

## *Sistemas Paralelos*

**Armando De Giusti<sup>1</sup>, Marcelo Naiouf<sup>2</sup>, Laura De Giusti<sup>3</sup>, Franco Chichizola<sup>4</sup>, Fernando Tinetti<sup>5</sup>, Andrés Barbieri<sup>6</sup>, Mónica Denham<sup>7</sup>, Horacio Villagarcía<sup>8</sup>, Alejandra Cosentino<sup>9</sup>**

{degiusti,mnaiouf,ldgiusti,francoch,fernando,barbieri,mdenham,hvw,acosenti}@lidi.info.unlp.edu.ar

*Instituto de Investigación en Informática LIDI<sup>10</sup>  
Facultad de Informática. Universidad Nacional de La Plata.*

### **Resumen**

El objetivo de esta línea es investigar en Sistemas Paralelos, esto es, la combinación de problemas de software asociados con la utilización de arquitecturas de procesamiento paralelo, especialmente sistemas multiprocesador distribuidos.

Los temas fundamentales abarcan la especificación, transformación, optimización y verificación de algoritmos ejecutables en sistemas paralelos/distribuidos, la optimización de clases de soluciones en función de modelos de arquitectura multiprocesador, las métricas de complejidad y eficiencia relacionadas con el procesamiento paralelo, la influencia del balance de carga y la asignación de tareas a procesadores, la escalabilidad de los sistemas paralelos, así como simulación y diseño de arquitecturas VLSI orientadas a multiprocesamiento.

Asimismo se ha iniciado el estudio de los modelos de predicción de performance en sistemas paralelos.

Interesa la aplicación de las investigaciones en áreas como el procesamiento de datos numéricos en cómputo científico, el procesamiento de imágenes digitales y las bases de datos distribuidas. Para esto, se trabaja experimentalmente con distintos modelos de arquitectura disponibles o accesibles desde el III-LIDI.

### **Introducción**

Actualmente es innegable la importancia y el creciente interés en el procesamiento paralelo dentro de la Ciencia de la Computación, por un gran número de razones (crecimiento de la potencia de cálculo dada por la evolución tecnológica, transformación y creación de algoritmos que explotan la concurrencia para obtener mejores tiempos de respuesta, la necesidad de tratar sistemas de tiempo real distribuidos, límite físico de las máquinas secuenciales que convierte a la solución paralela en la única factible, etc).

---

<sup>1</sup> Investigador Principal CONICET. Profesor Titular.

<sup>2</sup> Profesor Titular.

<sup>3</sup> Becaria de Perfeccionamiento UNLP. Jefe de Trabajos Prácticos.

<sup>4</sup> Becario de Iniciación UNLP. Ayudante Diplomado.

<sup>5</sup> Profesor Adjunto

<sup>6</sup> Becario de Iniciación UNLP. Ayudante Diplomado.

<sup>7</sup> Ayudante Diplomado.

<sup>8</sup> Profesor Asociado.

<sup>9</sup> Jefe de Trabajos Prácticos.

<sup>10</sup> III-LIDI - Facultad de Informática. UNLP - Calle 50 y 115 1er Piso, (1900) La Plata, Argentina.  
TE/Fax +(54)(221)422-7707. <http://www.lidi.info.unlp.edu.ar>

En términos generales las máquinas paralelas permiten resolver problemas de complejidad creciente y obtener resultados con mayor velocidad. Permite resolver algunas de las restricciones impuestas por las computadoras monoprocesador: además de ofrecer soluciones más rápidas, las aplicaciones paralelizadas pueden resolver problemas más grandes y complejos cuyos datos de entrada o resultados intermedios exceden la capacidad de memoria de una CPU, las simulaciones pueden ser corridas con mayor resolución, los fenómenos físicos pueden ser modelados de manera más realista. En algunos casos el costo del paralelismo puede ser alto en términos del esfuerzo de programación requerido, en muchos casos debe pensarse en la aplicación de técnicas novedosas reescribiendo completamente el código serial, y las técnicas de debugging y tuning de performance “secuenciales” no se extienden fácilmente al mundo paralelo.

Se debe hacer referencia a *sistemas paralelos* como la combinación de algoritmo y arquitectura, ya que a diferencia del cómputo secuencial donde el modelo RAM es aceptado prácticamente como standard, en el paralelismo no se encuentra un modelo unificador ya que cada uno enfatiza determinados aspectos en desmedro de otros. Las diferentes arquitecturas (con memoria compartida o distribuida, con procesadores homogéneos o heterogéneos, distintas topologías de conexión, etc.) hacen que en el desarrollo de las soluciones paralelas sea indispensable considerar la combinación de ambos factores.

La última etapa en el desarrollo de algoritmos paralelos es el mapeo de procesos en procesadores; los objetivos son mejorar la utilización de los procesadores y obtener el mejor tiempo de respuesta de la aplicación realizando la distribución de manera que la carga computacional tienda a ser equitativa (*balanceada*) en el tiempo. Este es uno de los aspectos centrales del procesamiento paralelo, pues tiene un impacto directo en el uso eficiente de los recursos (que implican costo) y las mejoras de performance alcanzables.

La performance obtenida en el sistema paralelo está dada por una compleja relación en la que intervienen una gran cantidad de factores: el tamaño del problema, la arquitectura de soporte, la distribución de procesos en procesadores, la existencia o no de un algoritmo de balanceo de carga, etc. Existe un gran número de métricas para evaluar sistemas paralelos. Entre ellas se encuentran el tiempo efectivo, speedup, eficiencia, producto procesador-tiempo, overhead paralelo, grado de concurrencia, escalabilidad, isoeficiencia, isospeed, etc.

En esta línea de investigación es de interés la evaluación de performance de distintas clases de aplicaciones sobre las arquitecturas disponibles, de modo de adecuar los resultados teóricos de las métricas a la realidad. Esto se basa en que muchos sistemas paralelos no alcanzan su capacidad teórica, y las causas de esta degradación son muchas y no siempre fáciles de determinar. El análisis permite estudiar el impacto que tienen algunos de estos factores sobre las implementaciones, y adecuar las métricas clásicas a las mismas. En particular, se estudia la influencia de las estrategias de distribución de procesos y datos, y la carga (estática o dinámica) asignada a cada procesador sobre el speedup, la eficiencia y la escalabilidad. Más allá de las características teóricas/algorítmicas de las aplicaciones, el rendimiento real y por lo tanto el tiempo necesario para resolver los problemas siempre o en la mayoría de los casos dependerá de la arquitectura de procesamiento elegida para la implementación.

Los temas presentados se enmarcan en el Proyecto “Procesamiento Concurrente y Paralelo” del Instituto de Investigación en Informática LIDI, y existen convenios de cooperación en estos temas de interés común con varias Universidades del país y del exterior.

## **Líneas de Investigación y desarrollo**

- Algoritmos paralelos. Paralelización de aplicaciones, incluyendo su adecuación a diferentes modelos de arquitectura. Optimización de algoritmos.
- Paradigmas de programación paralela orientados a la arquitectura de soporte.
- Análisis de los problemas de migración y asignación óptima de procesos y datos a procesadores. Migración dinámica. Balance de carga estático y dinámico.
- Estudio de las diferentes métricas de evaluación de performance de sistemas paralelos. Abarca el análisis de la complejidad de los algoritmos, el rendimiento de las soluciones paralelas y la caracterización de los modelos de performance asociados.
- Modelos de predicción de performance en sistemas paralelos. Análisis de la respuesta de los modelos para clases de arquitecturas reconocidas.
- Desarrollo de aplicaciones en clusters dedicados y no dedicados. Análisis de la eficiencia de las soluciones.
- Modelos de Memoria en arquitecturas paralelas (compartida, parcialmente compartida, totalmente distribuida).
- Simulación de arquitecturas VLSI orientadas a multiprocesamiento a nivel circuito integrado, tendientes a migrar a hardware procesos de administración de recursos compartidos propios de las capas bajas de los sistemas operativos. Lenguajes de simulación/especificación.
- Aplicación de las investigaciones en áreas como el procesamiento de datos numéricos en cómputo científico, el procesamiento de imágenes digitales, las bases de datos distribuidas, reconocimiento de patrones en bases de datos de secuencias (por ejemplo en los casos de investigación sobre genoma, ADN), reconocimiento de patrones de identificación (rostros, huellas digitales).

## **Resultados obtenidos/esperados**

- Desarrollo de soluciones a los problemas mencionados sobre diferentes modelos de arquitectura (cluster, multiprocesados con memoria compartida distribuida tipo SGI Origin 2000 y cubo de transputers). Evaluar la eficiencia, rendimiento y speedup obtenido.
- Resolver problemas concretos que requieren procesamiento paralelo en el área médica, en el área de tratamiento de señales en tiempo real, en cómputo numérico y en procesamiento de bases de datos distribuidas. Transferir resultados.
- Plantear modelos de arquitectura VLSI orientados a multiprocesamiento “on-chip” y estudiar su comportamiento teórico mediante simulación. Analizar la migración de algoritmos de software para tratamiento de señales en tiempo real a hardware dedicado.
- Formar recursos humanos en los temas de Procesamiento Concurrente y Paralelo, incluyendo Tesis de Postgrado.

## Formación de Recursos Humanos

En el marco de esta línea de investigación se ha concluido una Tesis Doctoral, y se espera concluir otras tres a la brevedad (1 en 2004 y 2 en 2005). En todos los casos hay Directores/Codirectores del exterior.

Asimismo, se desarrollan permanentemente Tesinas de Grado de Licenciatura en Informática (unas 5 por año) vinculadas con los temas de investigación y desarrollo presentados.

## Bibliografía Básica

- [AKL 97] Akl S, "Parallel Computation. Models and Methods", Prentice-Hall, Inc., 1997.
- [ALPE95] Alpern B., L. Carter, J. Ferrante, "Space-limited procedures: A methodology for portable high-performance", International Working Conference on Massively Parallel Programming Models, 1995.
- [ANDR00] Andrews. "Foundations of Multithreaded, Parallel and Distributed Programming", Addison Wesley, 2000
- [BAKE99a] Baker M., G. Fox, "Metacomputing: Harnessing Informal Supercomputers", in R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 154-185, 1999.
- [BANI98] Banikazemi M., V. Moorthy, D. Panda, "Efficient Collective Communication on Heterogeneous Networks of Workstations", Proc. International Conference on Parallel Processing, pp. 460-467, 1998.
- [BARA99] Barak A., La'adan O., Shiloh A., "Scalable Cluster Computing with MOSIX for LINUX". 1999.
- [CAMP99] Campos L.M., Scherson I., "Rate of Change Load Balancing in Distributed and Parallel Systems", Proceedings of the 13th International Parallel Processing Symposium and 10th Symposium on Parallel and Distributed Processing, San Juan, Puerto Rico, 1999, 701-707
- [CAST96] Castleman, K, "Digital Image Processing." Prentice Hall.1996
- [CHAN88] Chandy, Misra, "Parallel Program Design. A Foundation", Addison Wesley, 1988.
- [CHIO99] Chiola G., G. Ciaccio, "Lightweighth Messaging Systems", in R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 246-269, 1999.
- [COMMXX] Communications of the ACM (colección de revistas)
- [COMPXX] IEEE Computer (colección de revistas)
- [COMSXX] IEEE Communications society (colección de revistas)
- [CONCXX] IEEE Concurrency (colección de revistas)
- [CORR99] Corradi A., Leonardi L., Zambonelli F., "Diffusive Load-Balancing Policies for Dynamic Applications", IEEE Concurrency, 7(1), Jan-Mar 1999, 22-31.
- [ELEAXX] ACM eLearn (colección de revistas)
- [EVOLXX] IEEE Evolutionary Computation (colección de revistas)
- [FITC02] Fitch A.J., Kadyrov A., Christmas W.J., Kittler J., "Fast exhaustive robust matching," in: Proceedings of 16th International Conference on Pattern Recognition, Vol. 3, 2002,
- [GEIS94] Geist AI, Beguelin A., Dongarra J., Jiang W., Manchek R., Sunderam V., "PVM: Parallel Virtual Machine. A Users' Guide and Tutorial for Networked Parallel Computing ", The MIT Press, Cambridge, Massachusetts. 1994.
- [JAIN00] Jain A.K., Duin R.P.W., Jianchang M., "Statistical pattern recognition: a review," IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4-37.

- [KAFI98] Kafil M., Ahmad I., "Optimal Task Assignment in Heterogeneous Distributed Computing Systems", IEEE Concurrency, 6(3), Jul-Sep 1998, 42-51
- [KUMA94] Kumar V., Grama A., Gupta A., Karypis G., "Introduction to Parallel Computing. Design and Analysis of Algorithms", Benjamin/Cummings, 1994.
- [LEE 97] Lee C-H., Shin K., "Optimal Task Assignment in Homogeneous Networks", IEEE Transactions on Parallel and Distributed Systems, 8(2), February 1997, 119-129
- [LEIG92] Leighton F. T., "Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes", Morgan Kaufmann Publishers, 1992.
- [LEOP01] Leopold C., "Parallel and Distributed Computing. A survey of Models, Paradigms, and Approaches", Wiley Series on Parallel and Distributed Computing. Albert Zomaya Series Editor, 2001
- [LIAN00] Lian D-R., Tripathi S., "On Performance Prediction of Parallel Computations with Precedent Constraints", IEEE Transactions on Parallel and Distributed Systems, 11(5), May 2000, 491-508
- [MPI 97] MPI Forum, "MPI-2: Extension to the MPI", University of Tennessee Knoxville, 1997.
- [MURA95] Murat Tekalp, A., "Digital Image Processing Algorithms", Prentice Hall. 1995.
- [SIGAXX] IEEE Sigact News (colección de revistas)
- [SNIE96] Snier M., Otto S., Huss-Lederman S., Walker D., Dongarra J., "MPI: The Complete Reference", The MIT Press, Cambridge, Massachusetts, 1996.
- [SUN 02] Sun X-H., "Scalability versus Execution Time in Scalable Systems", Journal of Parallel and Distributed Computing, 62(), 2002, 173-192
- [SUN 99] Sun X-H., Pantano M., Fahringer T., Zhan Z., "SCALA: A Framework for Performance Evaluation of Scalable Computing", Proceedings of the 4th Workshop on High Level Parallel Programming Models and Supportive Environments (HIPS 99), Lecture Notes in Computer Science No. 1586, Springer Verlag, 1999, 49-62"
- [TINE98] Tinetti F., De Giusti A., "Procesamiento Paralelo. Conceptos de Arquitectura y Algoritmos", Editorial Exacta, 1998.
- [TPDPXX] IEEE Transactions on Parallel and Distributed Processing (colección de revistas)
- [WATT98] Watts J., Taylor S., "A Practical Approach to Dynamic Load Balancing", IEEE Transactions on Parallel and Distributed Systems, 9(3), March 1998, 235-248
- [WILK99] Wilkinson B., Allen M., Parallel Programming: Techniques and Applications Using Networking Workstations, Prentice-Hall, Inc., 1999.
- [ZOMA96] Zomaya A., "Parallel Computing. Paradigms and Applications", Int. Thomson Computer Press, 1996.