



#### Escola Tècnica Superior d'Enginyeria Informàtica Universitat Politècnica de València

# El producto matricial distribuido en entornos computacionales de alto rendimiento

TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

Autor: Rodrigo Huerta Gañán

Tutor: Pedro Alonso Jordá

Curso 2019-2020

#### Resum

????

#### Resumen

????

**Palabras clave:** ?????, ???, ?????????????

#### **Abstract**

????

**Key words:** ?????, ????? ?????, ??????????

# Índice general

Ín	dice general dice de figuras dice de tablas	V VII VII
1	Introducción general(Por decidir)	1
	1.1 Motivación	. 1
	1.2 Objetivos	. 1
	1.3 Metodología	. 1
	1.4 Organización de la memoria	
	1.5 Convenciones	. 2
2	Estado del arte(Por decidir)	3
	2.1 Estado del arte en entornos no distribuidos	. 3
	2.1.1 Operación mediante cpus	
	2.1.2 Operación mediante gpus	
	2.2 Estado del arte en entornos distribuidos	
	2.2.1 Operación mediante cpus	
	2.2.2 Operación mediante gpus	. 3
3	Análisis y mejora del problema	5
4	Primera aproximación mediante M.P.I	7
	4.1 Índices de bloque	. 7
	4.1.1 Row major order	. 7
	4.1.2 Column major order	. 7
	4.2 Distribución y recuperación de matrices	
	4.3 Algoritmo S.U.M.M.A	. 7
5	Solución del problema en GPUs	9
	5.1 Modelos de programación y elección de uno	. 9
	5.2 Abstracción de GPUS lógicas	
	5.3 Principales modificaciones	
	5.4 Mejoras de rendimiento	. 9
6	Creación de librerías	11
	6.1 C++(Por decidir)	. 11
	6.2 Matlab	. 11
	6.3 Python	. 11
7	Comparación de rendimiento	13
8	Conclusión	15
9	Trabajo futuro	17
	9.1 Versión conjunta de mpi y nccl con capacidad multinodo	
	9.2 Soporte para arquitecturas con tensorcores	
	9.3 Ampliación de tipos soportados	
	9.4 Ampliación de las operaciones soportadas	
Ril	hliografía	19

# Índice de figuras

#### Índice de tablas

### CAPÍTULO 1 Introducción general(Por decidir)

En este primer capítulo de presentación de la memoria, se introduce al lector en la temática del trabajo de fin de grado, empezando con la sección de motivación, en la que se describe el interés del tema. En el segundo apartado, objetivos, se explica que se pretende conseguir con este trabajo. A continuación se explicara la metodología empleada y la organización seguida en la memoria. Finalmente se expondrán las convenciones que se seguirán en el documento.

4	-	B 4	. •	• /
L.	. 1	IVI	otiva	cion

- 1.2 Objetivos
- 1.3 Metodología
- 1.4 Organización de la memoria
- 1.5 Convenciones

# CAPÍTULO 2 Estado del arte(Por decidir)

Explicación.

#### 2.1 Estado del arte en entornos no distribuidos

- 2.1.1. Operación mediante cpus
- 2.1.2. Operación mediante gpus
- 2.2 Estado del arte en entornos distribuidos
- 2.2.1. Operación mediante cpus
- 2.2.2. Operación mediante gpus

## CAPÍTULO 3 Análisis y mejora del problema

#### **CAPÍTULO 4**

#### Primera aproximación mediante M.P.I

Explicación.

#### 4.1 Índices de bloque

- 4.1.1. Row major order
- 4.1.2. Column major order
- 4.2 Distribución y recuperación de matrices
- 4.3 Algoritmo S.U.M.M.A

## CAPÍTULO 5 Solución del problema en GPUs

5.1	Modelos de programación y elección de uno
5.2	Abstracción de GPUS lógicas
5.3	Principales modificaciones
5.4	Mejoras de rendimiento

#### CAPÍTULO 6 Creación de librerías

- 6.1 C++(Por decidir)
- 6.2 Matlab
- 6.3 Python

## CAPÍTULO 7 Comparación de rendimiento

#### CAPÍTULO 8 Conclusión

# CAPÍTULO 9 Trabajo futuro

- 9.1 Versión conjunta de mpi y nccl con capacidad multinodo
- 9.2 Soporte para arquitecturas con tensorcores
- 9.3 Ampliación de tipos soportados
- 9.4 Ampliación de las operaciones soportadas

# Bibliografía