

## **ESTUDIO DE RENDIMIENTO DE APLICACIONES PARALELAS PARA ARQUITECTURAS DE TIPO GPU**

**MARCELO AARON ESPINA ROSAS  
INGENIERO CIVIL EN COMPUTACION**

### **RESUMEN**

En esta memoria se evalúa el uso de tarjetas de video Nvidia con arquitectura CUDA para la resolución de problemas de alto desempeño computacional. El modelo del Daisyworld se utilizó como un problema de referencia para poner a prueba las capacidades de esta arquitectura. El documento comprende un estudio de los principales aspectos relacionados con la computación de alto desempeño, con la evaluación de programas paralelos, así como de la arquitectura CUDA en particular, detallando la API y el modelo de programación relacionados. Asimismo, se describen las limitaciones a que se ve enfrentado el programador cuando trabaja con esta arquitectura. En los experimentos se trabajó con tres implementaciones diferentes de la simulación del Daisyworld: una implementación secuencial, una aplicación paralela que usa pthreads, y una implementación paralela que opera sobre los cores de la tarjeta de video usando la API de CUDA. Los resultados obtenidos permiten concluir que el tiempo de comunicación entre la memoria del computador y la memoria de la tarjeta de video afectan en gran medida el rendimiento de los programas. Por lo tanto no todos los programas paralelos pueden sacar provecho de arquitecturas como CUDA. En particular, es necesario que la cantidad de trabajo que llevan a cabo los cores de la tarjeta de video sea relativamente alto, para compensar el costo de la transmisión de los datos. Los resultados nos permiten concluir que las tarjetas con arquitectura CUDA son una alternativa altamente viable y relativamente fácil de trabajar cuando se trata de problemas de alto desempeño computacional, aunque todavía quedan algunas limitaciones que se deben resolver.

## ABSTRACT

This thesis evaluates the use of Nvidia video cards with the CUDA architecture for solving problems that require a high computational performance. The Daisyworld model was used as a benchmark problem to test the capacities of this architecture. The document includes a survey of the main aspects of high performance computing, particularly the evaluation of parallel programs and CUDA architecture, detailing the API and related programming model. It also describes the constraints faced by the developer when working with this architecture. In the experiments, we worked with three different implementations of the Daisyworld simulation: a sequential implementation, a parallel implementation that uses pthreads, and a parallel implementation that operates on the core video card using the CUDA API. The results show that the communication time between the computer's memory and the memory of the video card greatly affect the performance of programs. So not all programs can take advantage of parallel architectures like CUDA. In particular, it is necessary that the amount of work carried out by cores of the video card is relatively high, too set the cost of data transmission. The results demonstrate that the cards with the CUDA architecture are highly viable alternatives and are relatively easy to work with when solving high performance computing problems, although there are still some limitations that must be resolved.