

# Proyecto de Desarrollo e Investigación en Informática

## Neurocomputación y Supercómputo

*Humberto Carrillo Calvet*

Facultad de Ciencias

### **Contenido**

1. Resumen
2. Investigación Básica
  - 2.1. Introducción
  - 2.2. Antecedentes
  - 2.3. Objetivos
  - 2.4. Metas, Metodología e Hipótesis
  - 2.5. Referencias
3. Desarrollo de Software y Aplicaciones
  - 3.1. Software para la Investigación (Proyecto de Supercómputo)
  - 3.2. Aplicación de Redes Neuronales a la Minería de Datos
  - 3.3. Otras Aplicaciones en Perspectiva
4. Participantes
5. Colaboradores
6. Actividades Académicas
7. Infraestructura
8. Requerimientos Financieros
9. Duración del Proyecto
10. Cotizaciones

## 1. Resumen

Este proyecto constituye un esfuerzo (multidisciplinario) por integrar sinérgicamente un trabajo de investigación y desarrollo con miras a obtener un potencial de aplicación con sentido inmediato y local. Como tal es la continuación natural del trabajo iniciado y desarrollado en el marco del proyecto REDII-UNAM 1998-99. Así pues, en este proyecto se extenderán y profundizarán las investigaciones iniciadas y, sobre la base de los desarrollos de software realizados anteriormente, se construirán nuevos sistemas con más y mejores capacidades. Un elemento importante de mejoría del software será la incorporación de supercómputo. Se continuará el esfuerzo de vinculación con la industria para aplicar nuestras investigaciones y nuestros desarrollos de software a la solución de problemas de importancia práctica.

En la dirección de investigación básica nos proponemos continuar las investigaciones sobre neuronas y redes neuronales originalmente planteadas. Los resultados obtenidos el año pasado [16A], [16B] nos permiten ahora el desarrollo de una metodología de análisis de sistemas modelados con neuronas de integración y disparo. Esta metodología se aplicará al estudio de la dinámica y de las propiedades computacionales de redes basadas en este tipo de neuronas. Las simulaciones y los cálculos, que el análisis de estos sistemas requiere, exige un esfuerzo computacional extraordinario que se llevará a cabo en la supercomputadora Origin-Cray. Se ha concebido todo un proyecto de desarrollo de software que utilizará los recursos de supercómputo con que cuenta nuestra universidad.

En la dirección de las aplicaciones, se contempla la aplicación de tecnología de redes neuronales al tratamiento de problemas de minería de datos. En colaboración con expertos en Infometría del Instituto Finlay de la Habana, Cuba, se está evaluando la posibilidad de construir un "Observatorio de Ciencia y Tecnología", basado en esta tecnología. En la medida en la que se cuente con el apoyo técnico requerido, se considera la posibilidad de hacer un convenio con la sociedad civil "Laboratorios de Sistemas Complejos" para continuar el proyecto de desarrollo de los productos tecnológicos (e.g.: ATMBuilder) que fueron aplicados a la banca mexicana (Banamex) el año pasado. Se considera también la posibilidad de aplicar esta tecnología al desarrollo de software educativo (para la enseñanza de la Ciencia a nivel de secundaria), para esto se está discutiendo un proyecto de colaboración con la empresa DIDATEC, en coordinación con la SEP y la UNAM, vía la Coordinación de Vinculación.

En colaboración con el System Science and Engineering Research Center, de Arizona State University, se planea la realización de otro taller de investigación sobre los temas de este proyecto.

Siete participantes hacen tesis sobre varios aspectos del proyecto: dos de doctorado cuatro de maestría y uno de licenciatura.

## 2. Investigación Básica

*Redes Neuronales (Diseño y análisis de modelos matemáticos de procesadores y redes de procesadores neuronales)*

### 2.1. Introducción

Nuestro trabajo de investigación está primordialmente enfocado al estudio de *neuronas y redes neuronales*. Por estas entendemos, procesadores y redes de procesadores que están contruidos a semejanza de las neuronas y los circuitos neuronales que constituyen el sistema nervioso de los seres vivos. Este tema está en la frontera de dos disciplinas relativamente nuevas: la *Neurocomputación* (o *Neuroingeniería*) y la *Neurociencia Computacional*. En un lenguaje más tradicional esto significa que el avance en las investigaciones sobre neuronas y redes neuronales, contribuye al conocimiento en dos vertientes principales: hacia la Neurobiología, por una parte y hacia la Ciencia de la Computación y la Inteligencia Artificial, por la otra. Los problemas de esta frontera de investigación tienen mucho en común con los que clásicamente han sido identificados como problemas de la *Cibernética*

La investigación en este campo tiene un carácter *interdisciplinario* e.g.: la búsqueda de nuevas formas de computación y la investigación de sus fundamentos matemáticos, son de interés para los neurobiólogos y científicos cognitivos interesados en entender el funcionamiento de la mente y el cerebro; recíprocamente, la comprensión de los procesos biofísicos, responsables de la neurofisiología subyacente y que constituyen la base de los procesos mentales de alto nivel (memoria, aprendizaje, creatividad, etc.), ofrece paradigmas inspiradores a los ingenieros, especialistas en inteligencia artificial y/o científicos de la computación.

Una red neuronal (biológica o artificial) es un sistema dinámico [21]. Su función, propiedades, atributos y particularmente sus capacidades computacionales, están determinadas por lo que técnicamente llamamos su *dinámica*

Una característica fundamental de la dinámica de este tipo de sistemas es su carácter altamente *no lineal*. Esto enriquece sus capacidades, pero al mismo tiempo complica el análisis de sus modelos matemáticos sobrepasando rápidamente los alcances de la metodología y/o tecnología disponible.

## 2.2. Antecedentes

En investigaciones previas con Rinzel y Baer [1,2 y 3] hemos analizado la dinámica de *modelos "realistas"* de neuronas, es decir, muy apegados a la realidad biofísica de las células nerviosas. Estos muestran interesantes capacidades de procesamiento pero son demasiado sofisticados matemáticamente para poder analizar redes de muchos procesadores: constan de sistemas de tres y hasta cinco ecuaciones diferenciales, para cada neurona [23].

Existen otros modelos de neuronas como el del *integrador con fugas*, ampliamente utilizado por M. Arbib [4, 5, 6], que están dados en función de una sola ecuación diferencial por cada neurona. Con F. Cervantes Pérez, tenemos un proyecto de trabajo sobre el análisis de este tipo de redes de neuronas y hemos escrito un par de artículos [7, 8].

El *modelo de integración y disparo* [9, 10, 11] es otro tipo de modelo de neurona, que también tiene una dinámica más rica que las neuronas clásicas de McCulloch y Pitts [12], pero puede todavía ser modelada con una sola ecuación diferencial. Los modelos del integrador con fugas, anteriormente mencionados, aunque son no lineales, ofrecen la ventaja de ser lineales por tramos. Como hemos puesto de relieve en la tesis de Mendoza [13], los modelos de integración y disparo no tienen esta cualidad, pero tienen otra cualidad muy importante: sus propiedades de sincronización pueden reducirse al estudio de una simple dinámica cíclica, que puede ser investigada usando la teoría de los sistemas dinámicos en la circunferencia. Sobre este tema de la dinámica cíclica hemos escrito varios trabajos y tesis [14, 15, 16, 17, 18].

El estudio de redes de neuronas modeladas por medio de ecuaciones diferenciales implica un análisis matemático muy complejo que es evitado por un amplio sector de investigadores, quienes prefieren trabajar con redes de neuronas que obedecen una dinámica tan simple (e.g.: las neuronas de McCulloch y Pitts) que su modelación no requiere del uso de ecuaciones diferenciales.

Otros autores han analizado modelos matemáticos de sistemas de integración y disparo que, en lugar de obedecer una ecuación diferencial, operan de acuerdo a una regla geométrica (ver [24] y las referencias contenidas en él).

Nuestro enfoque hacia la construcción de redes neuronales artificiales usa al máximo el conocimiento biológico de frontera y modelos de neurona tan complejos como lo permita la metodología de análisis matemático (y la tecnología computacional) actual.

### 2.3. Objetivo

Nuestro objetivo principal aquí es estudiar modelos matemáticos de procesadores y redes de procesadores neuronales, mucho más complejos que las clásicas neuronas de MacCulloch y Pitts [12], pero suficientemente simples todavía, para que su dinámica y por ende, sus propiedades computacionales, puedan ser estudiadas con la ayuda de tratamientos analíticos. Con este fin se considerarán modelos basados en procesadores de *integración y disparo* y se buscará alcanzar la comprensión de las propiedades computacionales de las redes neuronales (dinámicas globales) en términos del conocimiento de la arquitectura de ellas y las capacidades de procesamiento que tiene cada neurona (dinámica individual).

El logro de este objetivo es fundamental para el *diseño y control de las redes neuronales artificiales* que es una de las aplicaciones de nuestro interés.

### 2.4. Metas, Metodología e Hipótesis

El estudio de los procesadores de integración y disparo que realizamos el año pasado se basó en las investigaciones de Keener-Hoppensteadt-Rinzel (KHR):

J.P. Kenner, F.C. Hoppensteadt, J. Rinzel. **Integrate and Fire Models of Nerve Membrane Response to Oscillatory Input**. *SIAM J. APPL. Math.* Vol. 41, No. 3, 1981.

Estos autores usan la ecuación diferencial

$$v' = -av + S + H \operatorname{sen}(t),$$

sujeta a una condición de disparo, para modelar cada neurona, donde  $a$ ,  $S$ ,  $H$ , son parámetros del sistema. Gracias a que la ecuación es lineal y el forzamiento es trigonométrico, esta ecuación puede resolverse analíticamente. Aprovechando este hecho, estos autores pudieron inventar un procedimiento para descubrir los patrones de cambios cualitativos (bifurcaciones) en la dinámica de la neurona.

En el marco del proyecto REDII del año pasado, se obtuvieron resultados que permiten generalizar la metodología de Keener-Hoppensteadt-Rinzel a un sistema de la forma

$$v' = -av + S + g(t),$$

donde la función periódica  $g(t)$  no está restringida a ser una función trigonométrica y se descubrieron importantes propiedades cualitativas que pueden ser conocidas directamente de la ecuación diferencial, sin tener que conocer soluciones analíticas. Estos resultados son parte de la tesis de Maestría en Matemáticas de Miguel Angel Mendoza (en proceso) y fueron reportados en:

**Neuronas de Integración y Disparo con Acumulación Lineal.** *Reporte de Investigación 99-3, Laboratorio de Dinámica no Lineal, Facultad de Ciencias, UNAM.* (H. Carrillo, F. Ongay y M. A. Mendoza).

Posteriormente, se descubrieron otros teoremas que permitieron el desarrollo de una metodología completamente distinta y que aplica para analizar sistemas de la forma más general, esto es, de la forma:

$$v' = F(t, v, I),$$

donde  $I$  es un vector de parámetros y la función  $F(t, v, I)$  no tiene que ser lineal en ninguna de sus variables. *Esta investigación es parte de la tesis doctoral en Ciencias (Matemáticas) del alumno Fernando Ongay (ya terminada y con todos los votos aprobatorios presentará su examen de grado la primera semana de Julio, en la UNAM) y ha sido reportada también en el artículo:*

**On the Firing Maps of a General Class of Forced Integrate and Fire Systems.** Artículo aceptado para publicación en *Mathematical Biosciences*. (H. Carrillo and F. Ongay).

Nos proponemos ahora aplicar esta metodología para estudiar el comportamiento sincrónico de estos procesadores de integración y disparo, modelados por ecuaciones no lineales. Se usarán estos métodos teóricos para guiar simulaciones computacionales que permitan determinar la estructura de las diferentes zonas de sincronización en el espacio de parámetros. *Esta investigación es parte del proyecto de investigación doctoral que el alumno Miguel Angel Mendoza, somete actualmente al Programa de Doctorado en Ciencia e Ingeniería de la Computación de la UNAM.*

Investigaremos la relación que existe entre este tipo de modelos, gobernados por una ecuación diferencial, y los modelos geométricos anteriormente mencionados, con miras a evaluar cual tiene mejores capacidades de procesamiento. *(En colaboración con Fernando Ongay, una vez que haya obtenido el grado de doctor).*

Los procesadores neuronales de integración y disparo que nos interesan, son osciladores no lineales y el análisis de los modelos, que se hacen en términos de ellos, plantea interesantes problemas de la Teoría Matemática de Oscilaciones. Un problema fundamental es la búsqueda de métodos que simplifiquen el análisis. *Como parte de la tesis doctoral del alumno José Ramón*

*Guzmán, se desarrolla un nuevo método de análisis basado en la "promediación" de las ecuaciones diferenciales, que promete ser potente y de mucha utilidad.*

Este método de promediación se basa en principios diferentes a aquellos en los que se fundamenta el "**Método del Promedio**" de la teoría de perturbaciones que constituye uno de los métodos más importantes de la teoría cualitativa de los sistemas dinámicos. Sobre este método y sobre su aplicación al estudio de sistemas débilmente acoplados hemos trabajado en el pasado [22A, 22B]. Hoy estamos evaluando las perspectivas de su aplicación al estudio de redes de neuronas débilmente acopladas.

Siendo los investigadores Hoppensteadt y Rinzel autores de uno de los trabajos fundamentales sobre los sistemas de integración y disparo que nos interesan, estamos en comunicación con ellos discutir nuestros proyectos e investigaciones sobre estos temas.

En colaboración con el profesor Hoppensteadt organizamos en el System Science and Engineering Research Center, de Arizona State University, un "*Workshop of Mathematical Neuroscience and Neuroengineering*".

La colaboración con el profesor Hoppensteadt es interesante también a propósito de sus recientes investigaciones sobre "*Modelos Canónicos*" y "*Redes de Neuronas Débilmente Acopladas*", temas expuestos en su más reciente libro:

***Weakly Coupled Neural Networks.*** *Applied Math. Sciences, No. 126. Springer, 1997.*

Junto con Hoppensteadt estamos evaluando las perspectivas de aplicación del Método del Promedio de la Teoría de Perturbaciones, al estudio de redes de neuronas débilmente acopladas.

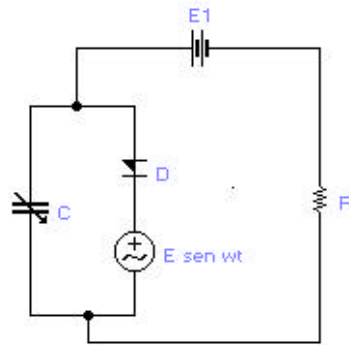
Dada la importancia que tiene implementar redes neuronales en hardware y la ventaja que representa el poder hacer experimentos, en paralelo a los modelos matemáticos, estudiaremos también modelos analógicos de estas neuronas (de integración y disparo).

Se continuará la experimentación (iniciada en el marco del proyecto del año pasado) con modelos analógicos y se compararán los resultados empíricos con los de los de la modelación.

Los procesadores neuronales de integración y disparo analógicos estarán constituidos por circuitos electrónicos con elementos no lineales que, debido a la presencia de rangos de resistencia negativa, dan lugar a fenómenos de *histéresis*. Así pues, estudiaremos sistemas cuyas neuronas están constituidas por circuitos electrónicos (ver figura) cuyos elementos no lineales son "diodos

shockley” o arreglos de transistores que permiten producir curvas características adaptables a los rangos de resistencia negativa requeridos [19, 20].

*Sobre este tema continua la elaboración de su tesis de maestría en Informática (DEPFI, UNAM) el Ingeniero en Electrónica, Francisco Chávez.*



*Se continuará también la investigación de las propiedades de sincronización que tienen este tipo de procesadores neuronales en respuesta a estímulos periódicamente aplicados (típico en la comunicación interneural de estas redes). Se continuará buscando la presencia y las causas, de los fenómenos de estabilidad, biestabilidad de equilibrios, existencia y coexistencia de múltiples atractores periódicos, sincronización racional (phase locking) así como respuesta asíncrona (desordenada o caótica).*

Para llevar a cabo este análisis se aplicará la teoría de número de rotación de Poincaré-Denjoy y sus generalizaciones entre otros temas de la teoría cualitativa de los sistemas dinámicos no lineales.

Sobre ésta parte del trabajo continúan la elaboración de sus tesis de Maestría en Matemáticas en la Facultad de Ciencias los alumnos Elizabeth Díaz Bobadilla y Miguel Angel Mendoza, quienes son egresados de la carrera de Matemáticas de la Facultad de Ciencias de la UAEMEX y actualmente alumnos de posgrado de la Facultad de Ciencias de la UNAM.

En todo este trabajo se parte de la hipótesis de que las propiedades de una red están determinadas por su arquitectura y por las capacidades dinámicas de cada uno de los procesadores que la integran. Asumimos también que la dinámica de cada uno de los procesadores puede explicada en términos de ecuaciones diferenciales y/o en diferencias.

## 2.5. Referencias

- [1] S. Baer, J. Rinzel y H. Carrillo. **Analysis of an Autonomous Phase Model for Neuronal Parabolic Bursting.** *Journal of Mathematical Biology*, 33: 309-333, 1995.
- [2] S. Baer, J. Rinzel y H. Carrillo. **A Three Variable Autonomous Phase Model for Neuronal Parabolic Bursting.** *Differential Equations and Applications to Biology and Industry* (pp.1-11), M. Martinelli (Editor), World Scientific, Singapore, 1996.
- [3] H. Carrillo, J. Rinzel y S. Baer. **Nonlinear Oscillations in Neurons Models.** *Memorias de EUROMECH- 2nd European Nonlinear Oscillations Conference*, Vol. 1, pp. 105-108, Praga, 1996.
- [4] M. Arbib. **Brain Theory and Neural Networks.** *The MIT Press*, Cambridge, M.A., 1995.
- [5] M. Arbib y F. Cervantes Pérez. **Stability and Parameter Dependency Analyses of a Facilitation Tectal Column (FTC) Model,** *J. Math. Biol.*, 1990.
- [6] M. Usher, H.G. Schuster and E. Niebur. **Dynamics of Populations of Integrate-and-Fire Neurons, Partial Synchronization and Memory.** *Neural Computation*, 5:313-326, 1993.
- [7] A. Herrera, F. Cervantes y H. Carrillo. **Análisis del Comportamiento Dinámico de Redes Neuronales.** *Memorias del 2do. Congreso Iberoamericano de Inteligencia Artificial.* 183-202. Limusa - Noriega, México, 1990.
- [8] A. Herrera, F. Cervantes y H. Carrillo. **Modulation of Neural Activation Dynamics.** Enviado a la revista *Neural Computation*.
- [9] J.P. Kenner, F.C. Hoppensteadt, J. Rinzel. **Integrate and Fire Models of Nerve Membrane Response to Oscillatory Input.** *SIAM J. APPL. Math.* Vol. 41, No. 3, diciembre 1981.
- [10] F.C. Hoppensteadt. **An Introduction to the Mathematics of Neurons.** *Cambridge University Press*, 1986.
- [11] F.C. Hoppensteadt, E. M. Izhikevich. **Weakly Connected Neural Networks.** *Springer*, 1997.
- [12] W. McCulloch y W. Pitts, *Bull. Math, Biophys.* 5, 115, 1943.
- [13] M.A. Mendoza Reyes. **Dinámica de las Neuronas de Integración y Disparo.** *Tesis Profesional. Facultad de Ciencias, UAEMEX*, 1997.

- [14] I.E. Díaz Bobadilla. **Sistemas Dinámicos en la Circunferencia.** *Tesis Profesional. Facultad de Ciencias, UAEMEX*, 1996.
- [15] J.R. Guzmán. **Sistemas Dinámicos en la Circunferencia: Aplicaciones a la Teoría de Números y Modelación de Neuronas.** *Tesis de Maestría. Facultad de Ciencias, UNAM*, 1994.
- [16A] H. Carrillo and F. Ongay. **On the Firing Maps of a General Class of Forced Integrate and Fire Systems.** Artículo aceptado para publicación en *Mathematical Biosciences*.
- [16B] H. Carrillo, F. Ongay y M. A. Mendoza. **Neuronas de Integración y Disparo con Acumulación Lineal.** *Reporte de Investigación 99-3, Laboratorio de Dinámica no Lineal*, Facultad de Ciencias, UNAM.
- [17] H. Carrillo, F. Ongay y J.R. Guzmán. **Dinámica de las Iteraciones de la Función de Arnold.** *Aportaciones Matemáticas*, Serie Comunicaciones, 14: 405-414, Sociedad Matemática Mexicana, 1994.
- [18A] H. Carrillo y J.R. Guzmán. **A Dynamical Systems Proof of the Little Theorem of Fermat.** *Aportaciones Matemáticas*. Serie Comunicaciones 22 (1998). pp 63-65. Sociedad Matemática Mexicana.
- [18B] H. Carrillo y J.R. Guzmán. **A Dynamical Systems Proof of Euler's Generalization of the Little Theorem of Fermat.** *Aportaciones Matemáticas*. Serie Comunicaciones 25. (1999) pp 199-202. Sociedad Matemática Mexicana.
- [19] L.O. Chua, C.A. Desoer, E.S. Kuh. **Linear and Nonlinear Circuits.** *McGraw Hill*, 1987.
- [20] C. Mead. **Analog VLSI and Neural Systems.** *Addison Wesley*, 1989.
- [21] H. Carrillo. **Cerebro, Redes Neuronales y Sistemas Dinámicos.** *Revista: Información Científica y Tecnología, CONACYT.* Abril de 1990. Vol. 12, Núm. 163.
- [22A] H. Carrillo. **The Averaging Method and the Persistence of Attractors.** *Applicable Analysis*, Vol. 29, pp. 191-208, 1988.
- [22B] H. Carrillo. **Averaging and Synchronization of Weakly Coupled Systems.** En: *Nonlinear Oscillations in Biology and Chemistry*, H. Othmer (ed.). *Lecture Notes in Biomathematics* No. 66, pp. 244-251, *Springer-Verlag*, Heidelberg 1986.

[23] E. Salinas y H. Carrillo. **Modelación y Simulación del Potencial Eléctrico de una Neurona**. *Reporte 92-3 del Laboratorio de Dinámica no Lineal, Facultad de Ciencias*, 1992.

[24] L. Glass. **Cardiac Arrhythmias and Circle Maps: A Classical Problem** *Chaos, Vol. 1, No. 1, 13-19*, 1991.

[25] H. Carrillo and David Lípman. **The Multiple Sequence Alignment Problem in Biology**. *SIAM Journal of Applied Mathematics*, Vol.48, No. 5, pp. 1073-1082, 1988.

### 3. Desarrollo de Software y Aplicaciones

#### 3.1. Software para la Investigación

##### *Proyecto de Supercómputo Cray Origin 2000*

##### Introducción

Estudiar la dinámica de sistemas interactuantes es un problema clásico en las ciencias básicas y la ingeniería. Estos aparecen al considerar un sistema compuesto por diferentes subsistemas acoplados (en interacción). El concepto de interacción permite realizar una reducción importante: conociendo la dinámica de cada subsistema, por separado, se busca conocer la dinámica global del sistema completo. Este enfoque es de interés, por ejemplo en Neurobiología y en Neurocomputación, pues en estos casos es deseable poder predecir el comportamiento (global) de una red neuronal (biológica o artificial) a partir del conocimiento de la arquitectura (estructura de conexiones) de la red y la dinámica de las neuronas (o procesadores) individuales. En Inteligencia Artificial se necesita tener la capacidad de predecir la dinámica global de la red para lograr que el diseño tenga todas las capacidades computacionales deseadas.

Típicamente, los procesadores neuronales son osciladores no lineales. El estudio de la interacción de sistemas que individualmente tienen un comportamiento periódico es de particular interés para la teoría de redes neuronales. Naturalmente, en este caso, nos preguntamos si el sistema compuesto (la red) tendrá también un comportamiento periódico o desordenado (caótico).

## Objetivo

Diseñar y desarrollar un sistema de software que sirva para analizar, visual e interactivamente, la forma en que los procesadores de integración y disparo responden a diversos estímulos externos.

## Descripción del Sistema

Se considerarán sistemas de integración y disparo modelados de manera geométrica o en términos de ecuaciones diferenciales. El sistema se operará desde una computadora personal (PC) a través de una interfaz visual y amigable.

El estudio teórico de la dinámica de estos procesadores requiere un gran esfuerzo computacional, pero la estructura lógica del problema permite la paralelización de sus algoritmos y por esto, el diseño del sistema contempla la capacidad de interaccionar, a través del servicio telnet de la red y del sistema de colas de entrada/salida (NQE/NQS) de la Cray *Origin 2000*, para hacer cálculo intensivo en esta computadora.

Este proyecto se llevará a cabo en el Laboratorio de Dinámica no Lineal de la Facultad de Ciencias de la UNAM, en colaboración con la Sección de Computación del Departamento de Ingeniería Eléctrica del CINVESTAV y Laboratorios de Sistemas Complejos, S.C.

Cabe mencionar que ya se cuenta con un sistema de software para el análisis de osciladores de integración y disparo modelados de manera geométrica, denominado ***Diente de Sierra***, y otro para analizar osciladores modelados en términos de ecuaciones diferenciales, denominado ***Integración y Disparo***.

El nuevo sistema que se propone desarrollar, integrará y ampliará las capacidades de análisis de los sistemas ***Diente de Sierra*** e ***Integración y Disparo***, paralelizando los algoritmos numéricos que sirven para encontrar soluciones y calcular bifurcaciones.

El sistema de software que desarrollaremos permitirá también el estudio de las propiedades de sincronización de neuronas del tipo "integración y disparo" que reciben estimulación periódica. Para el caso de estimulaciones periódicas el sistema permitirá estudiar los diferentes tipos de respuesta caótica o sincronizada del procesador. El cálculo de las sincronizaciones se hará mediante la detección de atractores periódicos de sistemas dinámicos en la circunferencia determinados por las funciones de disparo del oscilador forzado. En algunas regiones del espacio de parámetros los números de rotación de Poincaré son indicadores de la presencia de comportamientos sincronizados, pero incluso en

este caso particular, el esfuerzo computacional requerido para determinar las regiones de sincronización en el espacio de parámetros es muy grande.

## *Componentes y Módulos Principales*

Fundamentalmente el programa tendrá 2 componentes:

- Una componente para la visualización/interacción, construida en la plataforma Windows, utilizando C++ Builder en una PC [5], que proveerá un ambiente (gráfico) amigable para analizar este tipo de procesadores.
- Una componente para hacer cálculo intensivo en la supercomputadora *Cray Origin 2000*

### **Componente en la PC:**

Ofrecerá un ambiente, desarrollado en Windows, que ofrecerá al usuario la posibilidad de trabajar en seis distintos escenarios gráficos:

- *Sawtooth Oscillator*
- *Circle Dynamics*
- *Torus Dynamics*
- *Lifts*
- *Branches*
- *Tongues*

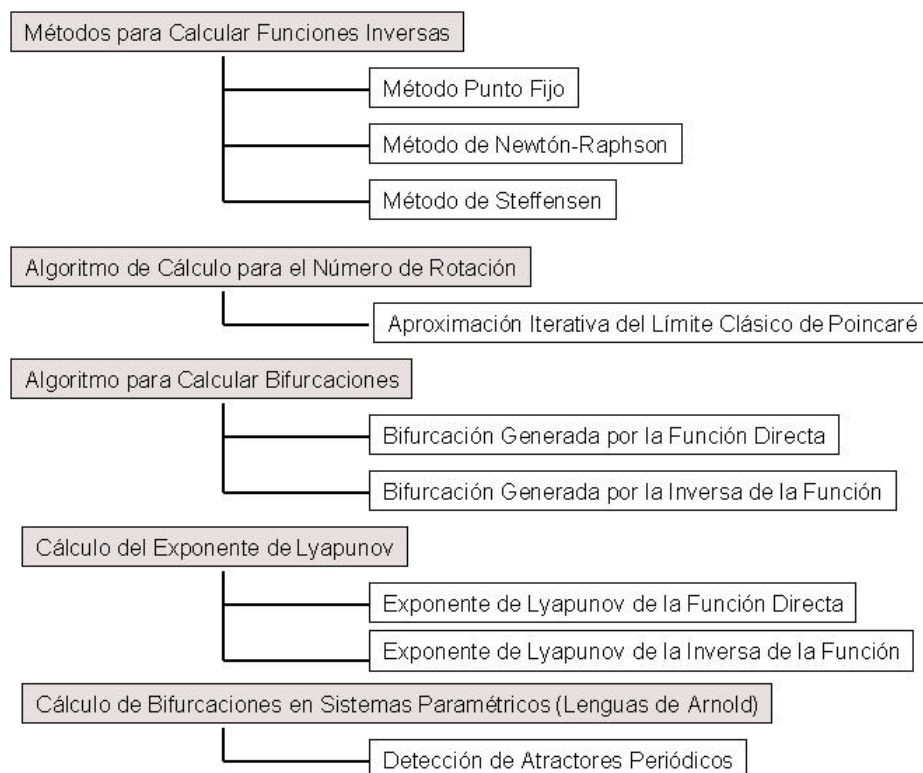
Estos escenarios son adecuados para el estudio de las iteraciones de las funciones de disparo del procesador estimulado [1]. Se conoce que estas funciones de disparo son levantamientos de funciones de la circunferencia, por lo que será conveniente analizar representaciones de tales funciones en la circunferencia y sus correspondientes gráficas en el toro [2].

Módulos que constituyen esta componente:

- Interfaz visual en plataforma Windows que permitirá observar los resultados del cálculo numérico
- Rutinas de cálculo para sistemas de integración y disparo modelados geoméricamente
- Rutinas de cálculo para sistemas de integración y disparo modelados por ecuaciones diferenciales
- Servicio de red Telnet para la comunicación (envío/recepción de información) de la PC con la *Cray Origin 2000* a través de RED-UNAM.

Los módulos de rutinas de cálculo para “sistemas *geométricos de integración y disparo*” y para “sistemas *modelados por una ecuación diferencial*” esencialmente deberán exhibir los mismos escenarios gráficos para analizar la dinámica de estos sistemas a través de funciones de la circunferencia. Con la adición de que en los sistemas modelados por una ecuación diferencial será necesario construir un “Módulo de Integración” que permita obtener la función de disparos a partir de la integración numérica de la ecuación diferencial.

Los algoritmos utilizados para nuestros cálculos son:



### Componente en la Cray Origin 2000:

Estará constituido fundamentalmente por:

- Módulo de rutinas paralelizadas de cálculo para sistemas geométricos
- Módulo de rutinas paralelizadas de cálculo para sistemas modelados por una ecuación diferencial [6].
- Utilización de las directivas de los compiladores MIPSpro C y MIPSpro C++, para la compilación y ejecución de las rutinas a paralelizar. Así también el uso de las directivas de paralelización automática de dichos compiladores (APO) y los lenguajes OpenMP y SHMEM para su

adecuada paralelización de las rutinas. De manera implícita a las directivas de OpenMP se realizará una adecuada distribución de tareas a los distintos procesadores, para evitar una sobrecarga de trabajo en ellos ([8], [9], [10] y [11]).

Se someterán los trabajos que requieran grandes recursos de cómputo al sistema de colas NQE/NQS para dejar trabajando automáticamente a la Origin 2000, y posteriormente recoger los resultados obtenidos.

Los módulos de rutinas paralelizadas de cálculo para “*osciladores geométricos de integración y disparo*” y “*osciladores modelados por una ecuación diferencial*” deberán realizar los cálculos necesarios para que en la PC se puedan desplegar los diagramas de bifurcación. Esto implica determinar atractores periódicos de mapeos de la circunferencia y calcular sus períodos y envolventes, así como los correspondientes exponentes de Lyapunov y los números (o intervalos) de rotación ([2], [3], [4] y [5]).

El sistema operativo utilizado por la *Cray Origin 2000* es IRIX, (UNIX versión de Silicon Graphics), convirtiéndola en una máquina multiproceso, multiusuario y permitiendo la programación en paralelo, razón por la cual es conveniente para nuestro proyecto, pues facilitará también hacer interfaz con las estaciones de trabajo Silicon Graphics con que cuenta nuestro laboratorio.

La supercomputadora, además de las características en hardware que de entrada la convierten en una máquina muy veloz, cuentan con dispositivos dedicados a un tipo muy específico de cálculo o acción. Entre tales podemos mencionar los registros vectoriales, escalares y de dirección; así como las unidades funcionales de suma, multiplicación y aproximación recíproca. Estos dispositivos permiten a la computadora realizar (en algunos casos) varias partes de un mismo cálculo simultáneamente. El procedimiento consiste en dividir el cálculo en secciones y operar dichas secciones en la unidad funcional o registro que corresponde al tipo de datos y operación a realizar.

Es importante hacer notar que las bibliotecas matemáticas y científicas ofrecidas por la *Cray Origin 2000* consisten de un conjunto de rutinas en su mayoría enfocadas a cálculos numéricos a las que es posible acceder desde nuestros programas. Para la construcción de los algoritmos para la detección de atractores haremos uso de estas bibliotecas. Esto conviene ya que estas rutinas están altamente optimizadas y el código que las conforma fue previamente compilado e incluso reescrito en lenguaje ensamblador.

Para observar el rendimiento de nuestro procesos y evaluar la eficiencia en la ejecución de los algoritmos, se necesitará monitorearlos a través de algunas herramientas disponibles en la supercomputadora, ***SpeedUp***, ***HPM*** y ***Perfex***, con la finalidad de darse cuenta de la cantidad y tipo de operaciones que se

realizaron, el número total de información que movió el CPU de/ y hacia la memoria, etc.

Se utilizarán los compiladores: ANSI C, C++ y Power C Accelerator (PCA), pues:

- Ofrecen capacidades de optimización y paralelización automática de código, y
- Permiten combinar paralelización automática y el estándar de programación paralela para convertir código secuencial en código paralelo denominado OpenMP.

Dado que pueden enviarse varios procesos para calcular distintas regiones (de distinto tamaño y complejidad) de sincronización en el espacio de parámetros, tendremos que especificar el tiempo y la memoria a usar para cada región. Será realizado haciendo uso del sistema de colas NQE/NQS de manera no interactiva, a través de un archivo (batch, script, etc) especificando toda la información requerida por la supercomputadora e instrucciones ejecutables de manera automática.

### **Referencias Bibliográficas**

- [1] R. Pérez, L. Glass. Bistability, Period Doubling Bifurcations and Chaos in a Periodically Forced Oscillator. *Physics Letters* 90 A9: 441-443 (1982).
- [2] Oscar R. García Regis. Software para el Análisis Visual Interactivo de Sistemas Dinámicos Discretos en la Circunferencia. Tesis Profesional. Facultad de Ciencias UAEM, 1999
- [3] J. R. Guzmán. Sistemas Dinámicos en la Circunferencia: Aplicaciones a la Teoría de Números y Modelación de Neuronas. Tesis de Maestría. Facultad de Ciencias UNAM, 1994
- [4] I. E. Díaz Bobadilla. Sistemas Dinámicos en la Circunferencia. Tesis Profesional. UAEM 1996
- [5] M. A. Mendoza Reyes. Dinámica de las Neuronas de Integración y Disparo. Tesis Profesional. UAEM 1998
- [6] F. Charte. C++ Builder. Adquiera los fundamentos básicos de programación. Ediciones Anaya Multimedia, S. A. 1997.
- [7] S. Akl. The Design and Analysis of Parallel Algorithms. Prentice-Hall, 1989.

- [8] G. S. Almasi and A. Gottlieb. Highly Parallel Computing. Benjamin/Cummings, second edition, 1994.
- [9] U. Banerjee. Dependence Analysis for Supercomputing. Kluwer Academic Publishers, 1988.
- [10] F. Berman and L. Snyder. On mapping parallel algorithms into parallel architectures. J. Parallel and Distributed Computing, 4(5):439--458, 1987.
- [11] G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. J. Parallel and Distributed Computing, 7:279--301, 1989.

### 3.2. Aplicación de Redes Neuronales a la Minería de Datos

La tecnología moderna permite la acumulación de grandes almacenes de datos (crudos) que necesitan ser explorados en búsqueda de información más elaborada (conocimiento). Un reto tecnológico de actualidad es desarrollar métodos y procedimientos para procesar estos datos y convertirlos en conocimiento útil para la toma de decisiones. A esta área de trabajo se le ha dado en llamar Minería de Datos y muchos problemas de la informática actual requieren la aplicación de técnicas de esta disciplina.

Las redes neuronales han probado ser de gran utilidad para resolver problemas de minería de datos. Estas han sido útiles particularmente para la organización creativa de información y el descubrimiento de conocimiento.

En colaboración con el Dr. Gilberto Sotolongo del Centro de Información del Instituto Finlay estamos evaluando la factibilidad, y la conveniencia, de diseñar un sistema basado en redes neuronales para resolver problemas de minería de datos en infometría.

#### **Objetivo**

Desarrollar un sistema (modular) de información bibliométrica (software basado en redes neuronales) como herramienta para el análisis de información.

#### **Descripción (Resumen)**

Dicho sistema debe incorporar las posibilidades de captura de datos, manipulación de datos, incluida la búsqueda y recuperación de información, así como la edición de bibliografía, el procesamiento estadístico básico (estadística descriptiva), cálculo de leyes bibliométricas, análisis exploratorio de datos incluyendo entre otros el análisis multivariado y técnicas de minería de datos basadas en redes neuronales no supervisadas. El sistema debe contar también

con una interfaz visual, amable e interactiva, que sirva para desplegar y analizar representaciones gráficas de los resultados.

## **Aplicación**

Se está evaluando la posibilidad de orientar el diseño de este desarrollo hacia la construcción de un Observatorio de Ciencia y Tecnología.

Referencia: A. Berson and S.J. Smith. Data Warehousing and Data Mining. McGraw-Hill. 1997

# Observatorio de Ciencia y tecnología

## **Introducción**

El constante crecimiento de la información y de los conocimientos reflejados en publicaciones científicas y técnicas (e.g. artículos, patentes, etc.) le impone a sus usuarios (investigadores, agentes de servicios de información, etc.) nuevos requerimientos, marcados por la impronta de las “nuevas tecnologías de la información”. Felizmente, también estas tecnologías complementan a otras técnicas y metodologías adecuadas, brindando la oportunidad de hacerle frente a tales desafíos. La Bibliometría se inserta en este marco como una disciplina relativamente nueva cuya actual denominación tiene escasamente treinta años. Aun cuando su gestación se inició a comienzos del Siglo XX, su auge es reciente. La misma ha influido sobremanera en otra especialidad vinculada a la organización de la ciencia, la llamada Cienciometría. Ambas, han dado lugar a otra disciplina un poco más reciente llamada Informetría.

*La Bibliometría ha capitalizado la esencia de la cuantificación de la información para diferentes fines.* De una parte, se destacan aquellas aplicaciones relacionadas con la gestión y uso de la información en bibliotecas donde surgió. Por otra parte, de la Bibliometría se han derivado aplicaciones en el campo de la política científica en lo que se conoce como Bibliometría Evaluativa. En su lógico desarrollo, mucho más recientemente la Bibliometría también aparece en la encrucijada de lo que se conoce como *vigilancia tecnológica* (como parte de verdaderos observatorios de ciencia y tecnología) como apoyo a la toma de decisiones en ambientes empresariales e institucionales (producción y servicios) movidos por los desarrollos científicos y tecnológicos, en donde el análisis exploratorio de datos ha encontrado en la “*minería de datos*” y el consecuente “descubrimiento de conocimientos en bases de datos” uno de sus modos de realización más prometedores en la actualidad. Para ello, la Bibliometría se apropia de técnicas conceptuadas en el “análisis exploratorio de datos” donde a las herramientas estadísticas ya establecidas se suman los más modernos

enfoques de análisis multivariado tales como las “*redes neuronales artificiales*” que ofrecen un gran potencial de aplicación para el análisis y procesamiento de grandes masas de “Datos” que requieren ser convertidos (cibernéticamente) en información útil para la toma de decisiones. Lo anterior hace que la Bibliometría vaya de la mano del desarrollo tecnológico, permitiendo que nos aproximemos, no sólo a interpretar esa realidad que constituyen los conocimientos (certificados en las publicaciones), sino que también retroalimenta para transformar dicha realidad.

## **Objetivos**

Creación de un Observatorio (o servicio de observación) de Ciencia y Tecnología, para la vigilancia científica y tecnológica de interés institucional.

## **Alcance del análisis**

Para una mejor comprensión de esta problemática, es necesario abordarla desde la óptica de lo que se conoce como proceso de "producción de conocimientos certificados", es decir, la producción de conocimientos científicos que se registran de forma escrita en documentos los que son sometidos a procesos de arbitraje y una vez aceptados son publicados.

Entre las formas más conocidas que adoptan tales documentos están las de artículos de revistas y patentes; éstas coinciden con las más utilizadas para realizar los trabajos de monitoreo. No se descartan otros tipos de documentos, como son los libros, informes, los trabajos presentados en congresos, etc. Más recientemente el auge alcanzado por el desarrollo de las nuevas tecnologías informativas y de comunicaciones —como Internet—, han potenciado variantes de los tradicionales documentos científicos con la primacía del hipertexto y multimedia.

Los indicadores derivados de las publicaciones científicas deben estar acompañados de indicadores contextuales, i.e. económicos, sociales, demográficos, etc. que aporten un enfoque más amplio al análisis, en su valor relativo, e.g. producto interno bruto, cantidad de profesionales, gastos en educación, población total, población dedicada a las actividades de Investigación y Desarrollo (I+D), etc.

En los últimos años se viene observando que algunos países, basándose en indicadores similares a los enumerados anteriormente, han creado entidades orgánicamente estructuradas para la medición de los resultados de la ciencia de los países. Al propio tiempo, las organizaciones que dependen de la ciencia y la técnica para su desenvolvimiento (investigación, desarrollo, producción y comercialización de productos y servicios) dedican no pocos esfuerzos también en esa dirección pero orientados hacia lo que se ha dado en llamar vigilancia científica y tecnológica como parte de las actividades de inteligencia corporativa o incluso de inteligencia económica. Las entidades que se dedican a realizar formal y organizadamente estos esfuerzos en un país, se han dado en llamar

"observatorios nacionales de ciencia y tecnología". La estructuración de actividades en el ámbito institucional con la óptica empresarial del sector en que se desenvuelven las instituciones es denominada, unidades de vigilancia científica y tecnológica o sencillamente de vigilancia tecnológica.

El común denominador de estas entidades es el de, utilizar fundamentalmente, indicadores basados en publicaciones científicas (conocimientos certificados mediante cuerpos de árbitros en el caso de los artículos de las revistas y mediante los procedimientos técnico-legales en el caso de las invenciones e innovaciones que dan lugar a las patentes) los cuales se someten a un análisis cualitativo y cuantitativo. La Infometría ha dotado a las actividades de observación y de vigilancia, de métodos y técnicas científicamente fundamentados para la elaboración de las hipótesis de análisis que cuerpos de especialistas toman como bases para sus estudios y recomendaciones en políticas y estrategias, según sea el caso.

Esta problemática es abordada de forma puntual por diferentes especialistas; sin embargo, no existe una entidad orgánicamente estructurada que realice tales trabajos. Al propio tiempo, existen especialistas capacitados para realizar tales trabajos; y además, la infraestructura informativa existente cuenta con un potencial científico-tecnológico importante que incluye los recursos informativos (fuentes, servicios y sistemas) disponible (o potencialmente disponibles).

La creación de un *observatorio de ciencia y tecnología* presupone la integración de diferentes recursos (humanos e informativos); así como poder contar con determinada infraestructura que incluya las nuevas tecnologías de información y comunicación.

### **Proyección**

Este proyecto germinal podría evolucionar en todo un observatorio de ciencia y tecnología que cumpla las tareas y funciones siguientes:

- Mantener un permanente monitoreo de la literatura nacional y extranjera independientemente del portador y medio en que ésta se presente.
- Determinar los términos de referencia de los umbrales de los indicadores sobre el desempeño científico de los investigadores del sector de interés y su comparabilidad internacional
- Realizar análisis periódicos de la dinámica de la producción científica del sector de interés.
- Determinar el nivel de impacto de la producción científica del sector en el plano nacional e internacional
- Realizar estudios (por encargo) sobre el desempeño científico y tecnológico de diferentes temáticas, tecnologías y entidades según se requiera.

### **Publicaciones recientes de G. Sotolongo**

1. Macías-Chapula, C. A.; Sotolongo-Aguilar, G. R.; Magde, B.; Solorio-Lagunas, J. Subject Content Analysis of AIDS Literature Produced in Latin America and the Caribbean. *Scientometrics*, 1999; 46 (3) :563 – 574
2. Macías-Chapula, C.A., Sotolongo-Aguilar, G.R., Madge, B. and Solorio-Lagunas, J. Subject content analysis of AIDS literature as produced in or about Latin America and the Caribbean. En: Macías-Chapula, César A., Ed. *Proceedings 7 ISSI Conference, Colima, México; 1999:313-322.*
3. Sotolongo-Aguilar, G.R. , Suárez-Balseiro, C. A. , Guzmán-Sánchez, M.V. Modular Bibliometric Information System with Proprietary Software En: Macías-Chapula, César A., Ed. *Proceedings 7 ISSI Conference, Colima, México; 1999:450-456.*
4. García-Díaz, I.C., Sotolongo-Aguilar, G.R. XVII. Las revistas científicas y su problemática. En: Cetto, Ana M. & Alonso, Octavio (comp.) *Revistas científicas en América Latina. 1999, p. 226-255. ISBN:968-16-5851-5*
5. Sotolongo-Aguilar, G.R., Guzmán-Sánchez, M.V., García, I., Sáenz-Casado, E. Vigilancia y evaluación de la actividad científico - tecnológica. *Reencuentros*. 1998; 21(abril):39-44.
6. García-Díaz, I.C., Sotolongo-Aguilar, G.R. La medición de la ciencia y sus cuestionamientos. *Reencuentros*. 1998; 21(abril):29-38.
7. Guzmán-Sánchez, M.V., Sáenz-Casado, E., & Sotolongo-Aguilar, G.R. Bibliometric Study on Vaccines (1990-1995) Part I. Scientific Production in Iberian - America. *Scientometrics*. 1998; 43(2):189-205.
8. Guzmán-Sánchez, M. V.; Sotolongo-Aguilar, G.R. Producción científica y tecnológica de vacunas en Iberoamérica. *Vaccinmonitor*. 1998; 7 (3):2-12.
9. Guzmán-Sánchez, M.V., Sotolongo-Aguilar, G.R. Gerencia de Información Tecnológica. *Ciencias de la Información*. 1997, 28(3):161-166.

### 3.3. Otras Aplicaciones en Perspectiva

Estamos trabajando con la Coordinación de Vinculación de la UNAM, para encontrar una fórmula adecuada de financiar (aportando apoyo técnico) proyectos de interés para la industria nacional. En particular se está estudiando la conveniencia de tratar de continuar los proyectos de desarrollo de software de interés para la Banca (ATM's/Unisys-Banamex) y el sector educativo (DIDATEC-SEP).

## 4. Participantes

### *Tesarios de Doctorado*

Fernando Ongay Larios. Profesor de la Facultad de Ciencias, UAEMEX.  
José Ramón Guzmán. Investigador del Instituto de Economía, UNAM.

### *Tesarios de Maestría*

Irma E. Díaz Bobadilla. Matemáticas, Facultad de Ciencias, UNAM.  
Francisco Chávez Castañeda. Facultad de Ingeniería, UNAM.  
Miguel Angel Mendoza Reyes. Matemáticas, Facultad de Ciencias, UNAM.  
Oscar García Regis. Ciencias de la Computación, IPN, CINVESTAV

### *Alumnos de Maestría que ingresan este año al Doctorado*

Miguel Angel Mendoza Reyes. Ingeniería y Ciencia de la Computación, IIMAS, UNAM.  
Oscar García Regis. Ciencias de la Computación, IPN, CINVESTAV

### *Técnico Académico*

Antonio Carrillo Ledesma.

## 5. Colaboradores

Nuestra investigación en este campo la hemos venido realizando en colaboración con las siguientes personas:

- *Francisco Cervantes Pérez*, Director del Laboratorio de Comportamiento Adaptativo Neuronal, Neurociencias y Simulaciones, ITAM.
- *Steven Baer*, profesor del Departamento de Matemáticas de Arizona State University y Editor del Journal of Theoretical Biology y Mathematical Biosciences.
- *John Rinzel*, del Courant Institute y Editor en Jefe del Journal of Computational Neurosciences.
- *Frank Hoppensteadt*, Director del Center for Systems Science and Engineering de Arizona State University y Editor del Journal of Mathematical Biology.
- *David Lipman*, Director del National Institute for Biotechnological Information, USA.

Con ellos nos mantenemos en contacto para analizar los problemas de interés común y discutir los avances de nuestras investigaciones.

El perfil de estos investigadores los hace adecuados para este propósito: el profesor Hoppensteadt y el profesor Rinzel son pioneros en el estudio de los modelos de redes de neuronas de integración y disparo [9], cuya investigación es objeto del presente proyecto. Con Baer, especialista en modelación de neuronas, también hemos escrito algunos artículos. Lipman es coautor de [25] y especialista en el análisis y comparación de secuencias biológicas, problema que requiere la aplicación de sofisticadas técnicas de minería de datos.

## 6. Actividades Académicas

En colaboración con el System Science and Engineering Research Center, de Arizona State University, se planea la realización de otro taller de investigación sobre los temas de este proyecto.

## 7. Infraestructura

El Laboratorio de Dinámica no Lineal que tenemos en la Facultad de Ciencias cuenta con el espacio y el equipo necesario (5 PC's y 2 estaciones de trabajo) para realizar las investigaciones planteadas. El proyecto de supercómputo ha sido aprobado por el comité técnico de la DGSCA, tenemos cuenta en la Origin-Cray y ya estamos haciendo las primeras pruebas.

## 8. Duración del Proyecto

Se propone una duración de tres años para el desarrollo de este proyecto