (34)$$

en analogía con la ecuación (26), donde

$$\Theta_{lm}^{\zeta} = \begin{cases} \Theta_0^{\zeta} \delta_{0m} + \Theta_m^{\zeta}, & (l=0) \\ \Theta_0^{\zeta} \delta_{lm} + \Theta_{l+m}^{\zeta} + \Theta_{|l-m|}^{\zeta}, & (l\ge 1) \end{cases}$$
(35)

en analogía con la ec. (27). En ambos cilindros el campo eléctrico tangencial debe ser cero, lo cual expresamos mediante

$$iq\zeta B^{\text{ex}}(\delta_{l0} - \Theta_l^{\zeta}) + \sum_{\eta} 2e_l^{\zeta\eta} K_l^{\eta} - \sum_{\eta} \sum_{m=0}^{\infty} \Theta_{lm}^{\zeta} e_m^{\zeta\eta} K_m^{\eta} = 0.$$
(36)

en analogía con la ec. (30).

Podemos resolver los sistemas de ecuaciones (26) y (30) para el caso de un solo cilindro, o el sistema de ecuaciones (34) y (36) para el caso de dos cilindros eligiendo un momento angular máximo $l, m < L_{max}$. Entonces, las ecs. (26) y (30) serían dos sistemas de L_{max} ecuaciones con L_{max} variables K_m , mientras que las ecs. (34) y (36) serían dos sistemas de $2L_{max}$ ecuaciones para las $2L_{max}$ variables K_m^{η} . En ambos casos, el sistema aparentemente está sobredeterminado. Sin embargo, no todas las ecuaciones descritas arriba son linealmente independientes y el número de ecuaciones independientes corresponde al número de incógnitas, por lo que los sistemas de ecuaciones pueden ser resueltos numéricamente usando métodos como la descomposición en valores singulares (conocido como SVD por sus siglas en inglés) y L_{max} puede incrementarse hasta obtener convergencia.

Una vez obtenidos las soluciones autoconsistentes de K_l , podemos usar la definición usual del momento dipolar magnético de un circuito (corriente por área/c) para obtener el momento dipolar magnético por unidad de longitud $m/h = \pi a^2 K_0/c$ para el caso de un cilindro o $m/h = \pi a^2 K_0^a/c + \pi b^2 K_0^b/c$ para el caso de dos cilindros, donde $h \to \infty$ es su altura. Para un sistema periódico con celdas unitarias 2D de área A, despreciando las interacciones entre celdas, la magnetización sería M = m/hA, de donde podemos obtener la suceptibilidad $\chi_M = M/H = M/B^{\text{ex}}$ y la permeabilidad $\mu = 1 + 4\pi\chi_M$,

$$\mu = 1 + \frac{1}{B^{\text{ex}}} \frac{4\pi}{c} \times \begin{cases} \frac{\pi a^2}{A} K_0, & \text{(un cilindro)}\\ \frac{\pi a^2}{A} K_0^a + \frac{\pi b^2}{A} K_0^b. & \text{(dos cilindros)} \end{cases}$$
(37)

3. Respuesta magnética y respuesta dieléctrica no-local

Existen dos enfoques para la descripción de las propiedades electromagnéticas de medios continuos. La mas común consiste en introducir una permitividad $\epsilon(\omega)$ y una permeabilidad $\mu(\omega)$ que relacionan entre sí al campo eléctrico E con el desplazamiento eléctrico

$$\boldsymbol{D}_{\omega} = \epsilon(\omega) \boldsymbol{E}_{\omega} \tag{38}$$

y al campo magnético H con el campo de inducción magnética

$$\boldsymbol{B}_{\omega} = \mu(\omega) \boldsymbol{H}_{\omega} \tag{39}$$

para cada frecuencia ω .

La dependencia de las funciones respuesta en ω se conoce como dispersión temporal y es consecuencia del tiempo finito que tarda el sistema en polarizarse y magnetizarse. Es decir, al tiempo t, D y B en un punto dado dependen de E y H no sólo en ese instante, sino también en tiempos anteriores t'. Esto puede apreciarse tomando la transfomada de Fourier temporal de las ecuaciones anteriores y usando el teorema de la convolución,

$$\boldsymbol{D}(t) = \int dt' \boldsymbol{\epsilon}(t - t') \boldsymbol{E}(t') \tag{40}$$

у

$$\boldsymbol{B}(t) = \int dt' \boldsymbol{\mu}(t - t') \boldsymbol{H}(t'), \qquad (41)$$

donde $\epsilon(t - t')$ y $\mu(t - t')$ son las transformadas de Fourier de $\epsilon(\omega)$ y de $\mu(\omega)$ respectivamente. En las ecuaciones previas dejamos implícita la posible dependencia de los campos y de las funciones respuesta en la posición r.

Las respuestas descritas arriba pueden ser insuficientes cuando los campos varían en el espacio con una escala de distancias comparable con la microescala del sistema. En este caso, podría suceder que, por ejemplo, el desplazamiento D(r) en una posición r dependa del campo eléctrico E(r') no sólo en esa posición, sino también en otros puntos cercanos r'. En este caso, es necesario generalizar la ec. (40) escribiéndola como una respuesta *no local*

$$\boldsymbol{D}(\boldsymbol{r},t) = \int dt' \int d\boldsymbol{r}' \boldsymbol{\epsilon}(\boldsymbol{r},\boldsymbol{r}',t-t') \boldsymbol{E}(\boldsymbol{r}',t'), \qquad (42)$$

donde además la respuesta dieléctrica está descrita por un tensor ϵ con componentes ϵ_{ij} . En un medio uniforme, con simetría frente a translaciones arbitrarias, podemos reemplazar la dependencia en $\mathbf{r} \ge \mathbf{r}'$ por una dependencia simple en la separación $\mathbf{r} - \mathbf{r}'$, que podemos simplificar tomando una transformada de Fourier espacio-temporal para obtener la relación algebráica

$$\boldsymbol{D}_{\boldsymbol{k}\omega} = \boldsymbol{\epsilon}(\boldsymbol{k},\omega)\boldsymbol{E}_{\boldsymbol{k}\omega}.$$
(43)

Por lo tanto, un medio no local homogéneo puede describirse por funciones respuesta que dependen tanto de la frecuencia como del vector de onda k. Por analogía con la dispersión temporal, la no localidad de las funciones respuesta se conoce también como *dispersión espacial*.

La posible dependencia de la respuesta dieléctrica tanto en ω como en k nos permite incoporar en ella la respuesta magnética del sistema. Para ello

advertimos primero que es posible escribir las corrientes eléctricas asociadas a la magnetización M, $j_m = c\nabla \times M \rightarrow ikc \times M$ como corrientes asociadas a una polarización P, $j_p = \partial P / \partial t \rightarrow -i\omega P$ si añadimos a P un término de la forma $-kc \times M / \omega$, i.e.,

$$P \to P - kc \times M/\omega, \quad M \to 0.$$
 (44)

Notamos en la ec. (44) que añadimos a la polarización un término proporcional a M y por tanto proporcional a la inducción magnética, según la ec. (39). Sin embargo, la ecuación de Faraday nos muestra que la inducción es proporcional al campo eléctrico

$$\boldsymbol{B}_{\boldsymbol{k}\omega} = \boldsymbol{k}\boldsymbol{c} \times \boldsymbol{E}_{\boldsymbol{k}\omega}/\omega. \tag{45}$$

Por lo tanto, a frecuencias finitas podemos describir completamente la respuesta del material en términos de una polarización que depende del campo eléctrico, elminando así la permeabilidad de la descripción de la respuesta electromagnética de un material y usando exclusivamente la permitividad. El *precio* a pagar por este procedimiento es que la permitividad resultante dependería necesariamente de \mathbf{k} , como se infiere de las ecs. (44 y (45). Por lo tanto, $\boldsymbol{\epsilon}$ sería una cantidad no local. Conversamente, la dependencia de $\boldsymbol{\epsilon}$ en el vector de onda \mathbf{k} puede reinterpretarse en términos de una respuesta magnética $\boldsymbol{\mu}$.

Por lo anterior, al calcular la respuesta macroscópica de un metamaterial fabricado de componentes no magnéticas, podemos considerar iguales a los campos \boldsymbol{B} y \boldsymbol{H} . De esta manera, el promedio de todas las corrientes microscópicas está incluido por completo en la definición del desplazamiento eléctrico \boldsymbol{D} , que se relaciona con el campo eléctrico \boldsymbol{E} mediante el tensor dieléctrico $\boldsymbol{\epsilon}$ el cual es necesariamente no local en presencia de efectos magnéticos.

Recordemos que el metamaterial es un cristal artificial con una celda unitaria que se repite periódicamente. Cuando el tiempo que tarda la luz en recorrer una celda unitaria del metamaterial se vuelve comparable al periodo, o equivalentemente, cuando la longitud de onda es comparable al parámetro de red, los los efectos no-locales se vuelven significativos. Esto puede conducir, como discutimos arriba, a una respuesta magnética efectiva, aún cuando los componentes del metamaterial no sean magnéticos [13]. Es por esta razón que resulta necesario tomar en cuenta la no-localidad en la descripción de la respuesta óptica de los metamateriales con IRN. El tensor dieléctrico no local $\epsilon(\mathbf{k}, \omega)$ describe tanto la respuesta dieléctrica como la respuesta magnética del sistema[14, 15]. El límite local de la permitividad $\epsilon(\omega)$ corresponde al límite en que el campo eléctrico varía lentamente, $\epsilon(\mathbf{k} \to \mathbf{0}, \omega)$ del enfoque no-local. Análogamente, el límite local de la permeabilidad puede obtenerse a partir de las primeras correcciones no locales a la permitividad. En particular, puede obtenerse a partir de una expansión de $\epsilon(\mathbf{k}, \omega)$ en potencias de \mathbf{k} para vectores de onda pequeños, como veremos mas adelante.

En la siguiente sección presentamos un formalismo basado en la recursión de Haydock para el cálculo de la permitividad macroscópica $\epsilon_M(\mathbf{k},\omega)$, el cual hemos implementado en un eficiente paquete computacional que permite calcular la función dieléctrica macroscópica de metamateriales a partir de la su geometría y de las propiedades de los materiales que los componen [16, 17, 18]. La dependencia de la respuesta dieléctrica resultante en el vector de onda nos permitirá obtener una aproximación para la permeabilidad magnética local macroscópica del sistema.

4. Cálculo recursivo

En esta sección desarrollaremos un esquema eficiente para el cálculo de la respuesta dieléctrica no local de metamaterials con geometrías y composiciones arbitrarias. Para ello empezamos por eliminar el campo magnético de las ecuaciones de Faraday y de Ampere-Maxwell llegando inmediátamente a una ecuación diferencial de segundo orden para el campo eléctrico E,

$$\nabla \times \nabla \times \boldsymbol{E} = \frac{4\pi i\omega}{c^2} \boldsymbol{J}_{\text{ex}} + \frac{\omega^2}{c^2} \hat{\boldsymbol{\epsilon}} \boldsymbol{E}, \qquad (46)$$

donde J_{ex} es una corriente eléctrica externa, no asociada a las corrientes de conducción, de polarización o de magnetización del sistema, c la velocidad de la luz en el vacío. Usando que el rotacional de un campo longitudinal es nulo, y que el rotacional del rotacional de un campo transversal es menos su laplaciano, podemos reescribir la ec.(46) como una ecuación de onda con fuentes,

$$\hat{\mathcal{W}}\boldsymbol{E} = -\frac{4\pi i}{\omega}\boldsymbol{J}_{\text{ex}} \tag{47}$$

donde

$$\hat{\mathcal{W}} = \left(\hat{\epsilon} + \frac{1}{q^2}\nabla^2\hat{\mathcal{P}}_T\right) \tag{48}$$

es el operador de onda microscópico, $\hat{\epsilon}$ es la función dieléctrica, que toma un valor característico en cada uno de los materiales que conforman al metamaterial, por lo cual la designaremos como la *respuesta microscópica*, y la cual escribimos como un operador lineal por economía de notación como apreciaremos adelante, $\hat{\mathcal{P}}_T$ es un proyector que actuando sobre un campo vectorial cualquiera \boldsymbol{F} nos lleva a su proyección transversal $\boldsymbol{F}_T = \hat{\mathcal{P}}_T \boldsymbol{F}$, y $q = \omega/c$ es el número de onda de una onda de frecuencia ω que se propaga libremente libre en el vacío.

Despejando formalmente al campo eléctrico de la ec. (47) obtenemos

$$\boldsymbol{E} = -\frac{4\pi i}{\omega} \hat{\mathcal{W}}^{-1} \boldsymbol{J}_{\text{ex}}$$
(49)

y promediando obtenemos el campo eléctrico macroscópico,

$$\boldsymbol{E}_{M} \equiv \boldsymbol{E}_{p} = -\hat{\mathcal{P}}_{p} \frac{4\pi i}{\omega} \hat{\mathcal{W}}^{-1} \boldsymbol{J}_{\text{ex}}, \qquad (50)$$

donde introdujimos el proyector *promedio* $\hat{\mathcal{P}}_p$ tal que al actuar sobre un campo cualquiera \boldsymbol{F} produce su promedio espacial $\boldsymbol{F}_p = \hat{\mathcal{P}}_p \boldsymbol{F}$. En este momento no es necesario aún especificar qué entendemos por promedio. Nos basta con cualquier operador que sea idempotente (el promedio del promedio es el promedio), que no dependa de la posición ni del tiempo para que conmute con los operadores diferenciales y que elimine de cualquier campo las fluctuaciones espaciales asociadas a la textura microscópica del metamaterial. Notamos en el lado derecho de esta ecuación que el promedio actúa sobre el producto de una constante multiplicada por un operador que actua linealmente sobre un campo vectorial. Típicamente, el promedio de un producto no es igual al producto de los promedios, pues puede haber *fluctuaciones correlacionadas*. Sin embargo, en el extremo derecho de la ec. (50) tenemos a la corriente externa, la cual, siendo *externa*, no tiene nada que ver con la textura del metamaterial. Por lo tanto, no tiene fluctuaciones espaciales provenientes del metamaterial, y es igual a su promedio espacial, $J_{ex} = \hat{\mathcal{P}}_p J_{ex}$. De no ser así, no tendría sentido elaborar una teoría macroscópica de la respuesta a dicha corriente. Por lo tanto, podemos reescribir la ec. (50) como

$$\boldsymbol{E}_{M} = -\frac{4\pi i}{\omega} \hat{\mathcal{W}}_{M}^{-1} \boldsymbol{J}_{\text{ex}}, \qquad (51)$$

donde identificamos la inversa del operador de onda macroscópico

$$\hat{\mathcal{W}}_{M}^{-1} = \hat{\mathcal{W}}_{pp}^{-1} \equiv \hat{\mathcal{P}}_{p} \hat{\mathcal{W}}^{-1} \hat{\mathcal{P}}_{p}, \qquad (52)$$

empleando la idempotencia del promedio. Podemos expresar este resultado diciendo que *el inverso del operador de onda macroscópico es el promedio del inverso del operador de onda microscópico*. En analogía a la ec. (48), interpretamos al operador de onda macroscópico en términos de un operador dieléctrico macroscópico

$$\hat{\mathcal{W}}_M = \left(\hat{\boldsymbol{\epsilon}}_M + \frac{1}{q^2} \nabla^2 \hat{\mathcal{P}}_T\right) \tag{53}$$

de donde podemos despejar fácilmente $\hat{\epsilon}_M$. En resumen, nuestro procedimiento de cálculo es partir de la respuesta dieléctrica microscópica para escribir el operador de onda microscópico, invertirlo, promediarlo, invertir el resultado y finlamente extraer la respuesta dieléctrica macroscópica.

Consideramos ahora un metamaterial binario, fabricado de dos materiales isotrópicos $A ext{ y } B$, con permitividades $\epsilon_a(\omega) ext{ y } \epsilon_b(\omega)$ respectivamente, las cuales son funciones complejas de la frecuencia en general. Podemos entonces reescribir el operador de onda microscópico como

$$\hat{\mathcal{W}} = \frac{\epsilon_a}{u} \left(u - \hat{\mathcal{B}} \right) + \frac{\nabla^2}{q^2} \hat{\mathcal{P}}_T \tag{54}$$

donde

$$u = \frac{1}{1 - \frac{\epsilon_b}{\epsilon_a}} \tag{55}$$

es conocida como la variable espectral, y el operador $\hat{\mathcal{B}}$ corresponde a la función característica $\mathcal{B}(\mathbf{r})$ del medio B, la cual toma el valor 1 cuando la posición \mathbf{r} está en la región ocupada por el material B y toma el valor 0 en caso contrario.

Vale la pena mencionar que la variable espectral $u(\omega)$ contiene la información sobre la composición del metamaterial y sobre la respuesta de sus componentes a distintas frecuencias ω pero no depende de la geometría, mientras que la función característica depende únicamente de la geometría y no de la composición. Podemos reescribir al operador de onda como

$$\hat{\mathcal{W}} = \frac{\epsilon_a}{u} \left(u \hat{g}^{-1} - \hat{\mathcal{B}} \right), \tag{56}$$

donde definimos el operador

$$\hat{g} = \left(\mathbf{1} + \hat{\mathcal{P}_T} \frac{\nabla^2}{q^2 \epsilon_a}\right)^{-1},\tag{57}$$

el cual veremos que juega el papel de una *métrica*.

Supondremos además que nuestro metamaterial es un cristal artificial con una celda unitaria que se repite periódicamente. Este cristal está caracterizado entonces por una red de Bravais $\{R\}$ y su correspondiente red recíproca $\{G\}$, de forma tal que, de acuerdo al *teorema de Bloch*, cualquier onda plana que se propague en este cristal con algún vector de onda k sería difractada, acoplándose con ondas con vectores de onda de la forma k + G. Así, las excitaciones del sistema tendrán la forma de una *onda de Bloch*. El hecho de que la red recíproca sea *discreta* nos sugiere definir el promedio de un campo cualquiera F(r) descrito por una onda de Bloch con vector de Bloch k,

$$\boldsymbol{F}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} \boldsymbol{F}_{\boldsymbol{k}\boldsymbol{G}} e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}},\tag{58}$$

como la contribución cuyo vector de onda es k,

$$\boldsymbol{F}_{p}(\boldsymbol{r}) = \hat{\mathcal{P}}_{P} \sum_{\boldsymbol{G}} \boldsymbol{F}_{\boldsymbol{k}\boldsymbol{G}} e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}} = \boldsymbol{F}_{\boldsymbol{k}\boldsymbol{0}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}.$$
(59)

Nuestra elección de promedio descarta todos los vectores de onda $\mathbf{k} + \mathbf{G}$ correspondientes a vectores recíprocos $\mathbf{G} \neq \mathbf{0}$ y sólo conserva aquella componente con $\mathbf{G} = \mathbf{0}$. Entonces podemos representar al proyector promedio como $\hat{\mathcal{P}}_p \rightarrow \delta_{\mathbf{G0}}$ en términos de una delta de Kronecker. Una interpretación de este operador promedio es como un filtro en el espacio recíproco que sólo conserva aquellos vectores de onda en la primera zona de Brillouin y elimina todos los demás.

Usando el promedio anterior, podemos emplear la ec. (52) para escribir el operador de onda inverso macroscópico como

$$\hat{\mathcal{W}}_{M}^{-1} = \frac{u}{\epsilon_{a}} \hat{g}_{pp} \left(u - \hat{\mathcal{B}} \hat{g} \right)_{pp}^{-1}, \tag{60}$$

donde usamos el hecho de que el operador \hat{g} no mezcla distintos vectores de onda, y por tanto *conmuta* con el operador promedio. Notamos en el lado derecho de esta ecuación que contiene una expresión análoga a un operador de Green,

$$\hat{\mathcal{G}} = (\epsilon - \hat{\mathcal{H}})^{-1},\tag{61}$$

como podría aparecer en un problema ordinario de mecánica cuántica, donde identificamos la energía compleja ϵ con la variable espectral $\epsilon \to u$ y al hamiltoniano $\hat{\mathcal{H}}$ con el producto $\hat{\mathcal{H}} \to \hat{\mathcal{B}}\hat{g}$. Una objeción a la identificación anterior es que el hamiltoniano es un operador Hermitiano, mientras que el producto $\hat{\mathcal{B}}\hat{g}$ parece no serlo. Sin embargo, este producto se vuelve hermitiano si reemplazamos el producto usual $\langle \psi | \psi \rangle$ entre dos vectores de estado $| \phi \rangle$ y $| \psi \rangle$ por un nuevo producto escalar

$$(\phi|\psi) \equiv \langle \phi|\hat{g}|\psi\rangle, \qquad (62)$$

en el que el operador \hat{g} juega el papel de una *métrica*. Bajo este producto escalar se verifica que

$$(\phi|\hat{\mathcal{B}}\hat{g}|\psi) = \langle \phi|\hat{g}\hat{\mathcal{B}}\hat{g}|\psi\rangle = \langle \psi|\hat{g}\hat{\mathcal{B}}\hat{g}|\phi\rangle^* = (\psi|\hat{\mathcal{B}}\hat{g}|\phi)^*, \tag{63}$$

por lo cual $\hat{\mathcal{B}}\hat{g}$ sí es hermitiano con respecto a esta métrica.

La analogía del operador de onda macroscópico con la función de Green de problemas cuánticos simples, nos permite tomar *prestados* resultados de la mecánica cuántica para obtener la respuesta macroscópica. En particular, recurriremos a la *recursión de Haydock* [16, 17, 18]. En este método elegimos un *estado* $|0\rangle$ inicial normalizado respecto a la métrica \hat{g} y correspondiente a una onda plana con el vector de onda \mathbf{k} y cierta polarización \mathbf{e} . A partir de $|0\rangle$ podemos generar nuevos estados actuando recursivamente con nuestro hamiltoniano. Dados los estados $|0\rangle, |1\rangle, |2\rangle \dots |n\rangle$, generamos el estado $|n + 1\rangle$ mediante

$$|\widetilde{n+1}\rangle \equiv \hat{\mathcal{B}}\hat{g}|n\rangle = b_{n+1}|n+1\rangle + a_n|n\rangle + b_n g_n|n-1\rangle, \tag{64}$$

donde elegimos los *coeficientes de Haydock* $a_n \ge b_n$ de manera que los estados $|n\rangle$ sean *ortonormales*

$$(n|m) = \langle n|\hat{g}|m\rangle = g_n \delta_{nm} \tag{65}$$

respecto a la métrica \hat{g} , y $g_n = \pm 1$. Necesitamos introducir el signo g_n pues la métrica \hat{g} no es necesariamente positiva definida. Notamos que en esta base el operador $\hat{\mathcal{B}}\hat{g}$ es tridiagonal y sus componentes se obtienen de los coeficientes de Haydock, por lo que podemos escribir

$$\left(u - \hat{\mathcal{B}}\hat{g}\right) = \begin{pmatrix} u - a_0 & -b_1g_1g_0 & 0 & 0 & \dots \\ -b_1 & u - a_1 & -b_2g_2g_1 & 0 & \dots \\ 0 & -b_2 & u - a_2 & -b_3g_3g_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$
(66)

De acuerdo a la ecuación (60) la respuesta macroscópica del sistema está dada en términos del promedio, que identificamos como el elemento 00, de la inversa de esta matriz. Este elemento se puede obtener en forma de la fracción continuada

$$\boldsymbol{e} \cdot \boldsymbol{\mathcal{W}}_{M}^{-1} \cdot \boldsymbol{e} = \frac{u}{\epsilon_{A}} \frac{g_{0}b_{0}^{2}}{u - a_{0} - \frac{g_{0}g_{1}b_{1}^{2}}{u - a_{1} - \frac{g_{1}g_{2}b_{2}^{2}}{u - a_{2} - \frac{g_{2}g_{3}b_{3}^{2}}{\cdot \cdot \cdot}}}$$
(67)

Los coeficientes de Haydock que aparecen en (67) provienen de la aplicación reiterada de la función característica, que depende exclusivamente de la geometría, y de la métrica, la cual tiene una dependencia trivial en el número de onda $k_a = \sqrt{\epsilon_a}q$ de las ondas libres en el medio A. Para cada valor de k_a , que a su vez corresponde a un valor del número de onda libre $q = \omega/c$, la dependencia en los componentes del metamaterial y su respuesta dinámica está contenida únicamente en la variable espectral u. De la ec. (67) obtenemos la proyección del operador de onda macroscópico \mathcal{W}_M^{-1} a lo largo del vector de polarización e. Repitiendo el cálculo anterior para suficientes direcciones e independientes, podemos encontrar todas las componentes de \mathcal{W}_M^{-1} , el cual podemos interpretar como un tensor cartesiano para cada valor de la frecuencia y del vector de onda k. A continuación podemos invertir este tensor para finalmente obtener el tensor dieléctrico macroscópico

$$\boldsymbol{\epsilon}_M(\boldsymbol{k},\omega) = \frac{1}{q^2} (k^2 \mathbf{1} - \boldsymbol{k} \boldsymbol{k}) + \boldsymbol{\mathcal{W}}^M(\boldsymbol{k},\omega).$$
(68)

El método descrito anteriormente ha sido implementado en un paquete computacional que emplea el expresivo lenguaje PERL y su eficiente extensión numérica *Perl Data Language* [19, 20] y ha sido codificado como un sistema de objetos en un paquete llamado *Photonic* que actualmente se encuentra disponible al público [21]. El elemento central en el cálculo es la función característica $\mathcal{B}(\mathbf{r})$ la cual describe la geometría de la celda unitaria y la cual se puede representar como un arreglo D-dimensional de unos y ceros, donde 1 corresponde a los pixeles o voxeles ocupados por el material B y los ceros a aquellos ocupados por el material A.

Es importante notar que la métrica incluye al operador laplaciano, que en la representación del espacio recíproco, donde los estados $|\psi\rangle$ son representados como una suma de ondas planas, corresponde a $\nabla^2 \rightarrow -||\mathbf{k} + \mathbf{G}||^2$. Por lo tanto, la respuesta dieléctrica obtenida en este esquema es no local. La dispersión espacial se manifiesta a través de la dependencia en el vector de onda \mathbf{k} adicional a la dependencia en la frecuencia ω , como se observa explícitamente en la ec. (68). De nuestro cálculo no local podríamos obtener una respuesta dieléctrica local tomando el límite de longitudes de onda largas, vectores de onda pequeños,

$$\boldsymbol{\epsilon}_M(\omega) = \boldsymbol{\epsilon}_M(\boldsymbol{k} \to \boldsymbol{0}, \omega). \tag{69}$$

Sin embargo, este límite local es insuficiente para el cálculo de propiedades ópticas, pues la relación de dispersión electromagnética en un material es cuadrática en el vector de onda. Por lo tanto, es indispensable aproximar la respuesta dieléctrica al menos hasta el orden cuadrático en potencias de k. De esta expansión es posible obtener, a un primer orden de aproximación, la respuesta magnética local asociada a la dispersión espacial de ϵ como veremos a continuación.

La forma más sencilla de hallar la permeabilidad magnética correspondiente se sigue de la relación de dispersión de las ondas electromagnéticas transversales a lo largo de los ejes de simetría y para campos polarizados a lo largo de direcciones principales. Llamando $\epsilon_M(k,\omega)$ al valor principal correspondiente de ϵ_M y habiendo fijado la dirección de k, la relación de dispersión se sigue inmediatamente de la ecuación de onda y es

$$k^2 = \epsilon_M(k,\omega)q^2. \tag{70}$$

Haciendo una expansión de Taylor hasta orden cuadrático,

$$k^{2} = \left(\epsilon_{M}(0,\omega) + \frac{1}{2}k^{2}\frac{\partial^{2}}{\partial k^{2}}\epsilon_{M}(k,\omega)\Big|_{k=0}\right)q^{2},$$
(71)

donde hemos eliminado el término lineal en k suponiendo que el sistema es invariante frente a inversiones temporales $t \to -t$, $k \to -k$. Despejando k^2 de esta ecuación, podemos reescribirla para obtener la forma convencional de la relación de dispersión

$$k^2 = \epsilon_M(\omega)\mu_M(\omega)q^2, \qquad (72)$$

donde identificamos la permeabilidad magnética como

$$\mu_M(\omega) = \frac{1}{1 - q^2 \left. \frac{1}{2} \frac{\partial^2}{\partial k^2} \epsilon_M(k, \omega) \right|_{k \to 0}}.$$
(73)

En resumen, hemos mostrado un esquema que nos permite calcular la permitividad y la permeabilidad locales macroscópicas de un metamaterial binario con una composición y geometría arbitrarias en el límite de longitud larga. Para esto, obtenemos primero la respuesta dieléctrica no local usando un formalismo numéricos como el método de Haydock presentado en esta sección, y posteriormente tomando su límite local y calculando su segunda derivada respecto al vector de onda con la cual encontramos una expressión aproximada para la permeabilidad magnética del sistema.

5. Resultados

Como un primer cálculo para probar el formalismo de la sección anterior, consideremos una red cuadrada con parámetro de red d de cilindros conductores huecos sumergidos en el espacio vacío sin apertura, con un radio exterior a = 0.5d y un radio interior 0.1d.

De acuerdo a nuestro modelo simple para conductores perfectos, (sección 2) cuando los cilindros no tienen aperturas (el ángulo $\alpha = 0$ en la Ec. (23)) las ecs. (26) desaparecen y las ecs. (30) conducen a $K_0 = -iqaB^{\text{ex}}/2e_0$ y $K_l = 0$ para $l \ge 1$, por lo cual la Ec. (37) lleva al resultado simple $\mu = 1 - f$, donde $f = \pi a^2/d^2$ es la fracción de llenado de la región encerrada por la superficie exterior de los cilindros.

Utilizando el formalismo presentado en la sección anterior realizamos cálculos numéricos de la permitividad no-local macroscópica ϵ_M de dicho metamaterial asignándole diferentes permitividades ϵ_a a los conductores. Recordemos que a bajas frecuencias ϵ_a es vuelve negativa y de mayor tamaño mientras mejor es el conductor.



Figura 2: Permitividad eléctrica ϵ_M a frecuencia $\omega = 0.1c/d$ obtenida numéricamente (círculos) y analíticamente (línea punteada), como función del cuadrado k^2 del vector de onda, para una red cuadrada de clindros metálicos huecos con radio exterior 0.4d y radio interior 0.1d, correspondiente a una llenado (exterior) f = 0.5, donde d es el parámetro de red. La línea continua muestra el límite para un conductor perfecto. Los resultados se muestran para diferentes valores de la permitividad eléctrica del metal como se indica en el interior de cada recuadro.

También llevamos a cabo un cálculo analítico resolviendo la ecuación de Helmholtz para el campo magnético en la región metálica y empatándolo con un campo magnético que no depende de la posición en la región del vacío, donde suponemos vale la aproximación de longitud de onda larga.

La figura 2 muestra los resultados de la componente transversal del tensor de permitividad ϵ_M no local a frecuencia $\omega = qc \operatorname{con} q = 0.1/d \operatorname{como}$ función del cuadrado k^2 del vector de onda para una red cuadrada de cilindros metálicos huecos con radio interior 0.1d y radio exterior a = .4d, correspondiente a la fracción de llenado $f \approx 0.5$. Si el conductor fuera perfecto la distancia de penetración del campo sería cero y el ancho de las paredes sería irrelevante, el resultado dependería exclusivamente del radio exterior de los cilindros. En la figura mostramos los resultados que esperaríamos para un cilindro conductor perfecto, $\mu = 1 - f$ y los resultados del cálculo analítico en la aproximación de longitud de onda larga.

Para una permitividad ϵ_a negativa, cuyo tamaño es del orden de 10², la aproximación analítica no describe satisfactoriamente las correcciones no locales a la respuesta macroscópica del sistema, pues la permitividad del cilindro es sólo dos órdenes de magnitud mayor que la del vacío, por lo que la distancia de penetración del campo en el cilindro es apenas un orden de magnitud menor que la longitud de onda en el vacío. Esto es consistente con la discrepancia que muestra la fig. 2 entre el resultado analítico y el resultado numérico. La discrepancia con el modelo de conductor perfecto es aún mayor. Sin embargo, para permitividades metálicas con valor absoluto del orden 10³ o mayor, los resultados numéricos tienen un excelente acuerdo con los del modelo analítico. Finalmente, observamos también, que al aumentar aún más la permitividad de los cilindros metálicos, cuando alcanza valores del orden 10^7 o mayores, obtenemos finalmente un acuerdo con el resultado esperado para conductores perfectos. Resultados similares se obtuvieron para otras fracciones de llenado de los cilindros. Sin embargo, en nuestro modelo numérico no podemos llegar a valores mucho mayores de ϵ_a por dificultades de convergencia. Recordemos que el cálculo numérico inicia a partir de la función característica $\mathcal{B}(\mathbf{r})$ que representamos como un arreglo de pixeles o voxeles. No podemos esperar que el método numérico funcione cuando la distancia de penetración del campo sea menor a un pixel, pues sería imposible representar dicho decaimiento con un arreglo discreto. Por lo tanto, al aumentar el valor absoluto de ϵ_a es necesario aumentar el tamaño de nuestros arreglos, con el consecuente aumento del tiempo de cómputo.

El hecho de que para este sistema los resultados numéricos coincidieran con los resultados de aproximaciones simples en aquellas regiones donde esperamos que las mismas sean válidas, sugieren que la permeabilidad magnética a bajas frecuencias sí está bien descrita por la dependencia en el vector de onda de la permeabilidad macroscópica no local.

Habiendo verificado la eficiencia y las limitaciones de nuestro método con un metamaterial muy sencillo, podemos proceder a analizar la respuesta magnética de un metamaterial más interesante, que permita obtener resonancias y valores negativos de la permeabilidad. En particular, sería importante investigar la geometría del tipo *split-ring* que analizamos en la sección 2, con el fin de



Figura 3: Permeabilidad magnética μ de una red cuadrada de cilindros conductores perfectos de radio a = 0.4d y fracción de llenado $f \approx 0.5$ para distintas semiaperturas angulares α entre 0.01 y 0.1. El cálculo fue realizado con $L_{\text{max}} = 100$ valores para el momento angular l.

comparar nuestros resultados y valorar la posibilidad de aplicar nuestro método en geometrías aún mas complejas.

Como un primer paso, exploraremos la respuesta de una red de cilindros conductores perfectos interrumpidos mediante una apertura de ancho angular 2α empleando el formalismo de la sección 2. La fig. 3 muestra la permitividad como función de la frecuencia para diversos valores de α . Vemos que para $\alpha = 0.1$ se ve claramente una resonancia tras de la cual la permeabilidad toma valores negativos. Sin embargo, dicha resonancia aparece a una frecuencia $qd = \omega d/c$ muy cercana a 1, por lo cual no está justificada la aproximación de longitud de onda larga. Intentamos recorrer la resonancia a frecuencias menores cerrando la brecha. Si bien para $\alpha = 0.05$ y $\alpha = 0.04$ la frecuencia de la resonancia disminuye, lo hace muy ligeramente. Para α más pequeña, la resonancia aparentemente desaparece.

Un análisis más cuidadoso nos hace ver que aperturas tan pequeñas como 0.01-0.03 difícilmente pueden representarse mediante una serie de Fourier (24) truncada con un valor tan pequeña de L_{max} . El principio de incertidumbre nos sugeriría que L_{max} debe ser mayor a $1/2\alpha$. Para verificar ésto, en la fig. 4 repetimos el cálculo anterior para una apertura pequeña pero para distintos valores de L_{max} . Observamos que al aumentar L_{max} , reaparece la resonancia que habíamos perdido al cerrar la apertura. Sin embargo, la posición de la resonancia sigue siendo muy alta, a pesar de que escogimos una apertura tan pequeña que difícilmente podría realizarse experimentalmente.



Figura 4: Permeabilidad magnética μ de una red cuadrada de cilindros conductores como los de la figura 3 con una semiapertura $\alpha = 0.025$ pero calculado con distintos valores de $L_{\rm max}$ entre 50 y 300.

Podemos entender el problema con los sistemas recién explorados si los visualizamos como unos circuitos LC. La capacitancia proviene de la brecha, pero como nuestros cilindros son infinitamente delgados, dicha capacitancia es muy pequeña y por tanto la resonancia $1/\sqrt{LC}$ corresponde a una frecuencia tan alta que las aproximaciones no son válidas. Para incrementar la capacitancia y contar con más parámetros que permitan entonar la resonancia es que se ha recurrido a cilindros dobles con aperturas, como los ilustrados en la fig. 1. Como muestra la fig. 5 con aperturas relativamente anchas podemos obtener resonancias bien desarrolladas y entonarlas hacia valores bajos de la frecuencia donde las aproximaciones realizadas conducen a errores pequeños.

Finalmente, en la figura 6 mostramos el resultado de nuestro cálculo numérico no local para una red de cilindros dobles abiertos. Para hacer el cálculo numérico obtuvimos para cada frecuencia la componente transversal del tensor de permitividad ϵ_M como función del cuadrado k^2 del vector de onda, ajustamos una recta al resultado y obtuvimos la permeabilidad μ_M usando la ec. (73). Tal como esperábamos tras analizar la fig. 5, obtuvimos una resonancia en la permeabilidad a frecuencias no demasiado altas arriba de la cual toma valores negativos. Como una referencia, también mostramos en la figura los resultados del modelo simple. Como la distancia de penetración del campo en un metal a bajas frecuencias es muy pequeña, empleamos el radio mayor del cilindro interior y el radio menor del cilindro exterior como los radios a y b que aparecen en dicho modelo. El acuerdo cualitativo entre los resultados de ambos modelos es muy bueno, aunque su frecuencia de resonancia no es idéntica. Esto se debe a que en



Figura 5: Permeabilidad magnética μ de una red cuadrada de cilindros conductores dobles con radio b = 0.4d y distintos valores del radio a y entre 0.3d y 0.37d y semiaperturas $\alpha = \beta$ entre 0.2 y 0.8.

el modelo numérico no podemos alcanzar el límite de un conductor perfecto ni podemos hacer que su ancho se anule, por lo que la respuesta microscópica y la geometría no son idénticas entre ambos modelos.

Notamos en las figs. 3-6 que a bajas frecuencias la permitividad $\mu_M \approx 1$, pues los anillos abiertos no pueden sostener una corriente DC, mientras que a altas frecuencias $\mu \approx 1 - f$ pues los capacitores se comportan como circuitos cerrados y los anillos se comportan como si no tuvieran aperturas.

Encontramos entonces un rango de frecuencias entonable en el cual la permeabilidad es negativa y sería posible construir un metamaterial izquierdo, siempre y cuando logremos al mismo tiempo una permitividad eléctrica negativa. Esto se puede lograr combinando el sistema que analizamos arriba con una malla de alambres metálicos muy delgados que hagan que el sistema se comporte macroscópicamente como un mal conductor.

6. Conclusiones

Hemos presentado un método simple para el cálculo de la respuesta magnética macroscópica de metamateriales que contienen fases conductoras perfectas con geometrías muy sencillas y un método que permite el cálculo de la respuesta dieléctrica no local de metamateriales de geometría y composición arbitraria. Asímismo, hemos mostrado cómo relacionar la no localidad de la respuesta dieléctrica con la respuesta magnética. Hemos aplicado ambos métodos a sistemas



Figura 6: Permeabilidad magnética μ calculada numéricamente para una red cuadrada de parejas de cilindros conductores abiertos. Un cilindro de cada pareja tiene radio interior 0.26d y radio exterior a = 0.32d. El otro tiene radio interior b = 0.34d y radio exterior 0.4d. La apertura corresponde a una separación 0.06d. La permitividad de los metales se tomó como $\epsilon_a = -10^4$. También se muestran los resultados del modelo simple para parejas de cilindros conductores perfectos con paredes infinitamente delgadas y radios a y b con la misma apertura.

simples y hemos comparado los resultados, lo cual nos permite concluir que el cálculo de la respuesta magnética a partir de la no localidad es viable y conduce a buenos resultados, con la ventaja de ser un método aplicable a geometrías variadas y complejas y permite atacar sistemas realistas fabricados con materiales dispersivos y disipativos.

Agradecimientos

Este trabajo se realizó con apoyo de DGAPA-UNAM a través de una beca postdoctoral para LJ y del proyecto IN113016.

A. Apéndice: Campo transversal

La corriente eléctrica (1) y el campo eléctrico longitudinal ((12) y (13)) producido por la correspondiente acumulación de carga da lugar a una inducción magnética \boldsymbol{B} ((18) y (21)) en la dirección z que a su vez da origen a un campo eléctrico transversal que cumple la ec. (19). Expandemos las componentes radiales y tangenciales de este campo de manera análoga a la mostradoa en las ecs. (6) y (12) definiendo así contribuciones $E_{rl}^T(r)$ y $E_{\theta l}^T(r)$ para cada l. Podemos entonces reescribir la Ec. (19) para $l \geq 1$ como

$$\frac{1}{r}(\partial_r r E_{\theta l}^T(r) - l E_{rl}^T(r)) = \frac{2\pi i q}{c} K_l \times \begin{cases} (r/a)^l, & (r < a) \\ -(a/r)^l, & (r > a) \end{cases}$$
(74)

Dado que E^T es un campo transversal, cumple con $\nabla \cdot E^T = 0$, lo cual reescribimos como

$$\partial_r r E_{rl}^T(r) = l E_{\theta l}^T(r), \tag{75}$$

que usamos para eliminar $E_{\theta l}^T(R)$ de la ec. (74) y obtener

$$\partial_r r \partial_r r E_{rl}^T(r) - l^2 E_{rl}^T(r) = \frac{2\pi i q}{c} l K_l r \times \begin{cases} (a/r)^l, & (r < a) \\ -(a/r)^l, & (r > a) \end{cases}$$
(76)

con una solución particular de la forma

$$E_{rl}^{T}(r) = \begin{cases} E_{r}^{d} r^{l+1}, & (r < a) \\ E_{r}^{f} r^{-l+1}. & (r > a) \end{cases}$$
(77)

Sustituyendo la ec. (77) en (76) obtenemos el campo radial

$$E_{rl}^{T}(r) = \frac{i\pi qr}{2c} lK_{l} \times \begin{cases} (r/a)^{l}/(l+1), & (r < a)\\ (a/r)^{l}/(l-1), & (r > a) \end{cases}$$
(78)

válida sólo para $l \ge 2$ (los casos l = 0, 1 se discutirán más adelante). Sustituyendo en la ec. (75) obtenemos el campo tangencial

$$E_{\theta l}^{T}(r) = \frac{i\pi qr}{2c} K_{l} \times \begin{cases} (r/a)^{l} (l+2)/(l+1), & r < a \\ -(a/r)^{l} (l-2)/(l-1) & r > a \end{cases}$$
(79)

Debemos añadir a las soluciones particulares (78) y (79) de las ecs. (74) y (75) las soluciones de las correspondientes ecuaciones homogéneas, proporcionales a r^{-l-1} y a r^{l-1} para r > a y r < a respectivamente. Los coeficientes de estos términos se obtienen de imponer condiciones de continuidad en r = a y conducen a la solución final

$$E_{rl}^{T}(r) = \frac{i\pi qr}{2c} \frac{K_{l}}{l^{2} - 1} \times \begin{cases} (r/a)^{l} (l(l-1) - (l-2)(l+1)(a/r)^{2}, & (r < a) \\ (a/r)^{l} (l(l+1) - (l+2)(l-1)(a/r)^{2}, & (r > a) \end{cases}$$
(80)

у

$$E_{\theta l}^{T}(r) = \frac{i\pi qr}{2c} \frac{K_{l}}{l^{2} - 1} \times \begin{cases} (r/a)^{l} ((l+2)(l-1) - (l-2)(l+1)(a/r)^{2}, & (r < a) \\ -(a/r)^{l} ((l-2)(l+1) - (l+2)(l-1)(a/r)^{2}), & (r > a) \\ (81) \end{cases}$$

para $l \geq 2$.

Evaluando estas expresiones en r = a y comparándolas con el campo longitudinal (11) y (13) observamos que el campo transversal es del orden de $(qa)^2$ veces más pequeño. Por lo tanto, en el límite de longitud de onda larga podermos despreciar al campo eléctrico transversal comparado el campo eléctrico longitudinal.

Para l = 1 y r < a las soluciones (80) y (81) no son válidas. Sin embargo, podemos seguir un procedimiento similar al anterior para resolver las ecs. (74) y (75) para hallar

$$E_{r1}^{T}(r) = -\frac{i\pi qa}{4c} K_1 \times \begin{cases} -(r/a)^2, & (r < a) \\ 4(\log(r/a) - 1) + 3(a/r)^2, & (r > a) \end{cases}$$
(82)

у

$$E_{r1}^{T}(r) = -\frac{i\pi qa}{4c} K_1 \times \begin{cases} -3(r/a)^2, & (r < a) \\ 4\log(r/a) - 3(a/r)^2, & (r > a) \end{cases}$$
(83)

Notamos que ahora aparece un término que diverge cuando $r \to \infty$, aunque muy lentamente, por lo cual no afectaría los cálculos en un metamaterial con una celda unitaria finita. A esta solución se le puede añadir un campo constante sin afectar ni las ecuaciones diferenciales ni las condiciones de continuidad. Al igual que los casos con $l \ge 2$, el campo transversal con l = 1 evaluado en la superficie r = a del cilindro es $(qa)^2$ veces menor que el correspondiente término longitudinal y es por tanto despreciable en el límite de longitud de onda larga.

Finalmente, en el caso l = 0 no hay campo longitudinal y el campo transversal fue evaluado explícitamente en el texto, ec. (22).

B. Apéndice: Función escalón

Expandendo en cosenos la función escalón (23), (24) y la corriente superficial (1), podemos expresar el hecho de que la corriente no fluye en la apertura $\Theta(\theta)K(\theta) = 0$ como

$$0 = \frac{1}{2} \sum_{n,m=0}^{\infty} \Theta_n K_m \cos((n+m)\theta) + \frac{1}{2} \sum_{n,m=0}^{\infty} \cos((n-m)\theta).$$
(84)

usando identidades trigonométricas para el producto de cosenos. Podemos reescribir la primera suma como

$$\sum_{n,m=0}^{\infty} \Theta_n K_m \cos((n+m)\theta) = \sum_{l=0}^{\infty} \sum_{m=0}^{l} \Theta_{l-m} K_m \cos(l\theta)$$
(85)

introduciendo el índice l = n + m. Para la segunda suma de la ec. (84) podemos tratar por separado los casos n > m,

$$\sum_{n=1}^{\infty} \sum_{m=0}^{n-1} \Theta_n K_m \cos((n-m)\theta) = \sum_{l=1}^{\infty} \sum_{m=0}^{\infty} \Theta_{l+m} K_m \cos(l\theta),$$
(86)

n = m,

$$\sum_{n=0}^{\infty} \sum_{m=n}^{n} \Theta_n K_m \cos((n-m)\theta) = \sum_{n=0}^{\infty} \Theta_n K_n,$$
(87)

y n < m

$$\sum_{n=0}^{\infty} \sum_{m=n+1}^{\infty} \Theta_n K_m \cos((n-m)\theta) = \sum_{l=1}^{\infty} \sum_{m=l}^{\infty} \Theta_{m-l} K_m \cos(l\theta), \qquad (88)$$

donde hicimos los cambios de variable l = n - m, l = 0 y l = m - n respectivamente. Sustituyendo las ecs. (85)-(88) en la ec. (84) y usando la independencia lineal de las funciones trigonométricas $\cos l\theta$, l = 0, 1, 2... podemos reescribir la Ec. (84) como un sistema de ecuaciones

$$0 = \sum_{m=0}^{\infty} \Theta_{lm} K_m, \quad (l = 0, 1, 2...)$$
(89)

donde la matriz Θ_{lm} se relaciona con los coeficientes de Fourier Θ_l de $\Theta(\theta)$ mediante la ec. (27). Este resultado es la versión para series de cosenos del conocido teorema de la convolución.

Referencias

- [1] David R Smith, John B Pendry, and Mike CK Wiltshire. Metamaterials and negative refractive index. *Science*, 305(5685):788–792, 2004.
- [2] Richard A Shelby, David R Smith, and Seldon Schultz. Experimental verification of a negative index of refraction. *Science*, 292(5514):77–79, 2001.
- [3] D. R. Smith, Willie J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz. Composite medium with simultaneously negative permeability and permittivity. *Phys. Rev. Lett.*, 84:4184–4187, May 2000.
- [4] Viktor G Veselago. The electrodynamics of substances with simultaneously negative values of ϵ and μ . Soviet physics uspekhi, 10(4):509, 1968.

- [5] David R. Smith and Norman Kroll. Negative refractive index in left-handed materials. *Phys. Rev. Lett.*, 85:2933–2936, Oct 2000.
- [6] J. B. Pendry. Negative refraction makes a perfect lens. *Phys. Rev. Lett.*, 85:3966–3969, Oct 2000.
- [7] Dylan Lu and Zhaowei Liu. Hyperlenses and metalenses for far-field superresolution imaging. *Nature Communications*, 3:1205, 2012.
- [8] S O'brien and JB Pendry. Magnetic activity at infrared frequencies in structured metallic photonic crystals. *Journal of Physics: Condensed Matter*, 14(25):6383, 2002.
- [9] John B Pendry, A J_ Holden, DJ Robbins, and WJ Stewart. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Transactions* on Microwave Theory and Techniques, 47(11):2075–2084, 1999.
- [10] Koray Aydin, Irfan Bulu, Kaan Guven, Maria Kafesaki, Costas M Soukoulis, and Ekmel Ozbay. Investigation of magnetic resonances for different split-ring resonator parameters and designs. *New Journal of Physics*, 7(1):168, 2005.
- [11] R. A. Shelby, D. R. Smith, and S. Schultz. Experimental verification of a negative index of refraction. *Science*, 292(5514):77–79, 2001.
- [12] Stefan Linden, Christian Enkrich, Martin Wegener, Jiangfeng Zhou, Thomas Koschny, and Costas M. Soukoulis. Magnetic response of metamaterials at 100 terahertz. *Science*, 306(5700):1351–1353, 2004.
- [13] Bruno Gompf, Barbara Krausz, Bettina Frank, and Martin Dressel. kdependent optics of nanostructures: Spatial dispersion of metallic nanorings and split-ring resonators. *Physical Review B*, 86(7):075462, 2012.
- [14] Vladimir M Agranovich and Yu N Gartstein. Spatial dispersion and negative refraction of light. *Physics-Uspekhi*, 49(10):1029, 2006.
- [15] VM Agranovich and Yu N Gartstein. Electrodynamics of metamaterials and the landau–lifshitz approach to the magnetic permeability. *Metamaterials*, 3(1):1–9, 2009.
- [16] W Luis Mochán, Guillermo P Ortiz, and Bernardo S Mendoza. Efficient homogenization procedure for the calculation of optical properties of 3d nanostructured composites. *Optics Express*, 18(21):22119–22127, 2010.
- [17] Bernardo S Mendoza and W Luis Mochán. Birefringent nanostructured composite materials. *Physical Review B*, 85(12):125418, 2012.
- [18] Ernesto Cortés, Luis Mochán, Bernardo S Mendoza, and Guillermo P Ortiz. Optical properties of nanostructured metamaterials. *Physica Status Solidi* (b), 247(8):2102–2107, 2010.

- [19] K. Glazebrook and F. Economou, *Pdl: The perl data language*, Dr. Dobb's Journal.
- [20] K. Glazebrook and J. Brinchmann and J. Cerney and C. DeForest and D. Hunt and T. Jenness and T. Luka and R. Schwebel and C. Soeller, *The perl data language v2.4.4*.
- [21] W. Luis Mochán, Guillermo Ortiz, Bernardo S. Mendoza, and José Samuel Pérez-Huerta. Photonic. Comprehensive Perl Archive Network (CPAN), 2016. Perl package for calculations on metamaterials and photonic structures.

La exploración del cosmos y los telescopios espaciales¹

Gloria Koenigsberger

Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Ave. Universidad S/N, Cuernavaca, Morelos, 62210, México

ABSTRACT

Se presenta una introducción general al tema de la Ciencia y la Tecnología y Espaciales, enfocada a las cuatro razones principales por las cuales la exploración espacial es importante: 1) la búsqueda del conocimiento, por el conocimiento mismo; 2) la seguridad nacional; 3) los beneficios comerciales; y 4) la sobrevivencia de la humanidad a largo plazo. Se incluye una recopilación de datos históricos sobre las sondas espaciales lanzadas por diversos países, así como datos sobre algunos de los riesgos involucrados en los vuelos tripulados.

Subject headings: Industria aeroespacial; estudios espaciales; telescopios espaciales

1. Introducción

El espacio interplanetario ha sido hasta hace poco tiempo una frontera inpenetrable para los seres humanos. A pesar de los muchos vehículos espaciales que han sido construidos por hombres y mujeres en la Tierra, ninguno ha transportado a un ser humano mas allá de la Luna, y las últimas pisadas del *Homo Sapiens* en este satélite natural van a cumplir medio siglo. En cuanto a vehículos no tripulados, el artefacto terrícola que mayor distancia ha recorrido es el *Voyager-1*, lanzado a finales de los años 1970's y que apenas recientemente ha entrado al espacio interestelar² Sin embargo, los avances tecnológicos y los intereses comerciales actuales comienzan a ofrecer la posibilidad de que, en un futuro cercano, finalmente viajemos nuevamente los humanos a alguno de los cuerpos celestes vecinos.

¹Este texto fue redactado y formado en LaTeX por G. Koenigsberger utilizando la clase aastex.cls de la AAS y el paquete graphicx. Cualquier error es responsabilidad de la autora.

²Véase *http://voyager.jpl.nasa.gov/where/* en donde se monitoréa día con día el avance de este artefacto y su "gemelo", el *Voyager-2*.

Por otro lado, ante la imposibilidad de explorar el cosmos fisicamente, hemos construido observatorios astronómicos que nos han permitido estudiar, no solo nuestro sistema solar, sino también estrellas y galaxias situadas a distancias inimaginables. Podemos caracterizar las últimas cuatro décadas como la Era de los Descubrimientos Astronómicos, por los fenómenos nuevos que se han observado, aunados al modelaje con simulaciones numéricas que permite explicar los procesos físicos observados.

Muchos de los nuevos descubrimientos astronómicos se deben a los telescopios espaciales. Estos son sistemas robóticos semi-autónomos, colocados en órbita alrededor de la Tierra a distancias suficientemente grandes como para eliminar los efectos limitantes de la atmósfera. Estos efectos son: 1) la turbulencia atmosférica, la cual degrada enormemente la resolución espacial, impidiéndo así ver muchos de los detalles en las imágenes; y 2) la opacidad a radiación en longitudes de onda del ultravioleta lejano, rayos-X y rayos- γ que hace imposible observar el cosmos desde la sueperficie terrestre en estas longitudes de onda.

Entre los ejemplos de los descubrimientos mas impactantes se destacan los discos protoplanetarios en la Nebulosa de Orión, en donde se pudo comprobar la teoría que predice la forma en que nacen las estrellas de baja masa acompañados de sus sistemas planetarios. También, se detectaron por primera vez una multitud de objetos que emiten enormes cantidades de rayos-X y rayos- γ , delatando la presencia de fenómenos de aceleración de material a velocidades imposibles de alcanzar en la Tierra, salvo quizás en los aceleradores mas potentes. Estos fenómenos se asocian a la presencia de estrellas de neutrones y hoyos negros, objetos predichos teóricamente pero que tuvieron que esperar a los observatorios espaciales a ser detectados. Cabe también destacar que el Sol también es una fuente de rayos-X y de luz ultravioleta, radiación que se origina en las zonas externas (llamadas *cromósfera* y *corona*). El poder observar la actividad solar en estas longitudes de onda ha permitido conocer mejor los efectos de la interacción del viento solar con el campo magnético de la Tierra.

Los telescopios espaciales se han apuntado también hacia nuestro propio planeta, La Tierra. Por ejemplo, una de las últimas generaciones de observatorios en órbita (llamado *GOES-16*) puede obtener una imágen de alta resolución espacial de todo el hemisferio terrestre cada 15 minutos en 16 bandas de energía. Esto permite trazar con gran detalle los movimientos y temperaturas de nubes y predecir los estados del clima a corto plazo.

En este capítulo presentamos una introducción al tema de la exploración del espacio. Esta incluye datos fríos y descripciones de la tecnología. Sin embargo, también mencionamos algunos de los principios básicos que deben regir todos los aspectos de esta empresa basándonos en algunos ejemplos históricos.



Fig. 1.— Concepción artística de la captura de un asteroide para poderle extraer metales y otros elementos y compuesto, o bien desviarlo de una órbita que lo podría llevar a colisionar con la Tierra.



Fig. 2.— Imágenes del Sol captadas en tres longitudes de onda: 304 Å, 171 Å, y en rayos-X. Los flujos en las diferentes energías son producidos por distintos procesos físicos. Por ejemplo, los rayos-X se deben a la aceleración de partículas en la región de la Corona solar, razón por la cual esta emisión aparece tan lejos de la superficie del astro. Créditos: NASA. Ver https://stereo.gsfc.nasa.gov/classroom/EUVsun.shtml para una explicación de los procesos visibles en este tipo de imágenes.

2. Porqué explorar el cosmos?

Hay al menos cuatro razones que han motivado la exploración espacial y que la siguen motivando, hoy mas que nunca.

2.1. El conocimiento por el conocimiento mismo

La primera y mas fundamental de todas es la necesidad que ha impulsado al ser humano a siempre ampliar las fronteras de su conocimiento y explorar lo desconocido. Aunque en gran medida esta necesidad encuentra su orígen en la lucha por la sobrevivencia, una parte de ella nace simplemente de la curiosidad por saber qué es lo que yace mas allá del territorio conocido. Esta es la búsqueda del conocimiento por el conocimiento mismo, y generalmente se refiere a la investigación que se efectúa en temas que no parecieran tener una aplicación "útil" para la sociedad contemporánea. Es importante remarcar, sin embargo, que un gran número de los avances tecnológicos que definen al inicio del Siglo 21 son resultado de la investigación efectuada en temas a los que no se les veía en su tiempo ninguna utilidad práctica.

2.2. Seguridad Nacional

Un segundo objetivo de la exploración espacial es la seguridad nacional. Observar a la Tierra desde arriba de su atmósfera proporciona una vista panorámica de los estados meteorológicos, movimientos de vehículos, incendios, entre otros. Además, existen redes de satélites que permiten el uso eficiento de los *GPS (Global Posicioning Sensors)* y satélites con detectores para la radiación infrarroja que permiten identificar la ubicación de seres humanos, manadas de animales, fallas geológicas, rupturas en gasoductos y oleoductos. No es coincidencia que los programas espaciales mas avanzados pertenecen a los países que poseen las potencias militares mas poderosas del mundo.



Fig. 3.— Distribución espectral de la radiación del Sol que llega a la distancia de la órbita de la Tierra (en amarillo) y la que logra propagarse por la atmósfera de la Tierra y llegar a la superficie (en rojo). Nótese que la cantidad de energía que llega a la superficie terrestre en la región ultravioleta del espectro es casi nula, así como en las bandas moleculares de H_2O y CO_2 . Es de notarse también que estas bandas moleculares son responsables, en gran medida, del efecto "invernadero" que mantiene a la superficie terrestre mas caliente de lo que sería si no tuviera una atmósfera con esta composición molecular.

2.3. Intereses comerciales

Un tercer objetivo es úno de interés comercial. La industria de las telecomunicaciones, fue de las primeras en ver los grandes beneficios de poseer satélites en órbita alrededor de la Tierra. Con los grandes avances tecnológicos, se prevee la explotación de cuerpos rocosos de nuestro sistema solar (asteroides, lunas, planetas) de los cuales se podrán extraer diversos minerales y metales. Se destacan entre éstos los elementos llamados "tierras raras", varios de los cuales son requisitos indispensables para el funcionamiento de los teléfonos celulares, las memorias de las computadoras, las baterías recargables y los aparatos de diagnóstico médico, entre otros. En la Tierra, estos elementos se encuentran en lugares muy limitados (principalmente en Rusia, China y EUA) y cada vez son mas escasos. Por ello, la posibilidad de extraerlos en cantidades mas abundantes de los cuerpos rocosos de nuestro sistema solar le ha dado un gran impulso a lo que ahora se llama *minería espacial*.

Nótese que el costo de los artefactos que son enviados al espacio es una fracción ínfima de la inversión que se hace en su fabricación. Es decir, la mayoría de los millones de dólares que cuesta una misión espacial se queda en la Tierra y le da un gran impulso al sector aeroespacial así como al desarrollo de nuevos materiales, computadoras, componentes electrónicas, y mucho mas.

2.4. Sobrevivencia de la humanidad

La cuarta razón para invertir en el desarrollo de la tecnología espacial es la sobrevivencia de la humanidad. Hay varios niveles de importancia de este rubro, pero destaquemos ahora el riesgo que implicaría la caída de otro objeto celeste a la Tierra, como el que la impactó hace aproximadamente 69 millones de años, noindent evento que ocasionó la extinción de los dinosaurios, entre muchas otras especies. Se sabe que el impacto ocurrió en el Golfo de México (el cráter que dejó se llama Chicxulub³ Existen hoy en día un número importante de observatorios dedicados a la detección y al monitoreo de los cuerpos celestes (en su mayoría asteroides) cuyas órbitas podrían representar un riesgo para la Tierra. También están en desarrollo tecnologías para contender con uno de estos cuerpos que se comenzara a aproximar demasiado a nuestro planeta. Algunas de estas tecnologías son semejantes a las que se planean para la minería de asteroides.

En otro orden de ideas, hay diseños para enormes estructuras espaciales que podrían ser utilizadas como ciudades en órbita, o bien como un especie de "parasol" para eclipsar una parte de la radiación solar que llega a la Tierra y de esta manera mitigar los efectos noscivos del calentamiento global.

Finalmente, cabe destacar que el Sol ha sido una estrella relativamente pasiva desde hace varios millones de años, lo cual ha repercutido en una relativa estabilidad para las

condiciones en la Tierra. Si el Sol entrara a una etapa de mucha mas actividad, como la que se observa en otras estrellas similares a nuestro astro, el bombardéo por rayos cósmicos aumentaría apreciablemente lo cual, aunado a una posible inversión del campo magnético terrestre (que nos protege de los rayos cósmicos), tendría resultados catastróficos para la vida en la Tierra. Este tipo de consideraciones, aunadas a las de la evolución geológica de nuestro planeta, llevan a la necesidad de planear la colonización de otros de los posibles *habitat* en nuestro Sistema Solar y mas allá.



Fig. 4.— Ilustración de varios modelos de cohetes con sus respectivas cargas. De izquierda a derecha: Saturno V, Space Transport System (STS; Shuttle), Ares I, Ares V, Ares IV, Space Launch System (SLS) Block I y SLS Block II. La familia Ares nunca se construyó y fue sustituida mas bien por la famila SLS. El Saturno V se lanzaba con unicamente combustible líquido. Los cohetes modernos de gran potencia utilizan, además, combustible sólido (los tubos blancos que se encuentran fijados a los costados del tanque de combustible líquido, éste último contenido en los cilindros rojos-anaranjados mostrados en la figura). Dibujo obtenido de https://en.wikipedia.org/w/index.php?curid=48782478.

3. Categorías de vehículos espaciales

Se pueden catalogar los vehículos espaciales dentro de dos grandes rubros: 1) el vehículo que contiene el equipo, los instrumentos y la tripulación, los cuales están siendo enviados al espacio para desarrollar alguna o algunas misiones específicas, comunmente llamado *carga útil*; y 2) el vehículo que sirve para llevar la carga útil desde la superficie terrestre hasta su destino, llamado *cohete* o *lanzador*.

3.1. Cohetes/lanzadores

Todos los lanzadores se basan en el diseño original de los cohetes alemanes de los años 1940s, cuyo objetivo era portar explosivos como carga útil, con fines bélicos. Las variaciones de este diseño han surgido gracias al desarrollo de mejores y mas eficientes combustibles, incluyendo los combustibles sólidos.

El tipo de lanzador depende del peso que éste deberá llevar a órbita y la distancia que deberá recorrer. Las cargas mas pesadas o las que han sido transportadas mas lejos han requerido de los lanzadores mas poderosos. Historicamente, el mas poderoso jamás construido fue el *Saturno-V*, utilizado para encaminar a las naves *Apollo* con sus tres tripulantes hacia la Luna en los años 1960's y 1970s. Hoy en día, tanto EUA como Rusia y China tienen en desarrollo lanzadores modernos con una potencia que se asemeja a la de los *Saturno-V*.

La Tabla⁴ 1 muestra una recopilación parcial de las familias de cohetes/lanzadores que han sido desarrollados por los mas de 15 países que han invertido esfuerzo y recursos en programas espaciales y que siguen activos. Las columnas 3-5 indican el peso que el cohete puede llevar hasta: $LEO=low \ earth \ orbit$, una órbita baja, aproximadamente 500 km; GTO: Geostationary transfer orbit, una órbita eccéntrica con apogeo a una distancia de ~42000 km de la Tierra; y TLI: trans lunar injection, lo cual implica que la primera etapa del cohete lleva primero al vehículo a una orbita alrededor de la Tierra desde donde se encienden los cohetes de una segunda etapa del cohete que eleva al vehículo a una órbita eccéntrica con apogeo a la distancia de la Luna. La columna 6 lista el costo (en millones de USD). La columna 7 indica el número de veces que un cohetes de ese modelo que ha sido lanzado y la 8 los lanzamientos que fueron exitosos. La tabla se enfoca primariamente a los lanzadores aún activos en 2014 e indica el año en que se llevó a cabo su primer lanzamiento. Hemos añadido 3 modelos muy exitosos que ya han sido descontinuados, en cual caso se lista también el año del último lanzamiento. El anexo contiene una tabla similar mas completa, que incluye muchas de las familias ya descontinuadas.

3.2. Carga útil: tipos de aplicaciones

Las Tablas 2 y 3 listan la clase de objetivo al que pertenecen los satélites y sondas que han sido lanzados desde 1944. Se puede observar que el porcentaje mas alto de las misiones corresponden a los rubros de Seguridad Nacional y de Intereses Comerciales.

⁴Datos tomados de https://en.wikipedia.org/wiki/Comparison_of_orbital_launchers_families



Fig. 5.— Dibujo del cohete que fue diseñado por los alemanes durante la Segunda Guerra Mundial y que sirvió de base para todos los diseños posteriores de cohetes.

3.3. Satélites astronómicos

Los primeros vehículos portadores de equipo astronómico se lanzaron en los años 1960's. Inicialmente eran detectores de luz y partículas muy primitivos, pero con el avance tecnológico, fueron convirtiéndose en verdaderos observatorios. Es decir, estaban dotados de

Familia	País	Carga Máxima		Costo	Num	Num	Año	
		LEO	GTO	TLI	M-USD	Total	Exitos	$1^{ero}-$
Angara A3,A5	Russia	35000	12500		na	1	1	2014 -
Ariane 5	Europe	21000	9600		220	76	74	2002 -
Atlas V	USA	18850	8900	2807	na	30	30	2002 -
Delta IV	USA	23040	13130	9000	na	18	18	2002 -
Falcon 1&1e	USA	420			7.9	5	4	2006 -
Falcon 9	USA	13150	4850		61.2	22	21	2010 -
H-II, IIA&IIB	Japan	19000	8000		na	28	26	1994 -
Long March 2-3-4	China	12000	5500	3300	na	167	158	1971 -
Minotaur IV&V	USA	1735	640	447	50	4	4	2010 -
Naro	S. Korea	100			na	3	2	2009 -
Pegasus	USA	450			56.3	42	37	1990 -
UR-500 Proton	USSR	23000	6920	5680	na	399	353	1965 -
PSLV	India	3800	1300		na	34	32	1993 -
UR-100N	Russia	2100				25	23	1994 -
Safir	Iran	50			na	6	4	2007 -
Shavit	Israel	225			15	9	7	1988 -
R-7 Soyuz	USSR	5500	2400	1200	na	1849		1957
Taurus	USA	1450			na	9	9	1989 -
R-36 Tsyklon	USSR	4100			na	259	_	1967 -
Unha	N. Korea	100			na	3	1	2006 -
VLS-1	Brazil	380			na	2	0	1997 -
Zenit	USSR	13740	6160	4098	na	82	71	1985 -
Ariane 6^*	Europe		10500		115	_		2020 -
Falcon Heavy	USA	53000	21200		80 - 125	0		_
GSLV Mk.III	India	8000	4000		na	1	1	2014 -
Saturn V	USA	118000		47000	185	13	13	1967 - 1973
Space Shuttle	USA	24400	3810		450	135	134	1981 - 2011
Titan I-II-III-IV	USA	21900	5773	8600	350	369	_	1959 - 2005

Table 1: Familias de cohetes (selección)

Esta tabla lista una selección de las familias de cohetes por país, carga, costo, número de lanzamientos y año de primer-último lanzamiento.

un telescopio, detectores avanzados (por ejemplo, CCDs de 2048-2048 pixeles o mas), sistemas de posicionamiento y guiado, computadoras de abordo capaces de almacenar grandes volúmenes de información, y sistemas de telecomunicaciones de banda ancha capaces de transmitir estos datos a la Tierra.

Los tres observatorios astronómicos espaciales mas complejos y, que mayor impacto científico tuvieron hasta principios del Siglo 21 fueron: 1) el Hubble Space Telescope (HST), 2) el CHANDRA X-ray Observatory, y 3) el SWIFT gamma Ray Observatory. En conjunto,

Aplicación	A partir de
	Año
Militares (armamentos, misiles)	1944
Investigación Científica	1947
Vigilancia y espionaje	$\sim \! 1960$
Telecomunicaciones	1958
Vuelos tripulados por humanos	1961
Navegación (GPS)	1978
Ensamblaje de estaciones espaciales (ISS)	1998
Turismo	2001
Minería de asteroides	2020

Table 2: Objetivos de las cargas útil

Table 3: Uso de los 6581 satélites/sondas lanzados hasta 2007

Comunicaciones	1814	27.6	militar y civil
Monitoréo, Espionaje	1775	27.0	militar
Navegación, servicios	547	8.3	militar
Investigación y desarrollo	540	8.2	militar
Programa tripulado humano	507	7.7	civil:
Investigación y desarrollo	365	5.6	civil
Estudio Tierra y Espacio	351	5.5	civil
Percepción remota Tierra	345	5.2	civil
Sistema Solar y Universo	323	4.9	civil

estos observatorios tenían la capacidad de efectuar observaciones en longitudes de onda que van desde los Rayos- γ hasta el infrarrojo lejano.

El HST fue colocado en órbita por el único modelo de nave espacial que ha tenido características totalmente distintas a los vehículos que han portado cargas útiles al espacio; es decir, fue colocado por el STS-Space Transport System (Shuttle)=Transbordador Espacial, el primer (y único hasta recientemente que el Falcon 9 ha logrado aterrizar) vehículo espacial re-utilizable.

Las vidas del HST y del Transbordador Espacial estuvieron estrechamente ligadas desde que fue diseñado el HST y hasta la preparación para su destrucción, que ocurrirá al reingresar de manera controlada por la atmósfera terrestre y quemarse.



Fig. 6.— **Izquierda:** Lanzamiento número 120 de uno de los 5 Transbordadores Espaciales (Challenger, Columbia, Atlantis, Endeavour, Discovery). **Centro:** Diagrama que muestra las maniobras efectuadas, desde despegue hasta aterrizaje. **Derecha** El *Transbordador Atlantis* siendo transportado sobre un avión Boeing 747 que fue especialmente adaptado para llevar a cabo esta tarea.

4. Viajar al espacio es muy riesgoso

La Era Espacial inició oficialmente a finales de 1957 cuando la Unión Soviética lanzó al Espacio el primer satélite artificial de la humanidad, llamado *Sputnik I*. A partir de entonces y hasta la fecha, se han lanzado miles de artefactos, satélites y sondas al espacio. Cabe destacarse que el trabajo previo que condujo a la posibilidad de lanzar vehículos mas allá de la atmósfera terrestre, inició en los años 1900's con el trabajo teórico de Konstantin Tsiokovsky (1903) y Hermann Oberth (1922), seguido de los lanzamientos de cohetes impulsados por combustible líquido de Goddard (1926) y continuó en los años 1940s con los avances logrados principalmente por científicos e ingenieros alemanes, durante la Segunda Guerra Mundial. De hecho, Alemania estuvo a punto de tener misiles, con los cuales muy probablemente habría ganado la guerra. Con la derrota de Alemania, sus científicos e ingenieros fueron repartidos entre los EUA y la entonces URSS, dándole un gran auge a sus programas espaciales respectivos.

En esta sección nos enfocaremos en forma muy resumida al tema del riesgo que conlleva un viaje al espacio.

Año	Suceso		
1903	Trabajo Teórico de K. Tsiokovsky		
1922	Trabajo Teórico de H. Oberth		
1926	Experimentos con combustibles líquidos de Goddard		
1930-40s	Desarrollo de los cohetes V2 Alemanes; WWII; W. von Braun		
$1950\text{-}60\mathrm{s}$	Desarrollo tecnología espacial S. Korolev		
1957	URSS lanza el Sputnik I y lo coloca en órbita		
1961	EUA lanza al espacio a Shephard		
1961	URSS lanza al espacio a Y. Gagarin (1 órbita)		
1962	Primer hombre que orbita la Tierra, J. Glenn		
1963	Primera mujer en el espacio: Valentina Tereshkova		
1967	Primeros 2 hombres en pisar la Luna (Apollo 11): Neil Armstrong y Buzz Aldrin		
1967	Primer deceso de un humano durante un vuelo espacial: V. Komarov		
1972	Últimos 2 hombres en pisar la Luna (Apollo 17): E. Cernan y H. Schmitt		
1981	Primer vuelo espacial de un vehículo re-utilizable: Space Shuttle Columbia (EUA)		
1986-96	Ensamblaje de la primera estación espacial (MIR) URSS		
1998	Inicia el ensamblaje de la Estación Espacial Internacional países		
2010	Primer vuelo espacial operado por una empresa comercial (Falcon9/SpaceX)		
2015	Primer aterrizaje vertical de un cohete $(1^a \text{ etapa del Falcon 9})$		

Table 4: Datos históricos

4.1. Viajes tripulados

El Sputnik I fue seguido poco después por el Sputkik II que llevó al Espacio a Laika, canino que se convirtió así en la primera tripulante terrícola en ser puesta en órbita alrededor de la Tierra y, lamentablemente, también fue la primera en morir ahí. La causa de muerte fue sobre-calentamiento de la cápsula en la que viajaba, pero esta falla simplemente precipitó un desenlace esperado: El regreso a la Tierra de Laika no estaba contemplado en los planes de vuelo ya que la tecnología necesaria para traer a un tripulante de regreso no se había aun desarrollado.

La justificación que se utilizó para haber lanzado al espacio un ser vivo en estas condiciones fue que en aquella época había dudas sobre la sobrevivencia de seres vivos en las condiciones de microgravedad, así como las condiciones de muy alta gravedad durante el despegue. Sin embargo, se cree que Laika falleció poco después del lanzamiento dado que es entonces que ocurrió la falla que llevó al sobrecalentamiento de la cápsula. Esta información se mantuvo oculta hasta muchos años después. De hecho, fue hasta 1998 (después del colapso de régimen Soviético) que uno de los científicos responsables, Oleg Gazenko, expresó su remordimiento por haber permitido el sacrificio, además de que por su muerte prematura no hubo respuestas a las preguntas sobre los efectos de la microgravedad sobre la vida.

Año	Tripulante	Sobrevivió?	Comentario
1947	moscas	sí	cohete V-2; descenso c/ paracaidas de cápsula
1948	simio (Albert I)	(no)	anestesiado; falla del cohete al despegar
1949	simio (2;Albert II y IV)	no	falla de paracaidas en descenso
1950	ratón	no	falla de paracaidas en descenso
1951	simio+11 ratones	sí	Yorick
1952	2 simios + 2 ratones	sí	misma misión
1958	ratón (3)	no	3 misiones distintas del Programa MIA
1958	simio (Gordo)	no	falla del paracaidas
1959	$2 \ simios$	sí	1 falleció posteriormente en
	(Able& Baker)		la cirujía para retirar sensores
1959	4 ratones	(no)	se comieron material tóxico antes del despegue
1959	4 ratones	no	explotó el cohete
1959	Simio (SAM)	sí	vivió muchos años después
1960	Simio (Miss-SAM)	sí	vivió muchos años después
1961	Simio (HAM)	sí	vivió muchos años después
1961	Simio (Goliath)	no	
1961	Simio	sí	orbitó la Tierra; misión abortada a la 2a órbita
1961	Alan Shephard	sí	Mercury-Redstone 3
1961	Gus Grissom	sí	Mercury-Redstone 4; escotilla voló en amarizaje
1962	John Glenn	sí	orbitó la Tierra; Mercury-Redstone 4
1963	Gordon Cooper	sí	mercury-Atlas 9
1966	Geminis 8-12	sí	5 misiones (10 humanos)
1967	Apollo 1	(no)	enero: 3 humanos fallecen durante pruebas en tierra
SUBTOT	hasta 1967		$20/55{\sim}36\%$ probabilidad de muerte
1968	Apolo 7-8	sí	2 misiones (6 humanos)
1969	simo	(si)	regreso después de 9 días en órbita con
			signos vitales deteriorantes; murió 8 hrs después de
			su retorno por deshidratación
1969	Apollo 9-12	sí	4 misiones (12 humanos)
1970	Apollo 13	(si)	falla de sistemas camino a la Luna
1971	Apollo 14,15,16,17	sí	4 misiones (12 humanos)
1982	STS3-5	sí	3 misiones (~ 15 humanos)
1983	STS6-9	sí	3 misiones (~ 15 humanos)
TOTAL			$19/116$ vidas; $\sim 16\%$ prob. de muerte

Table 5: Las primeras misiones espaciales tripuladas de USA

A partir de la década de los 1960, los Soviéticos tuvieron varios éxitos con su programa espacial bajo el mando de Sergey Korolev, quien se considera uno de los grandes pioneros
Año	Tripulante	Sobrevivió?	Comentario
1951-2	9 caninas	??	3 de ellas volaron 2 veces
1951	2 canino	sí	sub-orbital (Dezik&Tsyam)
1951	2 canina	no	Dezik & Lisa
1951	2 canina	sí	Smelaya & Malyshka
1951	2 canina	no	sin nombre
1951	2 canina	sí	Bubik & Zib
1951	2 canina	sí	sub-orbital
1957	canina (Laika)	no	sobre-calentamiento; poco después del despegue viaje programado sin retorno
1959	canina (Otvazhnaya)	sí	5 vuelos sub-orbitales
1960	2 caninas	no	Bars & Lisichka
1960	2 caninas	sí	18 órbitas
1960	2 caninas	no	Pchelka &Mushka
1960	2 caninas	(si)	lanzamiento falló pero ambas
			fueron rescatadas vivas
1960	126 humanos	(no)	R-16 ICBM explota en Baikonur durante preparativos de lanzamiento
1961	2 caninas	sí	Chernushka
1961	1 canina	sí	Zvevdoshska
1961	Yuri Gagarin	sí	Vostok 1; 1 órbita
1963	Valentina Tereshkova	sí	
1966	2 caninas	sí	22 días en órbita
1967	Vladimir Komarov	no	fallas en paneles solares, estabilidad, paracaidas
			sin contar explosión de 1960
TOTAL	hasta 1967		$10/30; \sim 33\%$ muerte
			sin contar explosión de 1960
	Francia		
1963	gato	sí	
1963	gato	no	recuperación tard'ıa de la cápsula
	Argentina		
1967	rata	sí	cohete Yarará
1969	simio	sí	cohete Canopus II
1970	simio	no	falla del paracaidas

Table 6: Las primeras misiones espaciales tripuladas de la USSR y otros países

de la Era Espacial. Él guió el programa de vuelos tripulados hasta su muerte en 1966 por causa del cancer. Lamentablemente, su sucesor, Vasily Mishin cometió algunos de los errores mas graves en la historia de los vuelos espaciales tripulados, todos ellos debido a

que antepuso criterios políticos a los criterios científicos. Específicamente, cedió ante las presiones políticas de "ganarle" a EUA a llegar a la Luna, y ordenó lanzar a un cosmonaúta, Vladimir Komarov, en el *Soyuz-1*, nave que se sabía tenía fallas de diseño. Aparentemente, los ingenieros trabajando sobre esta nave habían reportado mas de 200 problemas no-resueltos. Sin embargo, Komarov fue lanzado al espacio. Poco después se anunció que su nave había caído a la Tierra y que había muerto. También fracasaron cuatro intentos posteriores de lanzar cohetes diseñados para ir a la Luna (éstos sí, sin tripulación). Posteriormente, en 1971, fue puesta en órbita la primera estación espacial, llamada Salyut-1. Según cuentan las historias, para poder enviar 3 hombres en esta misión en vez de los 2 para los cuales la nave estaba diseñada, Mishin los envió sin sus trajes presurizados (para ahorrar peso). Esto tuvo consecuencias fatales ya que en el reingreso la cápsula en la que viajaban se despresurizó y Georgy Dobrovolsky, Vladislav Volkov, and Viktor Patsayev perecieron.

Las historias de arriba son destacables por varias razones. La primera cae en el ámbito de la valoración de la vida ante las necesidades impuestas por la investigación científica, en este caso, la exploración por humanos del espacio cósmico. La segunda es que se muestra que el precipitar intentar obtener un resultado, saltándose pasos intermedios, tiende a tener consecuencias catastróficas. La tercera muestra que cuando los criterios políticos se inmiscuyen y llegan a dominar en los procesos de desarrollo científico, la probabilidad de fracaso es grande.

Una razón por la que errores como los descritos arriba pudieron ocurrir es que en la URSS los esfuerzos asociados a proyectos espaciales se desarrollaban bajo distintas estructuras administrativas (la mayoría militares), y había poca coordinación entre ellas. No fue hasta 1974 que se creó una sola agencia para vuelos espaciales, *NGO Energia*, que ahora se conoce como *S.P. Korolev Rocket and Space Corporation Energia*.

En EUA, se consolidó el esfuerzo espacial desde un inicio en una sola agencia llamada National Aeronautic and Space Administration (NASA). Los tres primeros objetivos decretados⁵ para esta agencia, son: 1) La expansión del conocimiento científico sobre los fenómenos de la atmósfera y del Espacio; 2) La mejoría en la utilidad, funcionamiento, velocidad, seguridad y eficiencia de los vehículos aeronáuticos y espaciales; y 3) El desarrollo y la operación de vehículos con la capacidad de transportar instrumentos, equipo, provisiones y organismos vivos por el Espacio.

⁵National Aeronautics and Space Act of 1958, Public Law #85-568, 72 Stat., 426. Firmado por el Presidente el 29 de julio de 1958, Record Group 255, National Archives and Records Administration, Washington, D.C; disponible en: NASA Historical Reference Collection, History Office, NASA Headquarters, Washington, D.C.

Los proyectos espaciales tripulados de la *NASA* no estuvieron libres de fatalidades humanas, pero el primer desastre no ocurrió durante un viaje espacial, sino en la fase de preparativos. El *Apollo-1* sufrió un corto circuito en su interior, cuando estaban los 3 astronaútas efectuando pruebas con la compuerta de la cabina cerrada. El ambiente, que en esa época era casi puro oxígeno, propició que el incendio se propagara velozmente, calcinando a los 3 astronáutas.

El desastre del *Shuttle Challenger* se presentó cuando falló uno de los sellos de hule de uno de los cilindros de combustible sólido, provocando una explosión en cadena durante el despegue, matando a los 7 tripulantes.⁶ El último desastre fue el del *Shuttle Columbia* que sufrió una falla en su escudo térmico la cual, al calentarse durante el reingreso a la atmósfera de la Tierra, produjo la desintegración de la nave, matando así a sus 7 tripulantes.⁷

Década		La falla	es en:		Objetivo	Núm.	Exitosos	±s.d.
(Años)	Pista	Lanzam	Órbita	Final	incumplido	total		
1957-1966	3	146	18	25	39	933	72%	21%
1967 - 1976	2	101	22	6	24	1567	89%	4%
1977 - 1986	0	75	20	1	16	1574	92%	3%
1987 - 1996	0	57	16	0	16	1359	93%	3%
1997 - 2006	1	74	14	2	26	1159	90%	4%
2007-2011	0	10	1	1	1	629	96%	
Total	6	463	91	35	122	7221		
Porcentaje	0.8%	65.2%	12.4%	4.8%	16.8%		—	

Table 7: Estadísticas 1957–2014 de fallas

Número Total: suma de exitosos y fallidos; "pista": antes de ser lanzado; "órbita": la falla ocurre habiendo llegado al espacio; "misión incumplida": no cumple con la misión original por algún problema, aunque el lanzamiento y el aterrizaje hayan sido exitosos; "final": falla en reingreso, aterrizaje, recuperación. Tabla basada en datos de Claude LaFleur, Space Encyclopedia.

⁶Francis "Dick" Scobee, Michael Smith, Judith Resnik, Ellison Onizuka, Ronald McNair, Christa McAuliffe y Gregory Jarvis

⁷Rick D. Husband, William C. McCool, Michael P. Anderson, David M. Brown, Kalpana Chawla, Laurel Blair Salton Clark, e Ilan Ramon

4.2. Qué tan riesgoso es viajar al espacio?

Las estádisticas en la Tabla 7 muestran que el riesgo de falla en los vuelos espaciales ha ido disminuyendo conforme pasa el tiempo, lo cual es de esperarse, dado que cada fracaso ha llevado a aprender importantes lecciones de cómo evitar otro igual. Hoy en día, la probabilidad de que una misión sea exitosa es superior al 90%. También se vé que, en términos generales, los viajes tripulados (por humanos) han tenido una tasa de éxito mayor que el promedio. Esto también es natural, dado que se exigen márgenes de riesgo mucho mas pequeños en el caso de este tipo de viajes que cuando la pérdida, en caso de un percance, sea unicamente de un equipo. Evidentemente, estos márgenes de riesgo tan pequeños implican un costo mucho mas elevado de la misión, razón por la cual los viajes tripulados han sido relativamente escasos.

Las cifras actuales contrastan dramaticamente con las de la primera década de los vuelos espaciales, cuando la probabilidad de éxito era no mas que el $\sim 67\%$. Nótese que, cuando Alan Shephard se subió al cohete *Mercury-Redstone* que lo llevaría en un vuelo al espacio en 1961, las dos misiones previas habían fracasado, y en la primera de ellas el simio que iba a bordo había perecido. Realmente se requirió mucho valor para emprender un viaje al espacio bajo esas condiciones!

Al paso del tiempo, se han ido desarrollando normas muy estrictas que gobiernan el lanzamiento de naves al espacio, particularmente, normas enfocadas a la seguridad. El análisis de los riesgos es ahora un proceso muy complejo y detallado, pero que en el fondo evalua y asigna probabilidades de riesgo a las tres preguntas siguientes: 1) Qué puede fallar? 2) Qué tan probable es que ocurra esta falla? 3) Cuáles son las consecuencias asociadas a esta falla? Este análisis se efectúa pieza-por-pieza de todos los componentes que entran en la fabricación de los cohetes, vehículos espaciales y equipo de abordo.

5. El futuro: un universo de posibilidades

El panorama que se dibuja para el futuro es úno de grandes retos para la supervivencia de la vida como la conocemos pero, al mismo tiempo, de grandes oportunidades para continuar con los procesos de evolución y adaptación que la han caracterizado desde el inicio de los tiempos en la Tierra. El futuro ofrece aventuras y recompensas que apenas ahora podemos comenzar a vislumbrar. Entre ellas se encuentra, como tesoro inigualable, la esperanza de un futuro mejor, ya sea en la Tierra misma o en algún otro habitat. Tesoros mas tangibles que nos esperan en el Espacio incluyen las materias primas que se podrán extraer de asteroides y demás cuerpos de nuestro sistema solar, nuevos compuestos orgánicos y, quizás, hasta nuevas



Fig. 7.— Representación del Telescopio JWST en su configuración final, cuando comience a operar a finales de esta década.

formas de vida. Todo ello irá acompañado de e impulsado por un crecimiento explosivo del conocimiento y del desarrollo tecnológico.

Los avances desde la década de los 1960s han sido realmente espectaculares, pero los siguientes lo serán aún mas. En EUA y Europa han comenzado a surgir empresas del sector privado las cuales trabajan conjuntamente con la NASA y la European Space Agency (ESA) en el desarrollo y, desde hace ya un par de años, la operación de misiones espaciales. Por ejemplo, la empresa Space-X ha comenzado a efectuar los viajes de abastacimiento de provisiones a la Estación Espacial Internacional, en donde siempre hay astronáutas en residencia. Se espero que en fechas próximas, Space-X también comience a efectuar el traslado de los astronáutas.⁸. Space-X también ha tenido un éxito arrollador en fechas recientes al haber logrado que su cohete Falcon-9 regrese y aterrice verticalmente en la Tierra. Con ésto, se abre la posibilidad de re-utilizar estos lanzadores, reduciendo el costo y el tiempo entre lanzamientos.

Por otro lado, la NASA concretó recientemente el diseño del vehículo espacial *Orión* que se planea llevará a una tripulación de 4 seres humanos a misiones de larga duración y eventualmente a Marte.⁹ Habrá que ver si este plan no se vé modificado por los avances vertiginosos de *SpaceX* y su tecnología para traer de regreso de forma controlada y hacer

⁸Véase http://www.spacex.com/dragon y http://www.spacex.com/falcon9

 $^{^9}$ Véase https://www.nasa.gov/exploration/systems/orion/index.html

aterrizar suavemente (sin uso de paracaídas) sus naves espaciales. Si sigue esta empresa teniendo el éxito que hasta ahora ha tenido, es muy probable que la primera misión a Marte sea desarrollada por ella, utilizando sus naves espaciales. Una vez que haya naves seguras para viajar a Marte, los planes futuros preveen la colonización por seres humanos de este planeta.

Otra de las misiones de gran importancia que se avecinan es el aterrizaje de una nave robótica en la Luna Europa, la segunda luna mas masiva de Júpiter.¹⁰ Este cuerpo guarda en su interior un gran secreto que recientemente se ha ido revelando: bajo la capa de hielo superficial que vemos al observarla, esta luna esconde lo que parece ser un oceanos de agua (H_2O) , y no solo agua, pero agua salada. Hay una fuente de calor en el interior rocoso que permite derretir el hielo y mantenerlo en forma líquida. Los cálculos indican que la fuente principal de este calor es la variación en la fuerza gravitacional que sufre debida a la órbita excéntrica de Europa alrededor de Júpiter. Habiéndo un oceano de agua salada, la probabilidad de encontrar organismos vivos es muy grande.

En lo que se refiere a la exploración del Cosmos a distancias inal canzables para los seres humanos,¹¹ los avances son también espectaculares. En 2013 fue lanzado el observatorio $Gaia^{12}$ el cual ha comenzado ya a operar y enviar datos a la Tierra sobre millones y millones de estrellas.

Está previsto que el Telescopio James Webb será lanzado en 2018 el cual, al igual que Gaia, se colocará en una órbita alrededor del punto L2 de Lagrange; es decir, entre la Tierra y el Sol. Se espera que con este telescopio sea posible observar y estudiar los primeros objetos cósmicos que se formaron en nuestro Universo después del "Big Bang".

 $^{^{10}} V\'ease\ https://www.nasa.gov/press-release/all-systems-go-for-nasas-mission-to-jupiter-moon-europain$

¹¹La exploración por seres humanos e inclusive sondas robóticas de nuestra Galaxia y mas allá es inpensable ahora, a menos de que se pueda violar el principio fundamental de viajar a una velocidad máxima cercana a la de la luz.

 $^{^{12}}$ http://sci.esa.int/gaia/



Fig. 8.— El telescopio espacial *James Webb Space Telescope* será lanzado por un cohete *Arianne* antes de finalizar esta década. El espejo mide 6*m* de diámetro y está constituido por segmentos hexagonales. Para poderlo meter en el módulo de carga es necesario doblarlo, como se ilustra en la figura de la derecha. Una vez en su destino final, iniciará el proceso de des-doblar el espejo, antenas y páneles solares, extender su base protectora de la radiación solar, y echar a andar todos los sistemas para entrar en operación. La configuración final se muestra en la Figura 7.



Fig. 9.— El telescopio James Webb será colocado en una órbita alrededor del punto de equilibrio L2 de Lagrange, en el sistema Tierra-Sol. El JWST será el segundo observatorio astronómico colocado en esa posición, ya que actualmente se encuentra ahí ya en operación el Telescopio Gaia.

6. Referencias

- "The Vision for Space Exploration", NASA, February 2004, https://www.nasa.gov/pdf/55583main_vision_space_exploration2.pdf, consultado 16 de abril de 2016
- 2. Para estadísticas globales consultar: http://claudelafleur.qc.ca/Spacecrafts-index.html#Countries y http://claudelafleur.qc.ca/Spacecrafts-index.html#Table-1
- 3. "The Story of Laika", *moscowanimals.org*, archived from the original on August 16, 2006, retrieved, 26 September 2006
- 4. "Message from the First Dog in Space Received 45 Years Too Late", Dogs in the News, 3 November 2002, http://web.archive.org/web/20060108184335/http://www.dogsinthenews.com/issues/ consultado 25 de febrero de 2016.
- 5. http://history.nasa.gov/animals.html; consultado el 25 de febrero 2016
- Post-Challenger evaluation of Space Shuttle risk assessment and management, Committee on Shuttle Criticality Review and Hazard Analysis Audit, Aeronautics and Space Engineering Board, National Academy Press, National Research Council, January 1988, USA
- Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners, M. Stamatelatos & H. Dezfuli, NASA/SP-2011-3421, Second Edition, December 2011, Washington, D.C.

País	1^{er} vuelo	Ultimo	Núm.	Carga	Familia	
Brazil	1997	2003	2/0	380 kg	VLS-1	3^{ero} explotó en
						tierra matando 21
China	1971		167/158	3000-12000	Long March 2-3-4	
			0/0	1500-25000	Long March 5-6-7	
Europa	1979	1989	28	2650	Ariane 1-2-3	
	1988	2003	116	4720-7000	Ariane 4	
	2002		76/74	9600-21000	Ariane 5	
	2012		5/5	2300	Vega	
	2020			10500	Ariane 6	
Francia	1965	1975	?	?	Diamant	
India	1993		29/28	1300-3800	PSLV	
Iran	2007		4/4	50	Safir	
Israel	1988		8	225	Shavit	
Japon	1966	1995	27	770	Mu 1 -3-4	Nissan
	1975	1987	15	730-2000	N-I,II	Mitsubishi
	1994		28/26	8000-19000	H-II,IIA,IIB	Mitsubishi
	1986	1992	9/9	2300	H-I	Mitsubishi
N.Korea	2006		3/1	100	UhHa	
Rumania			,		Haas	
Rusia	1994		25/23	2100	RU-100N,Rokot	
	2014		1/1	14600-12500	Angara A3,A5	
S. Korea	2009	_	3	100	Naro	
UK	1969	1971	4/3	132	Black Arrow	
USA	1957	1959	12/3	23	Vanguard	Martin
	1957	1980	357	1270/-/-/	Thor	Douglas
	1957	1997	514	5900/2340/	Atlas	Lockheed
	1959	2005	369	21900/5773/8600	Titan	Martin-Marietta
	1960	1994	125/104	210/-/-	Scout	USAFNASA
	1960	1989	186	3848/1312/-	Delta	Douglas
	1961	1975	13	18600/-/-	Saturn I	Chrisler/Doublas
	1967	1973	13	118000/-/47000	Saturn V	Boeing/NorthAmerican/Douglas
	1981	2011	135/134	24400/3810/-	STS (Shuttle)	Alliant/Martin-Marietta/Rockwell
	1989		9/9	1450/-/-	Taurus	Orbital sciences
	1989		151	6000/2171/1508	Delta II	ULA
	1990		40	450/-/-	Pegasus	Orbital Sciences
	1991	2004	63	8618/3833/-	Atlas II	Lockheed
	2000	_	11/11	580/-/-	Minotaur I	Orbital Sciences
	2002	_	18/18	23040/13130/9000	Delta IV	ULA
	2002	_	30/30	18850/8900/2807	Atlas V	ULA
	2010	_	22:	13150/4850/-	Falcon 9	Space —
	2010	_	4/4	1735/640/447	Minotaur IV	Orbital Sciences
	2013		5	6000/-/-	Antares	Orbital Sciences
				53000/21200/-	Falcon Heavy	SpaceX
				130000/-/-	SLS	Alliant/Lockheed

Table 8: Apéndice:	Familias	de	cohetes	por	país
--------------------	----------	----	---------	-----	------

País	1^{er} vuelo	Ultimo	Núm.	Carga	Familia	
URSS	1967	2010	610	1500	Kosmos	
	1969	1972	4/0	90000-23500	N1	
	1987	1988	2/2	100000	Energia	
USSR/Rus	1957		1849	5500/2400/1200	Soyuz	
	1965		399/353	23000/6900/5680	Proton	
USSR/Ukr	1967	2009	259	4100	Tsyklon	
USSR/Ukr/						
Rus	1985		82	13700/6160/4098	Zenit	
Ukraina/						
Rusia	1999		17	3600//750	Dnepr	LEO/TLI

Table 9: Apéndice: Familias de cohetes por país (cont.)

Criticalidad en la dinámica de redes de señalización relevantes a la fecundación

Daniel Priego Espinosa^{1,2}, Alejandro Aguado^{1,2}, Jesús Espinal Enríquez^{2,3}, Alberto Darszon⁴ y Gustavo Martínez Mekler^{1,2}

¹Instituto de Ciencias Físicas, UNAM ²Centro de Ciencias de la Complejidad, UNAM ³Instituto Nacional de Medicina Genómica ⁴Instituto de Biotecnología, UNAM

1. Introducción

Desde mediados del siglo pasado el estudio de múltiples fenómenos de interés para la biología se ha visto enriquecido por el creciente éxito que ha tenido el enfoque de los sistemas complejos (1, 2), donde el comportamiento de sistemas multifactoriales, de muchos componentes fuertemente enlazados, con dinámicas y correlaciones a varias escalas de resolución espacial y temporal, da lugar a la emergencia de propiedades globales. En la biología lo anterior es la norma. La disponibilidad de una cantidad de información previamente insospechada a varios niveles de resolución, requiere de los desarrollos de herramientas de análisis de datos, modelación matemática y simulación, utilizadas para los sistemas complejos, muchas de ellas provenientes del estudio de los sistemas dinámicos y de la física estadística. Con este enfoque se busca incorporar ensayos experimentales puntuales, a manera de ir colectando piezas que eventualmente constituirán un "rompecabezas" completo que describa apropiadamente algún proceso biológico de interés. Siempre ha quedado de manifiesto la importancia y necesidad de pasar de colectar, a armar dichas piezas de conocimiento empírico, para entonces darles un sentido integrativo en el contexto de un sistema completo. Este es el propósito de la biología de sistemas.

En este contexto ha sido de interés el estudio de redes formadas por biomoléculas y sus interacciones, las cuales subyacen la regulación de una variedad de procesos moleculares que sostienen el funcionamiento de la célula, e.g. redes genéticas, metabólicas, de señalización bioquímica, entre otras. Las investigaciones se han centrado en caracterizar tres propiedades fundamentales de las redes: su estructura, dinámica y función.

En nuestro grupo de trabajo abordamos cuestiones pertinentes a la fecundación, concretamente estudiamos la fertilización externa del erizo de mar con miras a entender los mecanismos que le permiten a los espermatozoides localizar al óvulo. Dicha capacidad es regulada por una red de señalización que genera fluctuaciones de concentración de calcio en los flagelos, la cual es disparada por compuestos provenientes del entorno del huevo. Gran parte de nuestro trabajo se ha encaminado hacia el entendimiento de los procesos biomoleculares involucrados en el funcionamiento de esta red, esto mediante un modelo dinámico discreto de redes lógicas. Una descripción de algunos de los avances obtenidos bajo este programa de investigación se encuentra en el capítulo "Modelaje de la via de señalización de calcio relacionada con la motilidad de espermatozoides" de la presente serie de memorias: la XVIII Escuela de Verano en Física (3).

El enfoque de la biología de sistemas pone además de manifiesto propiedades globales de la dinámica de la red de señalización tales como su estabilidad, robustez, redundancia y criticalidad. En este escrito nos centramos en una de esas propiedades, al demostrar que la dinámica se encuentra en un régimen crítico que comparte propiedades con los fenómenos críticos estudiados en la física y que ocurren en los sistemas dinámicos.

En la próxima sección presentamos las bases fisiológicas de nuestro objeto de estudio para pasar en la sección 3 a la construcción de modelos en redes y mostrar en la sección 4 la criticalidad del sistema. En la última se ligan los resultados con otras acepciones de criticalidad.

2. Fertilización en erizo de mar como caso de estudio

Uno de los mecanismos adaptativos esenciales que caracterizan y ayudan a la subsistencia de gran parte de los organismos, es la capacidad de reaccionar ante un estímulo externo, va sea acercándose hacia o alejándose de éste, en donde el beneficio de dicho desplazamiento puede tener diferentes significados de acuerdo con el contexto, e.g., búsqueda de comida, relocalización de algún tipo de células con funciones reparadoras hacia un sitio de lesión, encuentro con otra célula para reproducirse con ella. La respuesta senso-motriz que permite a una célula desplazarse con un movimiento dirigido hacia (o lejos de) la fuente de un gradiente de moléculas es conocida como quimiotaxis. Nuestro modelo de estudio es el erizo de mar, un organismo que ha sido empleado como referencia en la biología del desarrollo por más de un siglo, en el cual nos centramos en entender elementos de la fisiología de sus espermatozoides relacionados con su motilidad. Este tipo particular de célula se vale de una estructura subcelular conocida como flagelo para desplazarse describiendo trayectorias helicoidales en el medio. Dichas trayectorias son moduladas por moléculas de un polipéptido pequeño conocido como speract, las cuales son liberadas desde el recubrimiento del óvulo y se difunden para formar un gradiente alrededor de éste último, que sirve como una señal de localización que el espermatozoide debe interpretar. La unión de dichas moléculas con receptores específicos localizados en el flagelo del espermatozoide desencadena una serie de reacciones bioquímicas y electroquímicas al interior de éste (Fig. 1), que en conjunto generan fluctuaciones en la concentración intraflagelar de iones calcio ($[Ca^{2+}]_i$), las cuales constituyen una señal capaz de inducir cambios directamente en la curvatura del flagelo y por consecuencia en el nado. De esta forma, el fenómeno de quimiotaxis se puede entender como el resultado de la interacción coordinada entre variables externas e internas, relacionadas respectivamente con los gradientes de la concentración de estímulo explorada por la trayectoria del espermatozoide y el conjunto de eventos moleculares que suceden al interior del flagelo disparados por el estímulo externo (4).



Figura 1: Red de señalización activada por speract. Las moléculas involucradas se encuentran en el flagelo del espermatozoide e involucran principalmente receptores específicos de speract, enzimas (GC, AC) que sintetizan segundos mensajeros (cGMP, cAMP), canales iónicos y transportadores que permiten el flujo regulado de potasio (K⁺), sodio (Na⁺), cloro (Cl⁻) y calcio (Ca²⁺), los cuales que a su vez generan cambios en el potencial de membrana (v) y pH intracelular. La tira gris representa la membrana celular con canales e intercambiadores iónicos señalados.

3. Modelación mediante redes lógicas discretas

Dada la naturaleza no lineal de la mayor parte de las interacciones que componen la red de señalización y la acumulación gradual de mediciones experimentales sobre propiedades puntuales relacionadas con los eventos moleculares que subyacen a la quimiotaxis, ha sido necesario recurrir a modelos teóricos para avanzar significativamente en la construcción de un panorama completo que integre las observaciones empíricas en una imagen coherente y que al mismo tiempo nos proporcione una idea clara de cuáles son las propiedades esenciales del sistema de señalización visto como un todo.

En nuestro grupo hemos abordado esta red de señalización desde una perspectiva sistémica por medio de modelos discretos lógicos (5–7) cuyo propósito es ayudar a entender la emergencia de oscilaciones de calcio que regulan la motilidad del espermotozoide y proponer predicciones (las cuales han sido comprobadas experimentalmente por nuestro grupo experimental) sobre el papel específico que juegan ciertos nodos en el contexto de la red completa. Este tipo de modelos son una extensión del formalismo Booleano propuesto por primera vez por Stuart Kauffman (8) para estudiar la dinámica de redes de regulación genética. En la propuesta inicial de Kauffman, mediante una aproximación de grano grueso, los niveles de expresión de cada uno de los genes eran representados por nodos binarios, mientras que las interacciones entre genes eran representadas por enlaces entre nodos. Este formalismo implica la discretización tanto de los valores que los nodos pueden adquirir como de las unidades de tiempo en que suceden los cambios de estado. Si bien el formalismo de Kauffman no fue desarrolado para algún sistema biológico particular sino que fué planteado como un modelo generalizable, ha resultado ser útil a la hora de ser adaptado a casos particulares en la biología, tales como la diferenciación celular en el desarrollo floral (9), diferenciación celular del sistema inmune (10) y el ciclo celular de la levadura (11), entre otros. Una presentación detallada, pedagógica al respecto se encuentra en el capítulo "Redes Complejas" de Maximino Aldana, Memorias de la XV Escuela de Verano en Física de la presente serie (12).

Para la modelación de la red de señalización del espermatozoide en términos de una red lógica como las descritas en el párrafo anterior, se consideraron 20 entidades moleculares (Fig.2) representadas mediante 18 nodos binarios y 2 nodos ternarios, es decir, que puedan cambiar entre 2 o 3 estados posibles según la naturaleza fisiológica particular del nodo a ser modelado. Denotando el estado de un nodo *i* a un tiempo *t* como $\sigma_i(t)$, la configuración a un tiempo *t* de la red completa queda expresada por un vector $\Sigma(t)$ con longitud N = 20:

$$\Sigma(t) = (\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t)) \tag{1}$$

La evolución de cada nodo sucede en tiempos de máquina y el valor que adquiere al siguiente instante depende del estado actual de sus respectivos nodos reguladores. Bajo la suposición de que los tiempos característicos de transición son similares para todos los nodos, éstos son actualizados síncronamente en cada paso de tiempo. Así, un nodo dado σ_i tendrá asociado un conjunto de k reguladores $(\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_k})$ y la función de actualización F_i toma como entrada la configuración de reguladores a un tiempo t:

$$\sigma(t+1) = F_n((\sigma_{i_1}(t), \sigma_{i_2}(t), \dots, \sigma_{i_k}(t))$$

$$\tag{2}$$

Las funciones de actualización propias de cada nodo son construidas como tablas de verdad que contienen toda la combinatoria posible de estados de sus reguladores y un valor de salida asignado a cada una de las frases regulatorias, tomando en cuenta observaciones experimentales o hipótesis basadas en conocimiento previo cuando hay información faltante. Establecer estas reglas fue un ejercicio interdisciplinario que requirió de un intercambio teórico-experimental a fondo.

Dada una condición inicial $\Sigma(0)$, la dinámica de la red completa eventualmente alcanza un patrón periódico, después de transcurrido un número finito de pasos, es decir, la configuración del sistema llega a un estado que se repite cada cierto número de pasos de tiempo. Este patrón periódico se define como atractor del sistema y el número de pasos de tiempo entre configuraciones que se repiten



Figura 2: Abstracción de la red de señalización bioquímica mediante una red lógica discreta. Las interacciones activadoras, represoras y duales están coloreadas respectivamente con negro, rojo y amarillo

define el periodo del atractor; mientras que las trayectorias dadas por la sucesión de configuraciones que preceden al atractor reciben el nombre de transitorios. El conjunto de condiciones iniciales que llegan a un atractor constituyen la cuenca de atracción, y el conjunto de atractores con sus respectivas cuencas forma el paisaje de atracción del sistema. En particular, para el nodo del calcio (dCa) se obtiene un patrón oscilatorio que puede asociarse con el tren de oscilaciones típicamente observado experimentalmente en el flagelo de espermatozoides (13). La evolución de la red para una condición inicial basal se muestra en la figura 3.



Figura 3: Simulación para una condición inicial en la cual todos los nodos están en estado basal. En esta gráfica, la configuración en el tiempo t = 0 corresponde al renglón superior, y los subsecuentes renglones corresponden a la evolución del sistema en cada paso de tiempo hasta llegar al atractor. Los colores de blanco, amarillo y rojo corresponden respectivamente a los estados 0,1 y 2

A partir de este tipo de modelos se ha logrado (5–7) un mejor entendimiento del funcionamiento de los canales iónicos luego de haber explorado *in silico* la activación o bloqueo constitutivo de algunos de ellos, seguido de su corroboración experimental. Uno de los casos más complejos explorados fue la caracterización del mecanismo de acción del ácido niflúmico (6) sobre la dinámica de la señal de calcio, ya que ésta droga tiene la peculiaridad de poder actuar sobre al me-

nos 3 tipos de canales, ya sea inhibiéndolos o activándolos. En otro artículo (7), gracias a la consideración de variantes de la red abordada inicialmente y el estudio por medio de simulaciones de diversos escenarios de interacción, buscando consistencia con resultados experimentales observados, se pudo predecir que CatSper es el canal de calcio determinante. Haber hecho una predicción de esta relevancia por medio de un modelo matemático, pone de manifiesto los alcances que pueden tener estudios teóricos en la biología. El trabajo va más allá de proveer un mejor entendimiento y corroborar experimentos previos. La teoría plantea predicciones cuantitativas y cualitativas, y se convierte en un motor para el desarrollo de nuevos experimentos. Por supuesto, la palabra final la tendrán los experimentos. En este tenor, a largo plazo se espera que este tipo de estudios, con su extensión a mamíferos, pueda contribuir a la elaboración de fármacos que alteren la fertilidad ya sea aumentándola o disminuyéndola. Por ejemplo, que contribuya al desarrollo de anticonceptivos masculinos.

4. Régimen Dinámico Crítico

El funcionamiento y la evolución de los seres vivos supone implícitamente que la forma en que éstos contienden ante cambios internos o ambientales depende de mecanismos suficientemente robustos como para que no se colapsen fácilmente y al mismo tiempo suficientemente flexibles como para codificar un repertorio de respuestas con ventajas adaptativas para diferentes escenarios. Si bien pareciera dificil imaginar un sistema en donde coexistan estas dos propiedades aparentemente opuestas: robustez y adaptabilidad, se ha visto que las redes que operan en un régimen dinámico crítico con respecto a la propagación de perturbaciones dan cabida a la optimizacion de ambas capacidades (14, 15).

Derrida propuso un cálculo que proporciona una medida de cómo perturbaciones sobre condiciones iniciales producen alteraciones en la configuración de una red que se propagan en el tiempo (16). Cuando $t \to \infty$, estas perturbaciones tienen tres posibles destinos: se propagan a toda la red (régimen caótico), mueren eventualmente (régimen ordenado) o llegan a un escenario intermedio donde se constriñen a una proporción de nodos semejante al tamaño de la perturbación inicial con comportamientos estadísticos similares. A este régimen se le ha denominado crítico (16), donde el origen del término proviene de los fenómenos críticos asociados a transiciones de fase, donde se presentan correlaciones a todas escalas. La siguiente sección presenta una discusión sobre criticalidad.

Definamos para una configuración inicial $\Sigma(0)$ una perturbación de tamaño *h* tal que genera una nueva configuración $\Sigma'(0)$ que difiere de la original en el valor de *h* nodos, la cual decimos que está a una distancia de Hamming *h*. Al normalizar esta perturbación por el tamaño de la red ($H = \frac{h}{N}$), podemos tener una medida que expresa el porcentaje en que fue alterada una configuración dada.

El mapeo de Derrida $M(H(t)) \equiv \langle H(t+1) \rangle$, que relaciona cómo cambian *en promedio* las alteraciones inducidas por una perturbación H realizada al tiempo t después un paso de evolución, puede indicarnos el régimen dinámico en el cuál

opera una red discreta. En la práctica, para calcular numéricamente este mapeo, se muestrea un conjunto de condiciones iniciales del espacio de configuraciones sobre las cuales se hace incidir perturbaciones que abarcan todos los tamaños posibles, se evolucionan los pares $\Sigma(0)$ y $\Sigma'(0)$ resultantes y se mide el tamaño promedio de la perturbación luego de ese paso de evolución. Es importante enfatizar que se trata de una propiedad promedio de la evolución de todo el sistema y no la de una condición inicial particular.

Para tiempo continuo la susceptibilidad de una red lógica discreta está definida por la pendiente al origen del mapeo:

$$S \equiv \left. \frac{dM(H)}{dH} \right|_{H=0} \tag{3}$$

donde S determina el régimen dinámico en que opera la red: ordenado (S < 1), desordenado (S > 1) o crítico $(S \approx 1)$. Para tiempos discretos normalizados:

$$S = \frac{M(H(0))}{H(0)} \Big|_{H(0) = \frac{1}{N}}$$
(4)

Al calcular el mapeo de Derrida para nuestra red, resulta que el valor de la suceptibilidad es muy cercana a 1, lo cual indica que su dinámica opera en el régimen crítico (Fig.4a).

En el cálculo anterior la criticalidad está relacionada con la invariancia a lo largo del tiempo de la representación de las alteraciones inducidas por una perturbacion en condiciones iniciales. Cabe hacerse la pregunta sobre qué tan robusto es este comportamiento, qué tan sensible es a cambios en la estructura y definición de la red. A nivel estructural, consideremos la respuesta del comportamiento de la red ante la supresión de un nodo. Sucede que al llevar a cabo esta esa supresión (conocida en la literatura como "knock-out" de un componente de la red) el comportamiento crítico de la red resultante se preserva. En la Figura 4b se muestra que ante la supresión de nodos (uno a la vez), se conserva la pendiente de uno en la curva de Derrida ($S \approx 1$), lo cual recordemos es indicativo de un régimen dinámico crítico. Por ende hemos mostrado que la criticalidad es una propiedad robusta ante modificaciones estructurales. En el lenguaje de los sistemas dinámicos a este comportamiento se le conoce como estabilidad estructural. Para la biología esta propiedad es altamente significativa en términos evolutivos. Cabe comentar que con la eliminación de los nodos dCa y v, se obtienen las pendientes S que menos se aproximan a uno (Fig.4b), esto puede estar ligado a que son los que tienen más conexiones salientes y entrantes, (Fig.2) y que guardan un significado central desde un punto de vista fisiológico (4). Consecuentemente su ausencia constituve una pertubación mayor a las demás del sistema.

5. Consideraciones sobre la criticalidad

En un inicio, el término de criticalidad se acuñó en la física para el estudio de transiciones entre fases termodinámicas, posteriormente se adoptó en el



Figura 4: Mapeos de Derrida calculados sobre 500,000 condiciones iniciales aleatorias. Con propósitos de referencia, se grafica una recta identidad con un patrón de colores rojo y azul. (4a) Red completa. Nótese que el valor de susceptibilidad S es cercano a 1, por lo cual esta red opera en el régimen crítico. (4b) Redes resultantes de *knock-outs* individuales. Entre paréntesis se encuentran los correspondientes valores de susceptibilidad. Nótese que, al igual que la red completa, las curvas se aproximan a la vecindad del origen con una pendiente cercana a 1, con excepción de dos casos: el calcio (dCa) y el voltaje (v), cuyas curvas son menos tangentes a la recta identidad y tienden hacia el régimen ordenado

contexto del comportamiento de sistemas dinámicos con transiciones entre fases dinámicas. El mapeo de Derrida, aunque se encuentra ligado a fases dinámicas, tiene características propias. La relación entre las diversas acepciones de la criticalidad aún no ha sido descifrada por completo y es tema de investigación.

Las transiciones de fase han sido motivo de intenso estudio dentro de la física: transiciones hielo/líquido, liquido/gas, ferro/paramagneto, transición superfluida o superconductura, son algunos de los ejemplos. Entre la variedad de las posibles transiciones de fase se distinguen los fenómenos críticos, los cuales caracterizaremos aquí en términos de la transición ferro/para-magnética. Los metales ferromagnéticos, al ser expuestos a un campo magnético generado externamente, poseen un rango de temperaturas a las cuales se magnetizan, esto es mantienen una magnetización en ausencia del estímulo externo, se comportan como imanes. Resulta que al ir aumentando la temperatura, la magnetización decrece, hasta desaparecer a una temperatura que denotaremos por T_c , a partir de ella, el material se comporta como paramagneto, al no retener magnetización inducida. La magnetización es la densidad de momentos dipolares magnéticos, si consideramos un modelo simplificado bidimensional en que los momentos adquieren sólo uno de dos valores (+/-), con interacciones de corto alcance (modelo de Ising), lo que ocurre es que debajo de T_c se tienen dominios magnéticos dominantes, ya sea positivos o negativos (uno de los dos), intercalados con zonas del signo opuesto. Cuando nos encontramos por arriba de T_c los dominios se van desintegrando y se presenta una distribución aleatoria de cúmulos muy pequeños de espines alineados que van desapareciendo al incrementarse la temperatura. En cambio, muy cerca de la temperatura crítica se tienen de dominios compuestos de dominios, compuestos de dominios, en forma iterativa, dando lugar a una estructura sumamente compleja. Estas estructuras se presentan a todos tamaños, y se preservan ante cambios en la resolución de observación, es decir hay una invariancia de escala, todo el sistema está estadísticamente correlacionado a todas las escalas (17). Las interacciones son de corto alcance, pero las correlaciones se presentan en todos los rangos. Para mayor claridad, decimos que dos sitios están correlacionados si la variación de uno de ellos es perceptible a nivel estadístico por el otro. O sea, las variables no son estadísticamente independientes. La presencia de este sin fin de correlaciones a la temperatura crítica T_c da lugar a singularidades en cantidades termodinámicas en la vecindad de T_c , como por ejemplo la susceptibilidad magnética, esto debido a relaciones conocidas como fluctuación-disipación. El estudio de los fenómenos críticos requirió del desarrollo de la teoría del grupo de renormalización, por la cual Kenneth Wilson obtuvo un premio nobel (18). Para los propósitos de esta discusión recalcamos que por debajo de T_c las estructuras se congelan, por arriba se pierden y en T_c se preservan.

Por otra parte en los sistemas dinámicos, tales como sistemas de ecuaciones diferenciales nolineales o mapeos no lineales acoplados, que se utilizan para modelar dinámica de poblaciones ecológicas (ecuaciones de Lotka-Volterra, mapeo logístico), osciladores no-lineales (ecuación de Van der Pol), dinámicas excitables (ecuaciónes de Hodgkin-Huxley y Fitzhugh-Nagumo), entre otros, presentan transiciones entre comportamientos dinámicos cualitativamente muy diferenciados al variar algunos de sus parámetros. En sistemas dinámicos a esas transiciones se les denomina bifurcacioness. Como ejemplo de una bifurcación se tiene la transición de comportamiento regular periódico a uno caótico (dinámica determinista, con sensibilidad a condiciones iniciales, sin predictibilidad con rasgos parecidos a los comportamientos azarosos) a un valor "critico" particular de su parámetro de bifurcación (19). En ese valor la dinámica es marginal, entre el orden y el caos, se presentan correlaciones temporales a todas las escalas, así como una invariancia de comportamiento ante cambios en la resolución temporal de observación. Las analogías con las transiciones de fase termodinámicas saltan a la vista; han sido y continúan siendo estudiadas formalmente.

En las redes que aquí estudiamos, la criticalidad se refiere al régimen entre dos fases dinámicas, y se cuantifica en términos de respuestas a alteraciones generadas por perturbaciones en condiciones iniciales. Para su estudio viene a la mente la terminología utilizada en el análisis de los sistemas dinámicos de una transición entre un comportamiento regular que elimina el efecto de las perturbaciones y otro caótico que las amplifica. Usualmente en los estudios sobre dinámica caótica el sujeto primordial de análisis es el comportamiento asintótico de las trayectorias del sistema dinámico y la sensibilidad ante variaciones de condiciones iniciales. En cambio, en el criterio de Derrida el enfoque es sobre comportamientos globales, promedio, a un paso de tiempo. En este sentido, más que una transición caos-orden, parece ser mas adecuada la nomenclatura de una transición expansión-contracción. Es más, bajo esta perspectiva la proximidad de la susceptibilidad S definida con anterioridad a 1, es reminiscente de la divergencia de la susceptibilidad de compresión volumétrica en el punto crítico termodinámico.

Un punto que merece ser recalcado es que, a diferencia de los casos como el magnético y la ecuación logística donde hay un parámetro de control (la temperatura y el parámetro de bifurcación respectivamente) que sintoniza al sistema en el punto crítico, en nuestro caso es la dinámica, *per se*, la que lleva al régimen crítico. Este tipo de comportamiento se conoce como "criticalidad auto-organizada" (20) y ha sido motivo de intenso estudio.

Enfatizamos que el régimen dinámico crítico de nuestra red preserva estructuras a todas las escalas tal como es el caso del magnetismo en punto crítico o la dinámica marginal de los sistemas dinámicos. Este comportamiento es indicativo de correlaciones y sensibilidad del sistema a todas escalas. Una propuesta, materia de controversia, es que en el régimen crítico la transferencia de información se optimiza. En la biología, una propiedad de esta índole tiene repercusiones profundas. Este comportamiento, aunado a la coexistencia de adaptabilidad y robustez bajo un régimen dinámico crítico da sustento a la hipótesis de que la vida se desarrolla en un punto crítico. Es por ello que no es de extrañarse que se haya manifestado en diversos tipos de redes regulatorias: redes genéticas (21–23), neuronales (24–27) y de señalización (28–30). Con nuestro trabajo mostramos que en un proceso biológico tan fundamental como lo es la fertilización, la criticalidad sale a relucir una vez más.

Referencias

- Gustavo Martínez-Mekler and Germinal Cocho. Al borde del milenio: caos, crisis, complejidad. In *Ciencias de la materia: Genesis y evolucion de sus conceptos fundamentales*, chapter 8, pages 265–299. Siglo XXI, México, 1 edition, 1999. ISBN 978-9682321627.
- Germinal Cocho. Ciencia, Humanismo, Sociedad. De Los Sistemas Complejos a la Imaginación Heterodoxa. Copit arXives, México, 2017. ISBN 978-1-938128-13-4.
- 3. Jesús Espinal, Adán Guerrero, Maximino Aldana, Chris Wood, Alberto Darszon, and Gustavo Martínez-Mekler. Modelaje de la Vía de Señalización de Calcio Relacionada con la Movilidad del Espermatozoide. In Rocío Jauregui and José Recamier, editors, XVIII Escuela de Verano en Física, Cuernavaca, Morelos, México, 1 al 5 de agosto de 2010, México, 2011. UNI-VERSIDAD NACIONAL AUTÓNOMA DE MÉXICO.
- 4. Alberto Darszon, Adán Guerrero, Blanca E. Galindo, Takuya Nishigaki, and Christopher D. Wood. Sperm-activating peptides in the regulation of ion fluxes, signal transduction and motility. *The International journal of developmental biology*, 52(5-6):595-606, jan 2008. ISSN 0214-6282. doi: 10.1387/ijdb.072550ad. URL http://www.ncbi.nlm.nih.gov/pubmed/18649273.
- 5. Jesús Espinal. Maximino Aldana, Adán Guerrero, Christopher Wood, Alberto Darszon, and Gustavo Martínez-Mekler. Discrete dynamics model for the speract-activated Ca2+ signaling net-PloS one, 6(8):e22619, jan work relevant to sperm motility. 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0022619. URL http://www.ncbi.nlm.nih.gov/pubmed/21857937.
- Jesús Espinal-Enríquez, Alberto Darszon, Adán Guerrero, and Gustavo Martínez-Mekler. In silico determination of the effect of multi-target drugs on calcium dynamics signaling network underlying sea urchin spermatozoa motility. *PloS one*, 9(8):e104451, jan 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0104451. URL http://www.ncbi.nlm.nih.gov/pubmed/25162222.
- 7. Jesús Espinal-Enríquez, Daniel Alejandro Priego-Espinosa, Alberto Darszon, Carmen Beltrán, and Gustavo Martínez-Mekler. Network model predicts that CatSper is the main Ca2+ channel in the regulation of sea urchin sperm motility. *Scientific Reports*, Manuscript, 2017.
- 8. S A Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*, 22(3):437-67, mar 1969. ISSN 0022-5193. URL http://www.ncbi.nlm.nih.gov/pubmed/5803332.

- 9. L Mendoza, D Thieffry, and E R Alvarez-Buylla. Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics (Oxford, England)*, 15(7-8):593-606, 1998. ISSN 1367-4803. URL http://www.ncbi.nlm.nih.gov/pubmed/10487867.
- Mariana Esther Martinez-Sanchez, Luis Mendoza, Carlos Villarreal, and Elena R. Alvarez-Buylla. A Minimal Regulatory Network of Extrinsic and Intrinsic Factors Recovers Observed Patterns of CD4+ T Cell Differentiation and Plasticity. *PLoS computational biology*, 11(6):e1004324, jun 2015. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004324. URL http://www.ncbi.nlm.nih.gov/pubmed/26090929.
- Maria I. Davidich and Stefan Bornholdt. Boolean Network Model Predicts Knockout Mutant Phenotypes of Fission Yeast. *PLoS ONE*, 8(9), 2013. ISSN 19326203. doi: 10.1371/journal.pone.0071786.
- Maximino Aldana. Redes Complejas. In José Recamier, Rocío Jáuregui, and Manuel Torres, editors, XIX Escuela de Verano en Física, Cuernavaca, Morelos, México, Julio 25-agosto 5, 2011, pages 13–68, México, 2012. UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO.
- Chris D Wood, Alberto Darszon, and Michael Whitaker. Speract induces calcium oscillations in the sperm tail. *The Journal of cell biology*, 161 (1):89–101, apr 2003. ISSN 0021-9525. doi: 10.1083/jcb.200212053. URL http://www.ncbi.nlm.nih.gov/pubmed/12695500.
- Maximino Aldana. Boolean dynamics of networks with scalefree topology. *Physica D: Nonlinear Phenomena*, 185(1):45-66, oct 2003. ISSN 01672789. doi: 10.1016/S0167-2789(03)00174-X. URL http://linkinghub.elsevier.com/retrieve/pii/S016727890300174X.
- 15. Jorge Flores and Gustavo Martínez-Mekler. *Encuentros con la complejidad.* 1 edition, 2011. ISBN 978-6070302787.
- B Derrida and Y Pomeau. Random Networks of Automata: A Simple Annealed Approximation. *Europhysics Letters (EPL)*, 1(2):45–49, jan 1986. ISSN 0295-5075. doi: 10.1209/0295-5075/1/2/001.
- Alastair Bruce and David Wallace. Critical point phenomena: universal physics at large length scales. In Paul Davies, editor, *The New Physics*, chapter 8, pages 236–267. Cambridge University Press, Cambridge, UK., first edition, 1989. ISBN 9780521304207.
- Kenneth G. Wilson. The renormalization group: Critical phenomena and the Kondo problem. *Reviews of Modern Physics*, 47(4):773–840, 1975. ISSN 00346861. doi: 10.1103/RevModPhys.47.773.
- 19. RobertM. May.Simplemathematicalmodelswithverycomplicateddynamics.Nature,261(5560):459-467,jun

1976. ISSN 0028-0836. doi: 10.1038/261459a0. URL http://www.nature.com/doifinder/10.1038/261459a0.

- Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: an explanation of the 1/ f noise. *Physical Review Letters*, 59(4):381–384, 1987. ISSN 0031-9007. doi: 10.1103/PhysRevLett.59.381.
- 21. Enrique Balleza, Elena R Alvarez-Buylla, Alvaro Chaos, Stuart Kauffman, Ilya Shmulevich, and Maximino Aldana. Critical Dynamics in Genetic Regulatory Networks: Examples from Four Kingdoms. *PLoS ONE*, 3(6): e2456, jun 2008. ISSN 1932-6203. doi: 10.1371/journal.pone.0002456. URL http://dx.plos.org/10.1371/journal.pone.0002456.
- 22. D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek. Morphogenesis at criticality. *Proceedings of the National Academy of Sciences*, 111(10): 3683-3688, 2014. ISSN 0027-8424. doi: 10.1073/pnas.1324186111. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1324186111.
- 23. Matti Nykter, Nathan D Price, Maximino Aldana, Stephen A Ramsey, Stuart a Kauffman, Leroy E Hood, Olli Yli-Harja, and Ilya Shmulevich. Gene expression dynamics in the macrophage exhibit criticality. Proceedings of the National Academy of Sciences of the United States of America, 105(6):1897–900, feb 2008. ISSN 1091-6490. doi: 10.1073/pnas.0711525105. URL http://www.ncbi.nlm.nih.gov/pubmed/18250330.
- John M Beggs and Dietmar Plenz. Neuronal Avalanches in Neocortical Circuits. The Journal of Neuroscience, 23(35):11167-11177, 2003. ISSN 0270-6474, 1529-2401. doi: 23/35/11167 [pii]. URL http://www.jneurosci.org/content/23/35/11167.abstract.
- John M Beggs. The criticality hypothesis: how local cortical networks might optimize information processing. *Philosophical transactions. Se*ries A, Mathematical, physical, and engineering sciences, 366(1864):329– 43, feb 2008. ISSN 1364-503X. doi: 10.1098/rsta.2007.2092. URL http://www.ncbi.nlm.nih.gov/pubmed/17673410.
- 26. Enzo Tagliazucchi, Pablo Balenzuela, Daniel Fraiman, and Dante R. Chialvo. Criticality in large-scale brain fmri dynamics unveiled by a novel point process analysis. *Frontiers in Physiology*, 3 FEB(February):1–12, 2012. ISSN 1664042X. doi: 10.3389/fphys.2012.00015.
- Ariel Haimovici, Enzo Tagliazucchi, Pablo Balenzuela, and Dante R. Chialvo. Brain organization into resting state networks emerges at criticality on a model of the human connectome. *Physical Review Letters*, 110(17):1–4, 2013. ISSN 00319007. doi: 10.1103/PhysRevLett.110.178101.
- Simone Gupta, Siddharth S Bisht, Ritushree Kukreti, Sanjeev Jain, and Samir K Brahmachari. Boolean network analysis of a neurotransmitter signaling pathway. Journal of theoretical biology, 244(3):463–

9, feb 2007. ISSN 0022-5193. doi: 10.1016/j.jtbi.2006.08.014. URL http://www.ncbi.nlm.nih.gov/pubmed/17010384.

- Junil Kim, Drieke Vandamme, Jeong-Rae Kim, Amaya Garcia Munoz, Walter Kolch, and Kwang-Hyun Cho. Robustness and Evolvability of the Human Signaling Network. *PLoS Computational Biology*, 10(7):e1003763, jul 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003763. URL http://dx.plos.org/10.1371/journal.pcbi.1003763.
- 30. Michael Nivala, Christopher Y Ko, Melissa Nivala, James N Weiss, and Zhilin Qu. Criticality in intracellular calcium signaling in cardiac myocytes. *Biophysical journal*, 102(11):2433-42, jun 2012. ISSN 1542-0086. doi: 10.1016/j.bpj.2012.05.001. URL http://www.ncbi.nlm.nih.gov/pubmed/22713558.

Simular en computadoras la evolución de fluidos: ejemplos y estrategias en Astrofísica

Frédéric S. Masset, Instituto de Ciencias Físicas, UNAM

27 de enero de 2017

1. Introducción

En muchas ramas de la Física el experimento juega un papel de primer plano. Sin embargo, en Astrofísica, casi nunca se puede experimentar. Si bien se pueden buscar meteoritos en desiertos para posteriormente analizarlos, o tratar de reproducir en laboratorio las condiciones de vacío extremo que reinan en algunas partes del medio interestelar, en la mayoría de los casos las teorías astrofísicas no se pueden comprobar mediante experimentos. ¿ Como hacer explotar una supernova, o agarrar un gran conjunto de masas casi puntuales que interactúan entre sí de manera exclusivamente gravitatoria, como las estrellas en una galaxia ? Ante esta imposibilidad, no nos queda de otra que recurrir al "experimento del pobre" que son las simulaciones numéricas. La Astrofísica, no es de sorprender, ha llevado a muchos progresos en el campo de las simulaciones numéricas, y hoy en día los astrofísicos son entre los mayores consumidores de tiempo de cómputo en las grandes facilidades computacionales (aunque no los mayores: a conocimiento del autor, en la época actual los mayores consumidores son los físicos que se dedican a la cromodinámica cuántica).

El campo de las simulaciones numéricas en Astrofísica es sumamente vasto, y ni un libro entero bastaría para agotar el tema. Los sistemas son muy diversos no solamente en las escalas espaciales y temporales que abarcan, sino también en los procesos que implican. No hay gran cosa en común entre el interior del Sol, el de una estrella de neutrones, y una nube molecular fría, por ejemplo. No solamente las ecuaciones descriptivas del problema, sino también la manera de resolverlas, van a ser muy distintas. Por ende, el presente capítulo cubre solamente una diminuta parte de la vasta disciplina de los simulaciones numéricas en Astrofísica.

De manera mas precisa, vamos a presentar aquí una manera de simular la evolución temporal de un fluido. Los fluidos astrofísicos, en su inmensa mayoría, son gases, y por ende son fluidos compresibles, al contrario de los líquidos. Existen dos grandes ramas, correspondiendo a visiones diferentes, para simular un fluido. La primera corresponde a los llamados esquemas¹ Lagrangianos. En éstos, se siguen las parcelas de fluido a lo largo de su evolución. La denominación proviene de la llamada derivada Lagrangiana o particular:

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \vec{v} \cdot \vec{\nabla},\tag{1}$$

¹En la jerga de la dinámica computacional de fluidos, un *esquema* es un conjunto de algoritmos enfocados a brindar una solución. Un *código* es la implementación particular de un esquema.



Figura 1: Malla alternada (izquierda) y malla centrada (derecha). En la malla centrada, todas las cantidades (vectoriales y escalares) están definidas en el centro de las celdas, mientras que en la malla alternada, solamente las cantidades escalares están definidas en el centro de las celdas. Las componentes de las cantidades vectoriales están definidas en el medio de la cara o arista a la cual son perpendiculares.

que determina la tasa de evolución de una cantidad relativamente a una parcela de fluido. Una gran ventaja de los esquemas Lagrangianos es que pueden no tener fronteras, como muchos sistemas astrofísicos, y expandirse conforme evolucionan. Cuando, al contrario, son acotados, puede ser conveniente usar otro tipo de esquemas, los esquemas Eulerianos (en dinámica de fluidos, la derivada Euleriana es una derivada parcial con respecto al tiempo y corresponde a la tasa de variación que vería un observador fijo con respecto al referencial en el cual se describe el fluido). Los esquemas Eulerianos suelen ser esquemas en los cuales el estado del fluido se representa en una malla discreta. Simular el fluido consiste entonces en obtener una secuencia de estados (sobre la malla) para un conjunto discreto de fechas, la diferencia entre dos fechas consecutivas siendo el paso de tiempo. Los esquemas de malla también se dividen en una plétora de técnicas muy diferentes entre sí. A continuación vamos a presentar un método llamado de volúmenes finitos y diferencias finitas, con malla alternada. Este método ha sido implementado en el Instituto de Ciencias Físicas en un código llamado FARGO3D. Es un código bastante versátil, tanto por la geometría de la malla como por los ingredientes físicos que puede incorporar. Por razones de espacio y simplicidad, nos vamos a conformar con presentar aquí su versión isotérmica y dejaremos de lado las sutilezas relativas a la curvatura de la malla.

2. Discretización sobre una malla alternada

En una malla alternada, las cantidades escalares y vectoriales no están definidas en el mismo lugar. Aunque esto representa una complejidad adicional en la implementación, tiene ventajas que vamos a entender a continuación. En la figura 1 (parte izquierda), examinemos el centrado de las cantidades vectoriales (es decir el lugar donde están definidas). Supongamos que éstas representan las componentes del vector de velocidad. Las velocidades radiales están definidas en las interfaces² radiales (las que son de radio constante), mientras que las velocidades acimutales están definidas en las interfaces acimutales (las que son de acimut constante).

3. Ecuación de continuidad

La ecuación que rige la evolución de la densidad del fluido, y que se llama ecuación de conservación de masa o de continuidad, se escribe:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0.$$
⁽²⁾

Veamos cómo podemos tratar de encontrar una solución aproximada de esta ecuación sobre la malla. Lo primero que vamos a hacer es integrar la Ec. (2) sobre una celda. Usando el teorema de la divergencia llegamos a³:

$$\frac{\partial}{\partial t} \iint_{\text{celda}} \rho \, d^3 \vec{r} = \int_{\partial \text{celda}} \rho \vec{v} \, d^2 \vec{r},\tag{3}$$

donde ∂ celda representa la orilla de la celda, es decir el conjunto de interfaces que la delimitan. La integral doble del miembro de izquierda representa la masa de la celda, mientras que la integral del miembro de derecha representa el flujo de masa a través de la orilla de la celda. La Ec. (3) es exacta, y ninguna aproximación se ha hecho. A partir de ahora, vamos a hacer tres aproximaciones:

- 1. vamos a sustituir la derivada parcial con respecto al tiempo por una diferencia finita: $\partial X/\partial t \approx [X(\vec{r}, t + \Delta t) - X(\vec{r}, t)]/\Delta t;$
- 2. vamos a suponer la densidad uniforme sobre la celda. Ésta siendo nuestro "elemento de resolución", no podemos pretender a una descripción mas fina del fluido⁴. Esta aproximación permite escribir la masa de la celda como $\rho L_x L_y$, L_x y L_y siendo los lados de la celda en las dos dimensiones (véase Fig. 2).
- 3. Finalmente, vamos a escribir el flujo de masa sobre una interfaz como el producto de su longitud por la velocidad perpendicular a dicha interfaz, definida en su mitad, y el producto de la densidad, evaluada también en la mitad de la interfaz. Al igual que en el primer punto, renunciamos a una descripción de la velocidad mas fina que el elemento de resolución, y nuestro procedimiento equivale a considerar la velocidad uniforme sobre la interfaz.
- A final de cuentas, nuestra Ec. (3) puede reescribirse como:

$$\rho_{i,j}(t + \Delta t) = \rho_{i,j}(t) + \frac{\Delta t}{\Delta x \Delta y} \left[F_i^x - F_{i+1}^x + F_j^y - F_{j+1}^y \right], \tag{4}$$

²Aristas o superficies que constituyen la frontera entre dos celdas adyacentes.

³Para mayor compacidad nos limitamos aquí al caso bidimensional. La generalización al caso tridimensional se hace sin dificultad.

⁴Notase que existen esquemas en los que los campos hidrodinámicos están descritos por polinomios de orden variable dentro de una celda, lo cual nos proporciona una resolución "sub-malla". Estos esquemas, considerablemente mas complejos que los que se describen aquí, no se consideran en esta presentación.



Figura 2: Representación de los flujos en las orillas de una celda, en una malla Cartesiana bidimensional.

donde usamos las notaciones de la Fig. 2. La principal dificultad consiste en evaluar los flujos. Como mencionado arriba, son el producto de la velocidad y de la densidad en el medio de la arista en la cual están definidos (y de la longitud de dicha arista). Aquí vemos inmediatamente el beneficio de tener una malla alternada: de entrada, la velocidad está definida en la mitad de la arista, y ningún tipo de interpolación es necesario para obtener su valor. En cambio, sí es necesario algo de trabajo adicional para evaluar la densidad interfacial. Su valor, a su vez, puede obtenerse mediante varias técnicas. La que implementamos en el código FARGO3D consiste en predecir el valor de la densidad que llega a la mitad del paso de tiempo en la mitad de la arista, usando un llamado método de característica. Una vez conocidos los flujos, las nuevas densidades se obtienen fácilmente con la Ec. (4). Es importante subrayar que este método, al compartir flujos entre celdas, conserva la masa a la precisión de la máquina. Esto, naturalmente, no implica que la masa total de fluido se quede constante a lo largo de la simulación: el fluido puede salir de (o entrar a) la malla a través de su orilla. No obstante, implica que si la orilla de la malla no permitiera el paso de materia (por imposición de una velocidad normal nula), entonces sí se conservaría la masa total a la precisión de la máquina. En términos mas concretos, no se crea ni desaparece masa dentro de la malla, como es de esperar de una versión discreta de la ecuación de continuidad.

4. Ecuación del momento

Resolver la ecuación de conservación del momento es una tarea mas ardua que la de la masa. Esta ecuación se escribe (suponiendo que el fluido es inviscido y no sometido a fuerzas externas):

$$\frac{\partial}{\partial t}(\rho \vec{v}) + \vec{\nabla} \cdot (\rho \vec{v} \otimes \vec{v}) = -c_s^2 \vec{\nabla} \rho, \qquad (5)$$

donde c_s es la velocidad del sonido isoterma. La primera pregunta que surge es ¿cómo definir el momento $\vec{p} = \rho \vec{v}$ en la malla, si $\vec{v} \ge \rho$ no están definidos en el mismo lugar? La estrategia adoptada en FARGO3D es definir dos versiones del momento:

$$p_i^{x,-} = \rho_i v_{i-1/2}^x \quad \text{y} \quad p_i^{x,+} = \rho_i v_{i+1/2}^x, \tag{6}$$

donde para simplificar consideramos una malla monodimensional en x (la generalización a varias dimensiones siendo inmediata). Posteriormente, se transportan las cantidades $p^{x,-}$ y $p^{x,+}$ como cualquier cantidad centrada (por ejemplo la masa, véase sección 3). Una vez que hayan sido obtenidos los nuevos momentos y la nueva densidad, se actualizan las velocidades mediante:

$$v_{i-1/2} = \frac{p_i^{x,-} + p_{i-1}^{x,+}}{\rho_{i+1} + \rho_{i-1}}.$$
(7)

El lector podrá fácilmente comprobar que si no se actualizan los momentos y la densidad, recuperamos las velocidades iniciales. Al usar las técnicas desarrolladas en la sección anterior, aseguramos que las dos versiones discretas del momento se conservan (excepto por las pérdidas o ganancias en la frontera), y que el momento, que podría definirse como el promedio aritmético de ambas versiones, se conserva también. Al conservar masa y momento, aseguramos ciertas propiedades deseables del esquema, como por ejemplo la correcta representación de los choques (la derivación de las ecuaciones de Rankine-Hugoniot se hace escribiendo la conservación de las cantidades a través de un volumen de control atravesado por un choque).

5. Sobre la implementación

El primer propósito para el cual se ha desarrollado el código FARGO3D en el ICF-UNAM es la simulación de discos protoplanetarios. La simulación de éstos es muy costosa, y puede representar millones de pasos de tiempo⁵, por lo que se hizo el desarrollo para GPUs (*Graphics Processing Units*). Son plataformas muy rápidas, que cuentan con miles de núcleos de cómputo. Se tienen que programar en paralelo y para ello usamos el lenguaje CUDA. El lector interesado puede consultar la página del código: http://fargo.in2p3.fr. Una fuente de información útil es también el artículo que describe con mucho detalle los algoritmos de FARGO3D, disponible en arXiv.

⁵Existe un criterio sobre el paso de tiempo, llamado *condición CFL* o *condición de Courant*, que dice que la información no puede propagarse de mas de una celda sobre un paso de tiempo; eso acota el paso de tiempo de manera drástica.

Efecto Casimir Dinámico

Ricardo Román Ancheyta, José Récamier Angelini

Instituto de Ciencias Físicas, UNAM

1 Introducción

En 1948 Casimir predijo una fuerza atractiva entre dos placas paralelas perfectamente conductoras colocadas en el vacío [1]. A lo largo de las últimas décadas, se han propuesto una amplia variedad de consecuencias fundamentales y medibles debidas a fluctuaciones cuánticas bajo la influencia de condiciones externas [2]. Hace relativamente pocos años se lograron mediciones de alta precisión de la fuerza de Casimir las cuales confirman los conceptos básicos de la teoría cuántica de campos en presencia de condiciones estáticas externas [3].

En 1970, G. T. Moore publicó el artículo Quantum theory of the Electromagnetic Field in a variable length one-dimensional cavity [4], en donde hace la predicción de que fotones reales pueden ser creados a partir del estado de vacío debido a las condiciones de frontera dependientes del tiempo de los espejos sobre la energía de punto cero del campo. La creación de fotones debida a las oscilaciones en las paredes del espejo se conoce actualmente con el nombre de Efecto Casimir Dinámico (ECD) y se le considera una prueba directa de la existencia de las fluctuaciones cuánticas del vacío del campo electromagnético. Múltiples trabajos se han reportado en donde se estudia el ECD debido a su relevancia desde el punto de vista fundamental de la mecánica cuántica [5]. Sin embargo, fue hasta hace relativamente poco tiempo que una realización experimental del efecto pudo llevarse a cabo mediante la arquitectura de circuitos cuánticos superconductores en donde la longitud efectiva del resonador es modulada rápidamente [6].

2 Teoría

El Hamiltoniano efectivo más simple que describe el ECD puede escribirse como: [7, 8] (usando unidades con $\hbar = 1$)

$$H = \omega(t)\hat{a}^{\dagger}\hat{a} + i\chi(t)\left(\hat{a}^{\dagger 2} - \hat{a}^{2}\right)$$
(1)

en donde

$$\chi(t) = \frac{1}{4\omega(t)} \frac{d\omega(t)}{dt}$$

y $\omega(t)$ es la frecuencia instantánea de la cavidad. El Hamiltoniano dado en la Ec 1 representa un modo de un campo electromagnético dentro de una cavidad no estacionaria con espejos planos perfectamente conductores. En el caso del ECD uno de los espejos está fijo mientras que el otro oscila con una pequeña amplitud ϵ . Eligiendo para la frecuencia una dependencia temporal oscilatoria

$$\omega(t) = \omega_0 (1 + \epsilon \sin(\eta t)) \tag{2}$$

en donde ω_0 es la frecuencia cuando las paredes están separadas una distancia fija L_0 y η es un parámetro arbitrario que será seleccionado más adelante, se obtiene:

$$H = \omega_0 \left(1 + \epsilon \sin(\eta t)\right) \hat{a}^{\dagger} \hat{a} + i \frac{\eta \epsilon \cos(\eta t)}{4(1 + \epsilon \sin(\eta t))} \left(\hat{a}^{\dagger 2} - \hat{a}^2\right)$$
(3)

usando que $\epsilon \ll 1$ aproximamos el Hamiltoniano a la forma:

$$H \simeq \omega_0 \hat{a}^{\dagger} \hat{a} + i \frac{\eta \epsilon}{4} \cos(\eta t) \left(\hat{a}^{\dagger 2} - \hat{a}^2 \right) = H_0 + V \tag{4}$$

El operador de evolución temporal queda entonces escrito como $U = U_0 U_I$ en donde U_0 es el operador de evolución temporal correspondiente al Hamiltoniano no perturbado H_0 y satiface la ecuación [9]

$$i\frac{dU_0(t,t_0)}{dt} = H_0 U_0(t,t_0),$$
(5)

con la condición inicial $U_0(t_0, t_0) = 1$ y puesto que H_0 es independiente del tiempo, podemos escribir

$$U_0(t,t_0) = \exp[-iH_0(t-t_0)] = \exp[-i\omega_0(t-t_0)\hat{a}^{\dagger}\hat{a}],$$
(6)
mientras que el operador de evolución temporal en la representación de interacción $U_I(t, t_0)$ satisface:

$$i\frac{dU_I(t,t_0)}{dt} = \left[U_0^{\dagger}(t,t_0)VU_0(t,t_0)\right]U_I(t,t_0),\tag{7}$$

con la condición inicial $U_I(t_0, t_0) = 1$. En lo que sigue elegiremos el tiempo inicial $t_0 = 0$.

Usando la fórmula BCH

$$e^{\alpha \hat{A}} \hat{B} e^{-\alpha \hat{A}} = \hat{B} + \alpha [\hat{A}, \hat{B}] + \frac{1}{2!} [\hat{A}, [\hat{A}, \hat{B}]] + \cdots$$

obtenemos:

$$U_0^{\dagger}(t)\hat{a}U_0(t) = \hat{a}e^{-i\omega_0 t}$$
$$U_0^{\dagger}(t)\hat{a}^{\dagger}U_0(t) = \hat{a}^{\dagger}e^{i\omega_0 t}$$

de forma que el Hamiltoniano en la representación de interacción queda:

$$H_I(t) = i\frac{\eta\epsilon}{4}\cos(\eta t) \left(\hat{a}^{\dagger 2}e^{2i\omega_0 t} - \hat{a}^2 e^{-2i\omega_0 t}\right)$$
(8)

Escribiendo ahora el $\cos(\eta t)$ en su forma exponencial y reemplazando en la ecuación (8) obtenemos

$$H_I(t) = i\frac{\eta\epsilon}{8} \left(\hat{a}^{\dagger 2} e^{i(\eta+2\omega_0)t} + \hat{a}^{\dagger 2} e^{i(2\omega_0-\eta)t} - \hat{a}^2 e^{i(\eta-2\omega_0)t} - \hat{a}^2 e^{-i(\eta+2\omega_0)t} \right), \quad (9)$$

elegimos ahora condiciones de resonancia haciendo $\eta = 2\omega_0$, obteniendo:

$$H_I(t) = i \frac{\omega_0 \epsilon}{4} \left(\hat{a}^{\dagger 2} e^{i4\omega_0 t} + \hat{a}^{\dagger 2} - \hat{a}^2 - \hat{a}^2 e^{-i4\omega_0 t} \right)$$

en donde se tienen dos términos independientes del tiempo y dos más que oscilan rápidamente (con frecuencia $4\omega_0$). Despreciendo a estos últimos (aproximación de onda rotante), obtenemos el Hamiltoniano de interacción independiente del tiempo:

$$H_I = i \frac{\omega_0 \epsilon}{4} \left(\hat{a}^{\dagger 2} - \hat{a}^2 \right) \tag{10}$$

cuyo operador de evolución temporal es

$$U_I(t) = \exp[-iH_I t] = \exp[\frac{\omega_0 \epsilon}{4} (\hat{a}^{\dagger 2} - \hat{a}^2)t].$$
(11)

Podemos entonces escribir el operador de evolución temporal como

$$U(t) = U_0(t)U_I(t) = e^{-i\omega_0 t\hat{n}} e^{\frac{\omega_0 \epsilon}{4} (\hat{a}^{\dagger 2} - \hat{a}^2)t}$$
(12)

en donde $\hat{n} = \hat{a}^{\dagger}\hat{a}$ es el operador de número [9].

El conjunto de operadores $\{\hat{a}^2, \hat{a}^{\dagger 2}, \hat{n}, \hat{a}, \hat{a}^{\dagger}, 1\}$ es cerrado ante la operación de conmutación. Esto quiere decir que el conmutador de cualquier pareja de elementos en el conjunto es también un elemento del conjunto:

	\hat{a}^2	$\hat{a}^{\dagger 2}$	\hat{n}	\hat{a}	\hat{a}^{\dagger}	1
\hat{a}^2	0	$4\hat{n}+2$	$2\hat{a}^2$	0	$2\hat{a}$	0
$\hat{a}^{\dagger 2}$	$-4\hat{n}-2$	0	$-2\hat{a}^{\dagger 2}$	$-2\hat{a}^{\dagger}$	0	0
\hat{n}	$-2\hat{a}^2$	$2\hat{a}^{\dagger 2}$	0	$-\hat{a}$	\hat{a}^{\dagger}	0
â	0	$2\hat{a}^{\dagger}$	â	0	1	0
\hat{a}^{\dagger}	$-2\hat{a}$	0	$-\hat{a}^{\dagger}$	-1	0	0
1	0	0	0	0	0	0

Debido a que $U_I(t) = \exp\left[\frac{\omega_0\epsilon}{4}(\hat{a}^{\dagger 2} - \hat{a}^2)t\right]$ y que el conjunto $\{\hat{a}^2, \hat{a}^{\dagger 2}, \hat{n}, 1\}$ es cerrado ante conmutación, es posible escribir esta exponencial en forma de un producto de exponenciales [10, 11]:

$$\exp\left[\lambda(\alpha_1(t)\hat{a}^{\dagger 2} + \alpha_2(t)\hat{a}^2)\right] = \exp[\phi_1(\lambda, t)\hat{a}^{\dagger 2}]$$
$$\times \exp[\phi_2(\lambda, t)\hat{n}] \exp[\phi_3(\lambda, t)\hat{a}^2] \exp[\phi_4(\lambda, t)]$$
(13)

en donde λ es un parámetro auxiliar que haremos tender a 1 al final del cálculo y $\alpha_1(t) = \omega_0 \epsilon t/4$, $\alpha_2(t) = -\alpha_1(t)$. El conjunto de funciones $\{\phi_i(\lambda, t)\}$ se determina con la constricción $\phi_i(\lambda = 0, t) = 0$. Derivando la ecuación 13 con respecto a λ y usando independencia lineal se obtiene el conjunto de ecuaciones acopladas:

$$\frac{d\phi_1}{d\lambda} = \alpha_1 + 4\alpha_2\phi_1^2, \quad \frac{d\phi_2}{d\lambda} = 4\alpha_2\phi_1,$$
$$\frac{d\phi_3}{d\lambda} = \alpha_2 e^{2\phi_2}, \quad \frac{d\phi_4}{d\lambda} = 2\alpha_2\phi_1,$$

La primera es una ecuación de Riccati la cual es posible resolver [12] y conocida $\phi_1(\lambda, t)$ el resto se obtiene por integración directa. La solución es:

$$\phi_1(\lambda, t) = \frac{1}{2} \tanh[2\alpha_1(t)\lambda], \quad \phi_2(\lambda, t) = -\log[\cosh[2\alpha_1(t)\lambda]],$$
$$\phi_3(\lambda, t) = \frac{1}{2} \tanh[2\alpha_1(t)\lambda], \quad \phi_4(\lambda, t) = \frac{1}{2}\phi_2(\lambda, t),$$

Haciendo $\lambda = 1$ en la expresión para el operador de evoloción temporal U_I obtenemos:

$$U_I(t) = e^{\phi_1(t)\hat{a}^{\dagger 2}} e^{\phi_2(t)\hat{n}} e^{\phi_3(t)\hat{a}^2} e^{\phi_4(t)}.$$
(14)

Una forma alternativa para obtener el operador de evolución temporal U_I es aplicando el teorema de Wei-Norman [13] el cual establece que: Si el Hamiltoniano puede escribirse como una combinación lineal de operadores, independientes del tiempo, que cierran un álgebra de Lie, entonces el operador de evolución temporal puede escribirse en forma de un producto de exponenciales, esto es:

$$H_{I} = \sum_{n=1}^{N} f_{n}(t)\hat{X}_{n}, \quad U_{I}(t) = \prod_{n=1}^{N} e^{\alpha_{n}(t)\hat{X}_{n}},$$

y las funciones complejas $\alpha_n(t)$ satisfacen un conjunto de ecuaciones diferenciales ordinarias acopladas, no lineales, que se obtiene al sustituir el producto de exponenciales en la ecuación de Schrödinger.

Conocido el operador de evolución temporal podemos calcular el valor esperado de cualquier observable del sistema. Por ejemplo, para el valor medio del operador de número de fotones partiendo de un estado de número $|n\rangle$, se obtiene

$$\langle n|\hat{n}(t)|n\rangle = \langle n|U_I^{\dagger}U_0^{\dagger}\hat{n}U_0U_I|n\rangle = \langle n|U_I^{\dagger}\hat{n}U_I|n\rangle,$$

aplicando la transformación, obtenemos

$$U_I^{\dagger} \hat{n} U_I = \hat{n} (1 - 8\phi_1 \phi_3 e^{-2\phi_2}) - 2\phi_3 (1 - 4\phi_1 \phi_3 e^{-2\phi_2}) \hat{a}^2 + 2\phi_1 e^{-2\phi_2} \hat{a}^{\dagger 2} - 4\phi_1 \phi_3 e^{-2\phi_2}$$

de donde:

$$\langle n|\hat{n}(t)|n\rangle = n(1 - 8\phi_1 e^{-2\phi_2}) - 4\phi_1 \phi_3 e^{-2\phi_2}$$
(15)

si el estado inicial es el estado de vacío, entonces:

$$\langle 0|\hat{n}(t)|0\rangle = -4\phi_1\phi_3 e^{-2\phi_2} = \sinh^2(\epsilon\omega_0 t/2)$$
 (16)

En la figura 1 mostramos, en escala semi-logarítmica, la dependencia temporal del valor medio del número de fotones a partir del estado de vacío con parámetros $\epsilon = 0.1$, $\omega_0 = 1$; es claro el comportamiento exponencial correspondiente al ECD.



Figure 1: Evolución temporal del valor medio del operador de número para el estado de vacío $\langle 0|\hat{n}(t)|0\rangle$.

3 Compresión y dispersiones

Para analizar el comportamiento temporal de diferentes observables físicas consideremos primero la evolución temporal de los operadores de creación y aniquilación. En la representación de Heisenberg éstos están dados como:

$$\hat{a}(t) = U_I^{\dagger} U_0^{\dagger} \hat{a} U_0 U_I, \qquad \hat{a}^{\dagger}(t) = U_I^{\dagger} U_0^{\dagger} \hat{a}^{\dagger} U_0 U_I, \qquad (17)$$

aplicando las transformaciones obtenemos:

$$U_{0}^{\dagger}\hat{a}U_{0} = ae^{-i\omega_{0}t}, \qquad U_{0}^{\dagger}\hat{a}^{\dagger}U_{0} = a^{\dagger}e^{i\omega_{0}t},$$
$$U_{I}^{\dagger}\hat{a}U_{I} = \hat{a}(e^{\phi_{2}} - 4\phi_{1}\phi_{3}e^{-\phi_{2}}) + 2\phi_{1}e^{-\phi_{2}}\hat{a}^{\dagger}$$
$$U_{I}^{\dagger}\hat{a}^{\dagger}U_{I} = e^{-\phi_{2}}\hat{a}^{\dagger} - 2\phi_{3}e^{-\phi_{2}}\hat{a}$$

de donde:

$$\hat{a}(t) = e^{-i\omega_0 t} \left(\hat{a} (e^{\phi_2} - 4\phi_1 \phi_3 e^{-\phi_2}) + 2\phi_1 e^{-\phi_2} \hat{a}^{\dagger} \right)$$
(18)

$$\hat{a}^{\dagger}(t) = e^{i\omega_0 t} \left(e^{-\phi_2} \hat{a}^{\dagger} - 2\phi_3 e^{-\phi_2} \hat{a} \right)$$
(19)

Las ecuaciones anteriores se pueden escribir:

$$\begin{pmatrix} \hat{a}(t) \\ \hat{a}^{\dagger}(t) \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{a}^{\dagger} \end{pmatrix}$$
(20)

con:

$$t_{11} = e^{-i\omega_0 t} (e^{\phi_2} - 4\phi_1 \phi_3 e^{-\phi_2}), \quad t_{12} = 2\phi_1 e^{-i\omega_0 t} e^{-\phi_2}$$
$$t_{21} = -2\phi_3 e^{-\phi_2} e^{i\omega_0 t}, \quad t_{22} = e^{i\omega_0 t} e^{-\phi_2}$$

Las cuadraturas q(t) y p(t) están dadas por:

$$q(t) = \frac{1}{\sqrt{2}} \left[\hat{a}(t) + \hat{a}^{\dagger}(t) \right], \quad p(t) = \frac{i}{\sqrt{2}} \left[\hat{a}^{\dagger}(t) - \hat{a}(t) \right]$$
(21)

sustituyendo los resultados anteriores obtenemos:

$$q(t) = \frac{1}{\sqrt{2}} \left[(t_{11} + t_{21})\hat{a} + (t_{12} + t_{22})\hat{a}^{\dagger} \right], \ p(t) = \frac{i}{\sqrt{2}} \left[(t_{22} - t_{12})\hat{a}^{\dagger} - (t_{11} - t_{21})\hat{a} \right]$$

Para calcular las dispersiones es necesario obtener $q^2(t) \ge p^2(t)$

$$q^{2}(t) = \frac{1}{2} \left[(t_{11} + t_{21})^{2} \hat{a}^{2} + (t_{12} + t_{22})^{2} \hat{a}^{\dagger 2} + (t_{11} + t_{21})(t_{12} + t_{22})(2\hat{n} + 1) \right]$$
$$p^{2}(t) = -\frac{1}{2} \left[(t_{22} - t_{12})^{2} \hat{a}^{\dagger 2} + (t_{11} - t_{21})^{2} \hat{a}^{2} - (t_{22} - t_{12})(t_{11} - t_{21})(2\hat{n} + 1) \right]$$

Los valores esperados entre estados de número son:

$$\langle n|q(t)|n\rangle = 0, \quad \langle n|q^{2}(t)|n\rangle = (n + \frac{1}{2})(t_{11} + t_{21})(t_{12} + t_{22}),$$

$$\langle n|p(t)|n\rangle = 0, \quad \langle n|p^{2}(t)|n\rangle = (n + \frac{1}{2})(t_{22} - t_{12})(t_{11} - t_{21})$$
(22)

de donde obtenemos las dispersiones:

$$\Delta q(t) = \sqrt{\langle n | q^2(t) | n \rangle}, \quad \Delta p(t) = \sqrt{\langle n | p^2(t) | n \rangle}.$$

En la figura 2 se muestra la evolución temporal de las dispersiones en las coordenadas q(t), p(t) y su producto, con el parámetro $\epsilon = 0.1$ Nótese



Figure 2: Evolución temporal de las dispersiones Δq (azul), Δp (oro) y el producto $\Delta q \Delta p$ (verde) para el estado n = 3 con parámetros $\epsilon = 0.1$, $\omega_0 = 1$.

que para ciertos instantes del tiempo, la dispersión en alguna de las coordenadas es cercana a cero pero el producto de las dispersiones es siempre mayor que ambas dispersiones y mayor que el valor mínimo permitido por el principio de incertidumbre de Heisenberg. El hecho de que en alguna de las cuadraturas, la dispersión pueda ser menor a la permitida por el principio de incertidumbre de Heisenberg se conoce como squeezing (*compresión*) y se debe a los términos no lineales en los operadores de creación y aniquilación que contiene el Hamiltoniano.

Los estados coherentes [14] son los estados cuánticos que más se parecen a los estados clásicos ya que saturan la relación de dispersión $\Delta x \Delta p = \hbar/2$. Estos estados han sido ampliamente utilizados en óptica cuántica [15]. En términos de su evolución temporal, éstos permanecen localizados alrededor de su correspondiente trayectoria clásica y cuando interactúan con potenciales armónicos no cambian su forma en el tiempo. Esto es, un estado coherente al tiempo t_0 sigue siendo un estado coherente al tiempo t. De acuerdo con Glauber, los estados coherentes son tales que:

- (1) Minimizan la relación de incertidumbre
- (2) Son estados propios del operador de aniquilación del oscilador armónico
- (3) Se crean a partir del estado de vacío mediante la aplicación del operador de desplazamiento $\hat{D}(\alpha) = \exp[\alpha \hat{a}^{\dagger} \alpha^* \hat{a}].$

Con cualquiera de las definiciones anteriores, se obtienen los estados:

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$$
(23)

que han sido expresados en términos de estados de número (estados propios del oscilador armónico). Si $\alpha \to 0$ entonces los términos de la suma con n > 0 son despreciables y nos queda solamente el estado con n = 0 obteniendo entonces $|\alpha\rangle = |0\rangle$ y el estado $|0\rangle$ es un estado coherente y por lo tanto es un estado de mínima dispersión.

A partir de las expresiones que obtuvimos para los operadores q(t), p(t), $q^2(t)$ y $p^2(t)$ podemos evaluar los valores esperados entre estados coherentes, el resultado es:

$$\begin{split} \langle \alpha | q(t) | \alpha \rangle &= \frac{1}{\sqrt{2}} \left[(t_{11} + t_{21})\alpha + (t_{12} + t_{22})\alpha^* \right], \\ \langle \alpha | p(t) | \alpha \rangle &= \frac{i}{\sqrt{2}} \left[(t_{22} - t_{12})\alpha^* - (t_{11} - t_{21})\alpha \right], \\ \langle \alpha | q^2(t) | \alpha \rangle &= \frac{1}{2} \left[(t_{11} + t_{21})^2 \alpha^2 + (t_{12} + t_{22})^2 \alpha^{*2} + (t_{11} + t_{21})(t_{12} + t_{22})(2|\alpha|^2 + 1) \right], \\ \langle \alpha | p^2(t) | \alpha \rangle &= -\frac{1}{2} \left[(t_{22} - t_{12})^2 \alpha^{*2} + (t_{11} - t_{21})^2 \alpha^2 - (t_{22} - t_{12})(t_{11} - t_{21})(2|\alpha|^2 + 1) \right], \end{split}$$

En la figura 3 se muestra la evolución temporal de las dispersiones $\Delta q(t)$, $\Delta p(t)$ y su producto $\Delta q(t)\Delta p(t)$ calculados entre estados coherentes. En la figura se muestra el caso en que $\alpha = \sqrt{3}$ ya que el valor medio del operador de número entre estados coherentes es:

$$\langle \alpha | \hat{n} | \alpha \rangle = \langle \alpha | \hat{a}^{\dagger} \hat{a} | \alpha \rangle = \alpha^* \alpha \langle \alpha | \alpha \rangle = |\alpha|^2$$

en donde usamos que $\hat{a}|\alpha\rangle = \alpha |\alpha\rangle$ y $\langle \alpha | \hat{a}^{\dagger} = \alpha^* \langle \alpha |$ y para calcular las dispersiones entre estados de número tomamos como ejemplo n = 3.

Podemos notar que al tiempo inicial (t = 0) las dispersiones en $q \ge p$ son iguales y corresponden a un estado de mínima dispersión, su producto siendo 1/2 (usamos $\hbar = 1$). Esto corresponde a un estado coherente de acuerdo a la primera definición de Glauber. Debido a que el Hamiltoniano contiene términos de segundo orden en los operadores de creación y aniquilación, los estados coherentes no evolucionan en estados coherentes y las dispersiones



Figure 3: Evolución temporal de las dispersiones Δq (azul), Δp (oro) y su producto (verde) para el estado coherente $|\alpha\rangle \operatorname{con} \alpha = (\sqrt{3}, 0) \operatorname{con} \operatorname{parámetros} \epsilon = 0.1, \omega_0 = 1.$

dejan de tomar el valor más pequeño consistente con el principio de incertidumbre. Esto se manifiesta en el hecho de que el producto de las dispersiones oscila alejándose del valor 1/2 y la amplitud de las oscilaciones crece en el tiempo. Con respecto a las dispersiones $\Delta q(t)$ y $\Delta p(t)$ notamos que éstas presentan compresión y esta compresión se alterna entre las dos cuadraturas. Hay sin embargo ciertos instantes de tiempo para los cuales el producto de las dispersiones se acerca a 1/2, esto sucede cuando la compresión en alguna de las cuadraturas es máxima.

4 Conclusiones

En estas notas hemos presentado la construcción del operador de evolución temporal, para un sistema que consiste en un modo de un campo electromagnético inmerso en una cavidad no estacionaria con espejos planos perfectamente conductores. El Hamiltoniano es transformado a la representación de interacción y eligiendo de forma adecuada la frecuencia de oscilación de la cavidad, se llega a un Hamiltoniano de interacción independiente del tiempo. Este Hamiltoniano consta de operadores que cierran un algebra de Lie lo cual permite escribir el operador de evolución correspondiente en forma de un producto de exponenciales. Calculamos el valor medio del operador de número en la representación de Heisenberg entre estados de vacío encontrando la generación exponencial de fotones que corresponde al efecto Casimir Dinámico; evaluamos también las dispersiones en las cuadraturas del campo encontrando que al tomar los promedios entre estados de número las dispersiones son siempre mayores que el valor mínimo permitido por el principio de incertidumbre. Al calcular las dispersiones entre estados coherentes, encontramos que los valores obtenidos al tiempo inicial son tales que $\Delta q(t_0) = \Delta p(t_0)$ y además $\Delta q(t_0)\Delta p(t_0) = 1/2$. Al evolucionar en el tiempo, los estados dejan de ser estados de mínima dispersión y se presenta el efecto de compresión debido a la presencia de operadores cuadráticos de creación y aniquilación.

Agradecimientos: Agradecemos a la DGAPA-UNAM proyecto IN108413

References

- [1] H. B. G. Casimir, Proc. K. Ned. Akad. Wet. **51**, 793 (1948).
- G. Plunien, B. Müller, and W. Greiner, Phys. Rep. 134, 87 (1986); M. Bordag, *The Casimir Effect 50 years later* (World Scientific, Singapore, 1999).
- [3] U. Mohideen and A. Roy, Phys. Rev. Lett. 81, 4549 (1998); S. K. Lamoreaux, Phys. Rev. Lett. 78, 5 (1997).
- [4] G. Moore, J. Math. Phys. (N.Y.) **11**, 2679 (1970).
- [5] P D Nation, J R Johansson, M P Blencowe, F Nori, Rev. Mod. Phys. 84, 1-24 (2012); V V Dodonov, Physica Scripta 82(3) 038105 (2010); V V Dodonov, A B Klimov, Phys. Rev. A 53(4) 2664-2682 (1996).
- [6] C M Wilson, G Johansson, A Pourkabirian, M Simoen, J R Johansson, F Duty, T Nori, Nature 479, 376-379 (2011).
- [7] V V Dodonov, Physics Letters A **207**, 126-132 (1995).
- [8] C K Law, Phys. Rev. Lett. 73, 1931 (1994); C K Law, Phys. Rev. A 51(3), 2537 (1995).
- [9] Quantum Mechanics Vol. I. Claude Cohen Tannoudji, Bernard Diu, Frank Laloë, Wiley, 2005.
- [10] R. R. Puri, Mathematical Methods of Quantum Optics, (New York, Springer, 2001) p. 48
- [11] R. Gilmore, Lie Groups, Lie Algebras, and some of their applications, (New York, Wiley, 1974) p.460
- [12] R Román-Ancheyta, M Berrondo, J Récamier, J. Phys. Conference Series 698 (2016) 012008
- [13] J Wei and E Norman, Proc. Am. Math. Soc. 15, 327 (1964).
- [14] R J Glauber, Phys. Rev. Lett. **10**, 84 (1963).
- [15] L Mandel and E Wolf, Optical Coherence and Quantum Optics (Cambridge: Cambridge University Press, 1995)