

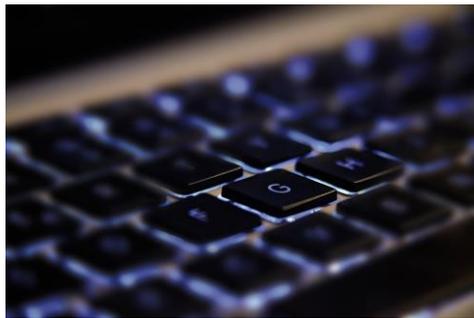
¿Por qué debemos seguir trabajando en álgebra lineal?

Sandra Catalán Pallarés

Universitat Jaume I

¿Qué es el álgebra lineal?

Rama de las matemáticas que se ocupa del estudio de los espacios vectoriales y las transformaciones lineales entre ellos.

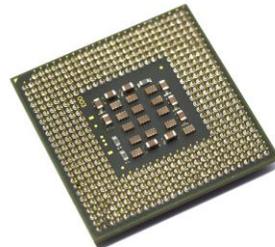


Software para álgebra lineal

Aplicaciones

LAPACK

BLAS



Software para álgebra lineal

BLAS

Basic Linear Algebra Subprograms

LAPACK

Linear Algebra PACKage

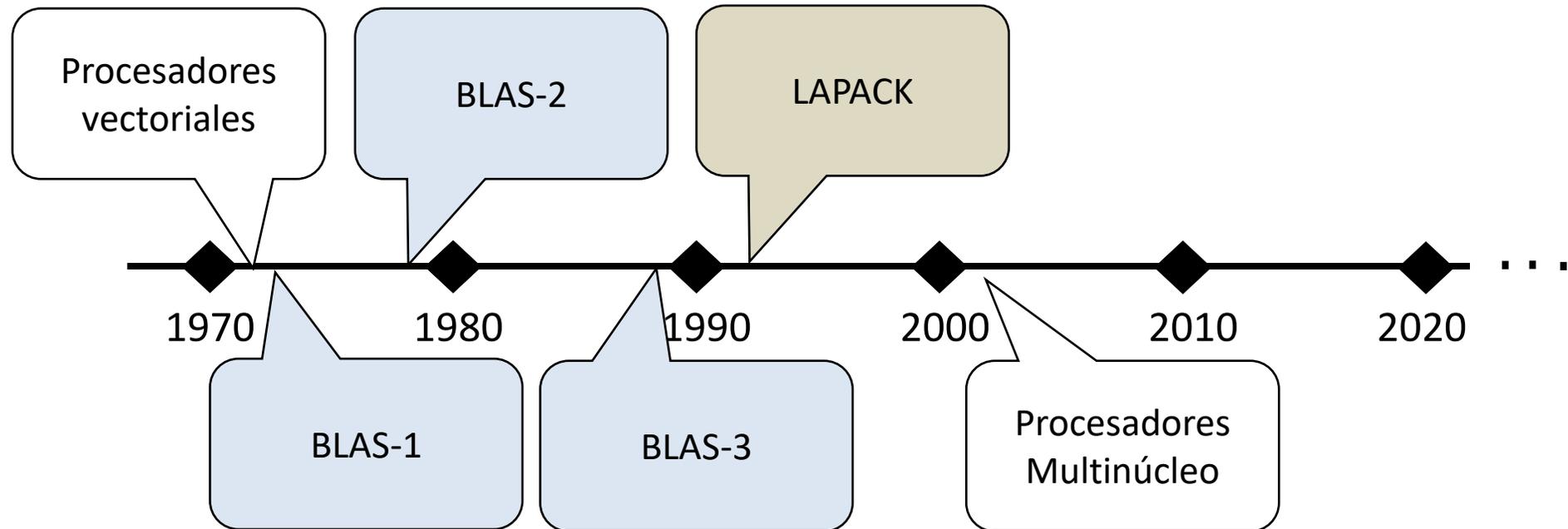
Estándar de facto

Implementaciones específicas: Intel MKL, AMD ACML, IBM ESSL...

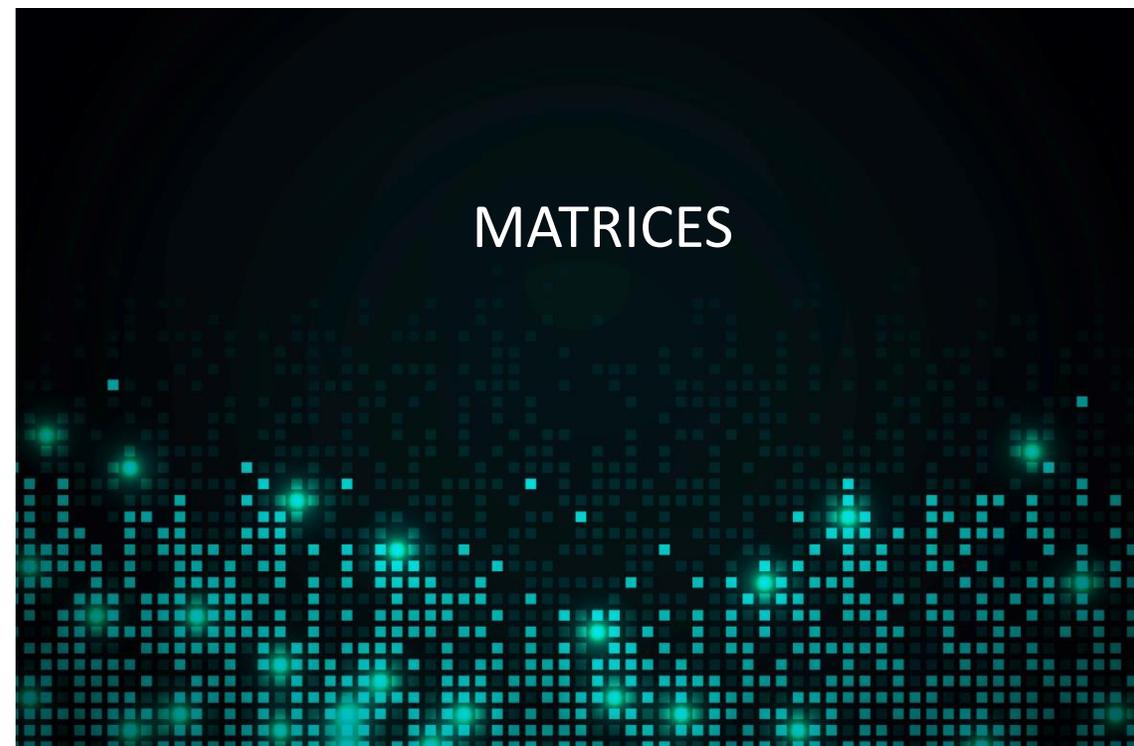
Rutinas para realizar operaciones sencillas como el producto matriz-vector

Rutinas para realizar operaciones complejas como factorizaciones de matrices o resolución de sistemas lineales

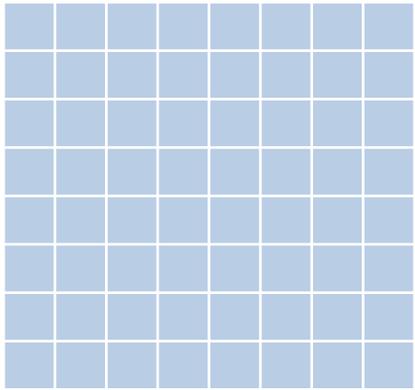
Historia



¿De qué vamos a hablar?



Matrices y el producto de matrices



$$B = \begin{pmatrix} 1 & 5 & 2.3 \\ -2 & 0 & 0 \\ 3 & -1.1 & 2 \end{pmatrix}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}$$

Representación y manipulación de imágenes

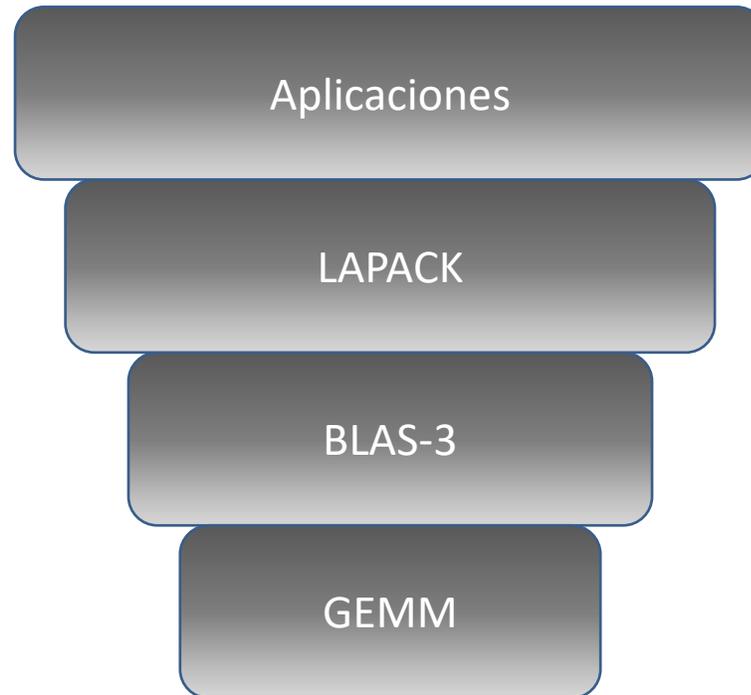
Redes neuronales

Optimización y análisis de datos

Resolución de sistemas de ecuaciones lineales

Procesamiento de señales

Software para álgebra lineal

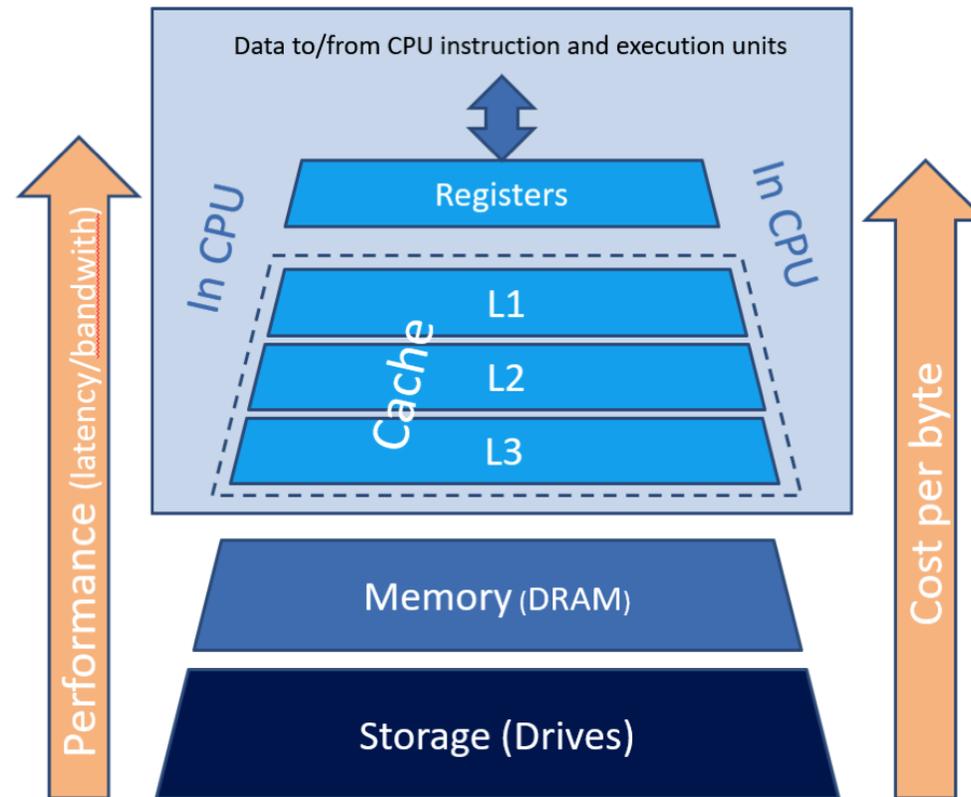


General Matrix-Matrix Multiplication (GEMM)

```
for (i=0; i<m; i++)  
{  
    for (j=0; j<n; j++)  
    {  
        C[i][j]=0;  
        for (l=0; l<k; l++)  
        {  
            C[i][j]+=a[i][l]*b[l][j];  
        }  
    }  
}
```

- ✓ Resultado correcto
- ✗ Bajo rendimiento

Jerarquía de memoria



- <https://www.cadalyst.com/design-related-technologies/workstations/traditional-computer-memory-hierarchy-%E2%80%94-and-its-impact-cad->

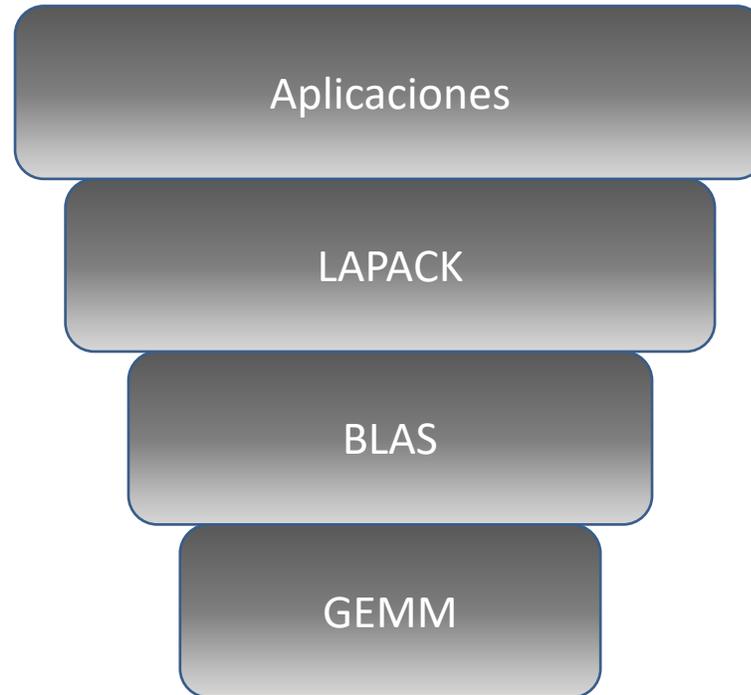
GEMM Optimizada

```
1 void blocked_GEMM( matrix C, A, B;
2                   int m, n, k,
3                   mc, nc, kc, mr, nr )
4 {
5     for (jc=0; jc<n; jc+=nc)           // Loop G1
6         for (pc=0; pc<k; pc+=kc){      // Loop G2
7             // Pack B
8             Bc:=B(pc:pc+kc-1, jc:jc+nc-1);
9             for (ic=0; ic<m; ic+=mc){  // Loop G3
10                // Pack A
11                Ac:=A(ic:ic+mc-1, pc:pc+kc-1);
12                for (jr=0; jr<nc; jr+=nr) // Loop G4
13                    for (ir=0; ir<mc; ir+=mr) // Loop G5
14                        mukernel( C(ic+ir:ic+ir+mr-1,
15                                  jc+jr:jc+jr+nr-1),
16                                  Ac(ir:ir+mr-1,0:kc-1),
17                                  Bc(0:kc-1,jr:jr+nr-1),
18                                  mr, nr, kc );
19            }
20        }
21 }
```

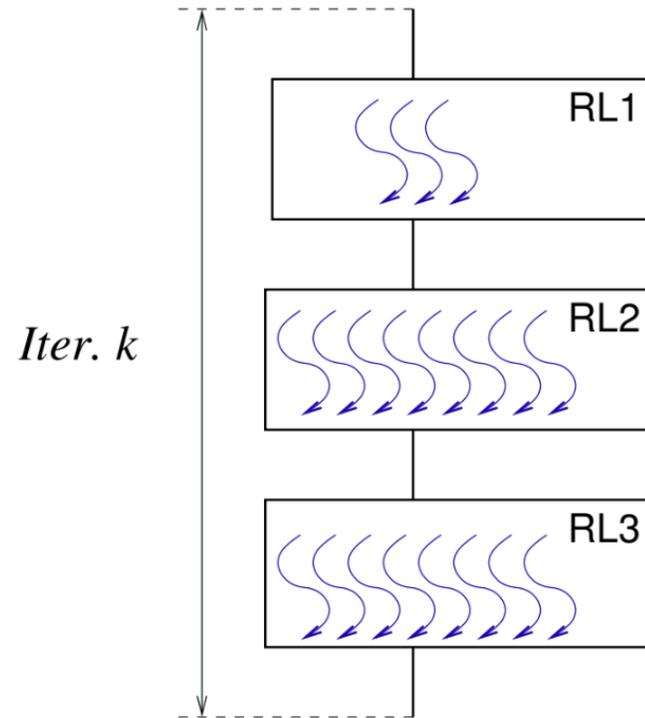
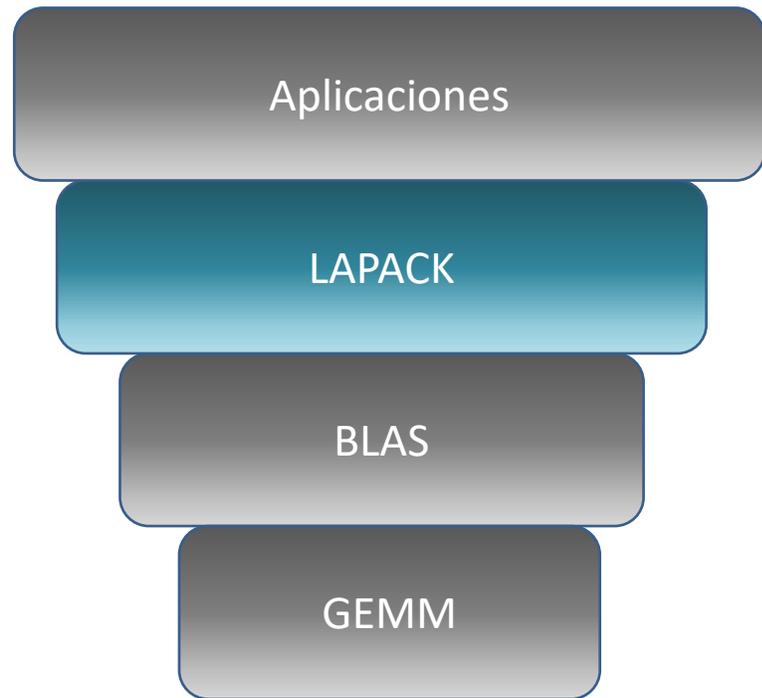
```
1 void mukernel( matrix Cr, Ar, Br;
2               int mr, nr, kc )
3 {
4     for (pr=0; pr<kc; pr++) // Loop M1
5         Cr += Ar(0:mr-1,pr)
6             * Br(pr,0:nr-1);
7 }
```

- ✓ Resultado correcto
- ✓ Buen rendimiento

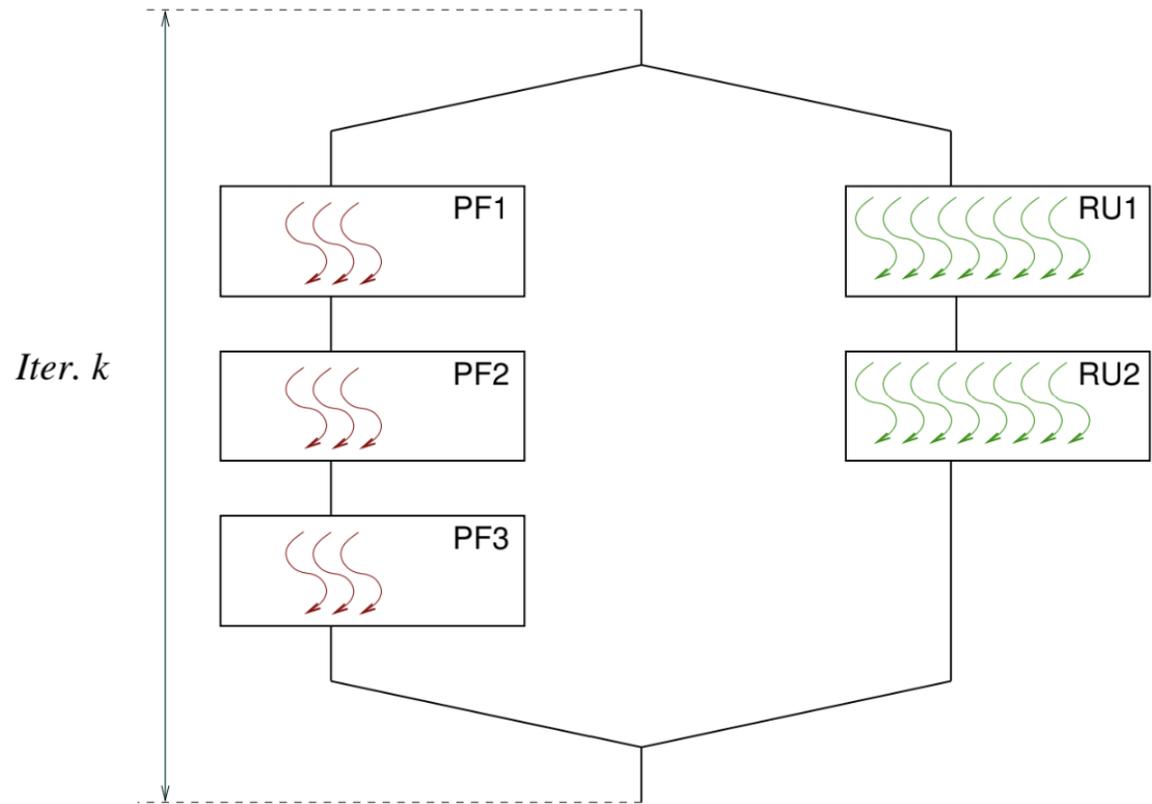
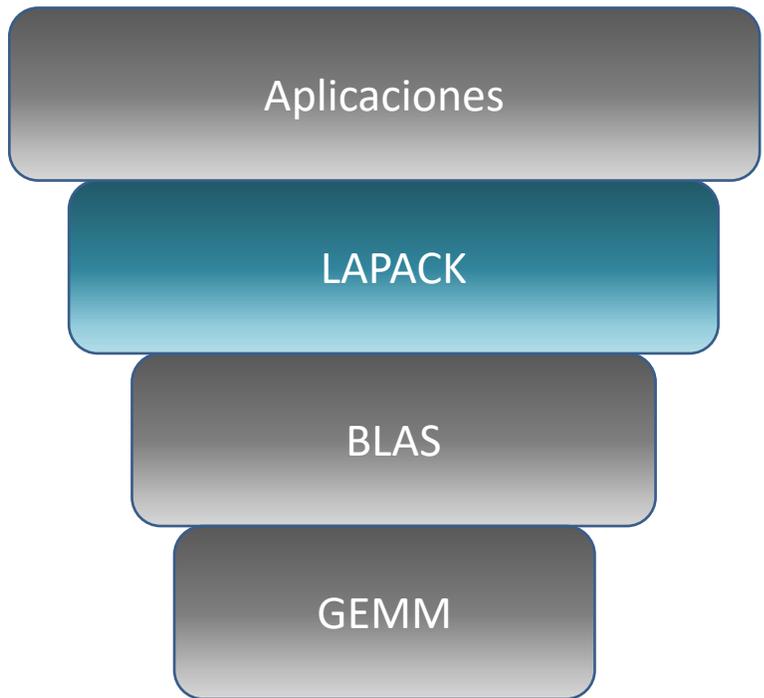
Optimizaciones en LAPACK



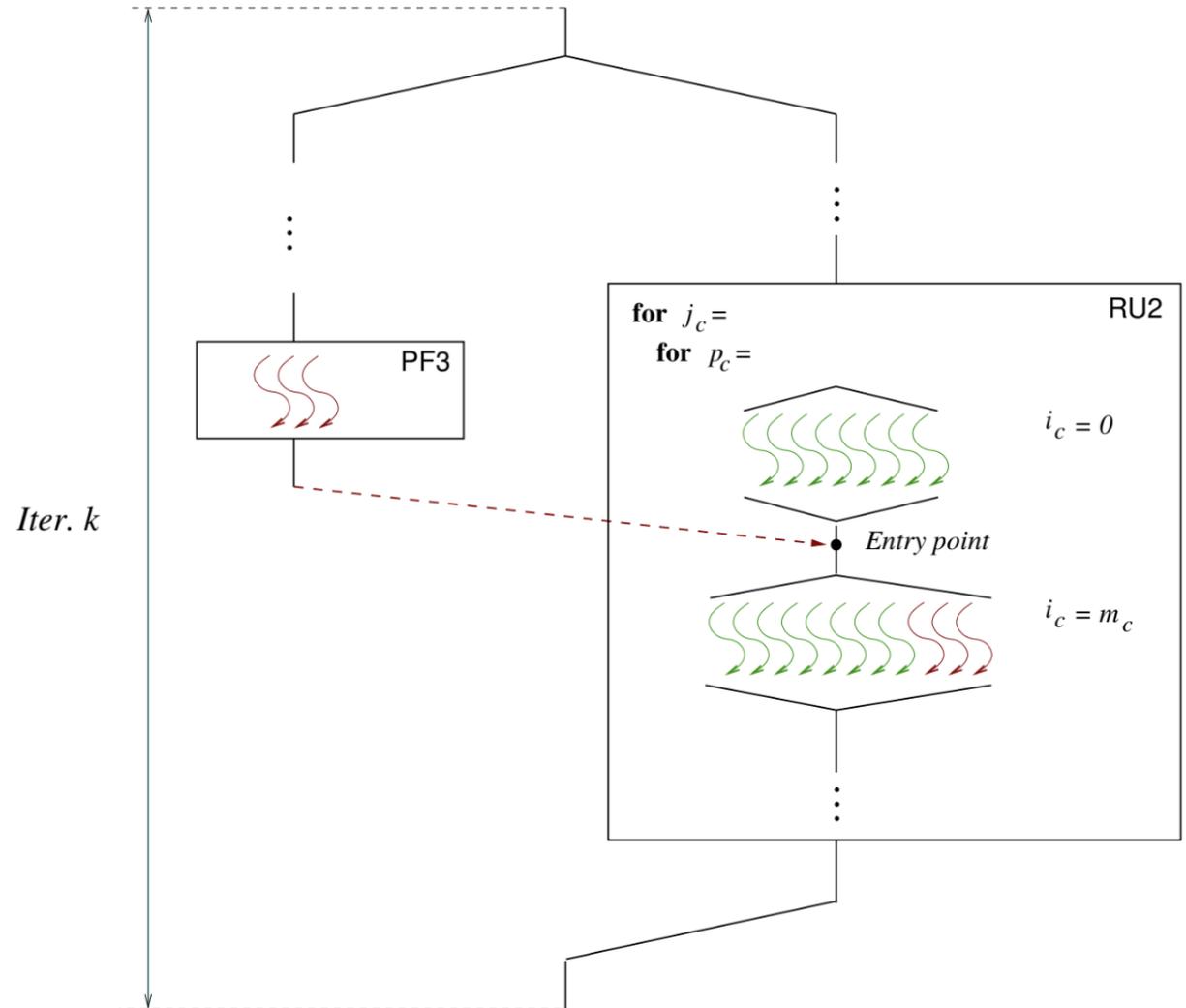
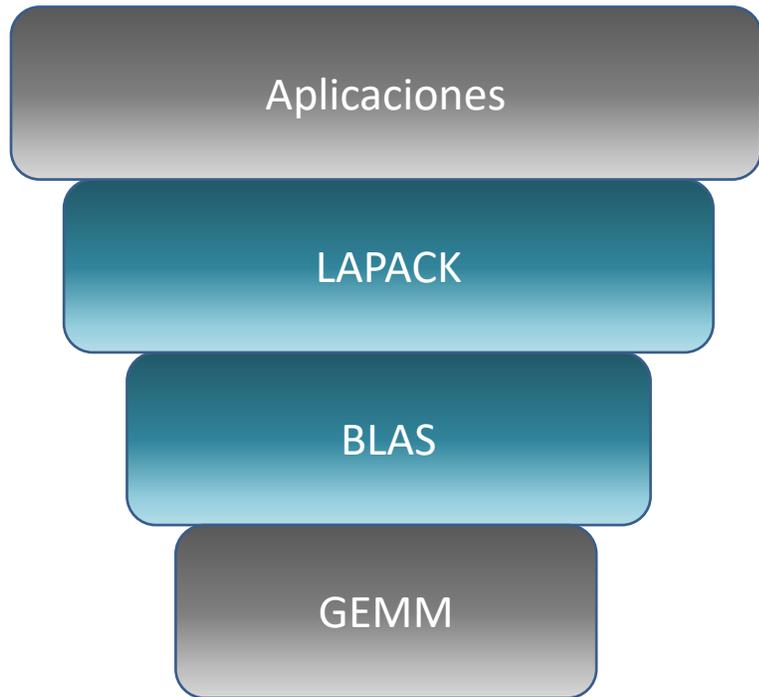
Aproximación tradicional: Fork-join



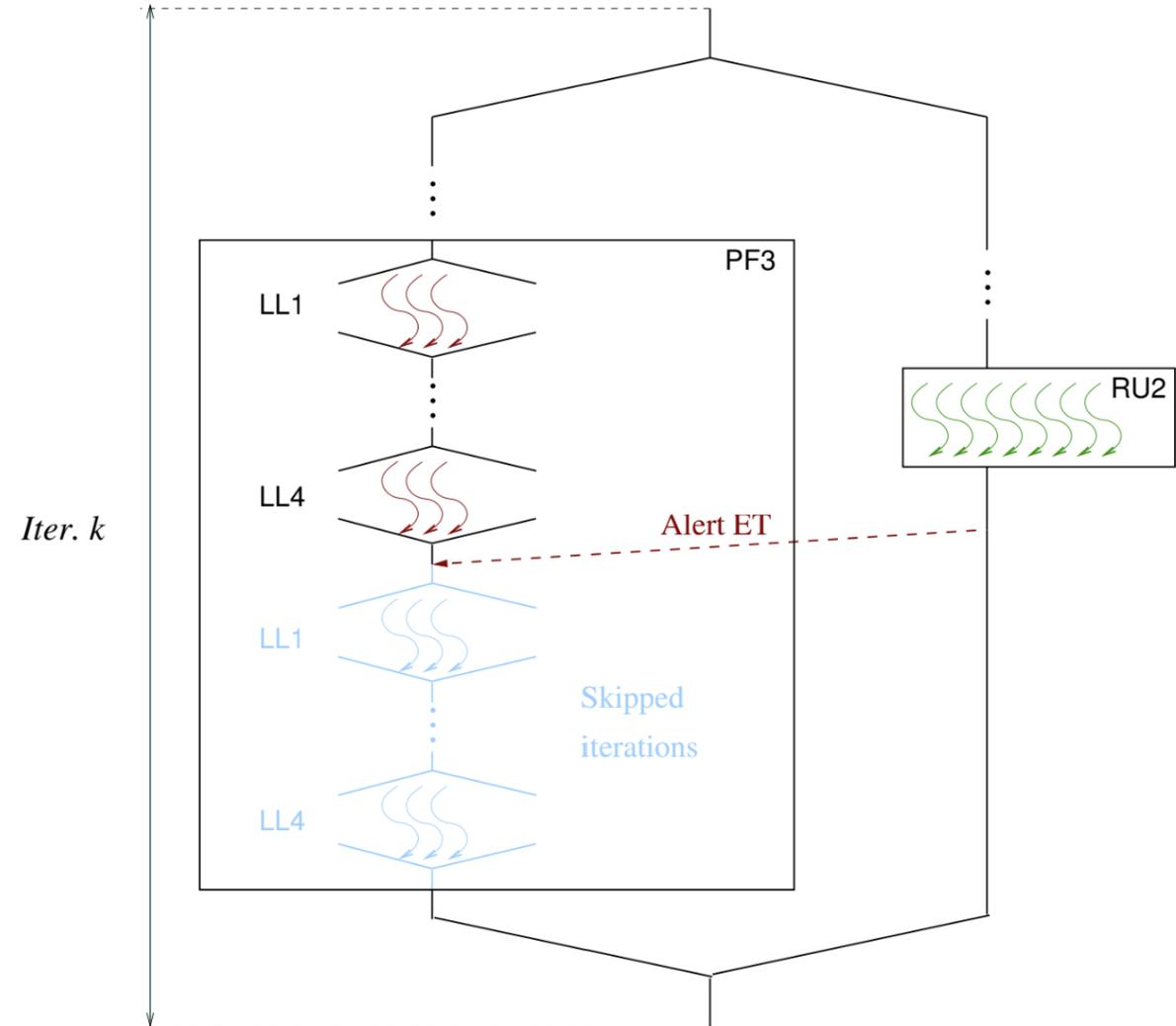
Alternativa: look-ahead



Alternativa: maleabilidad

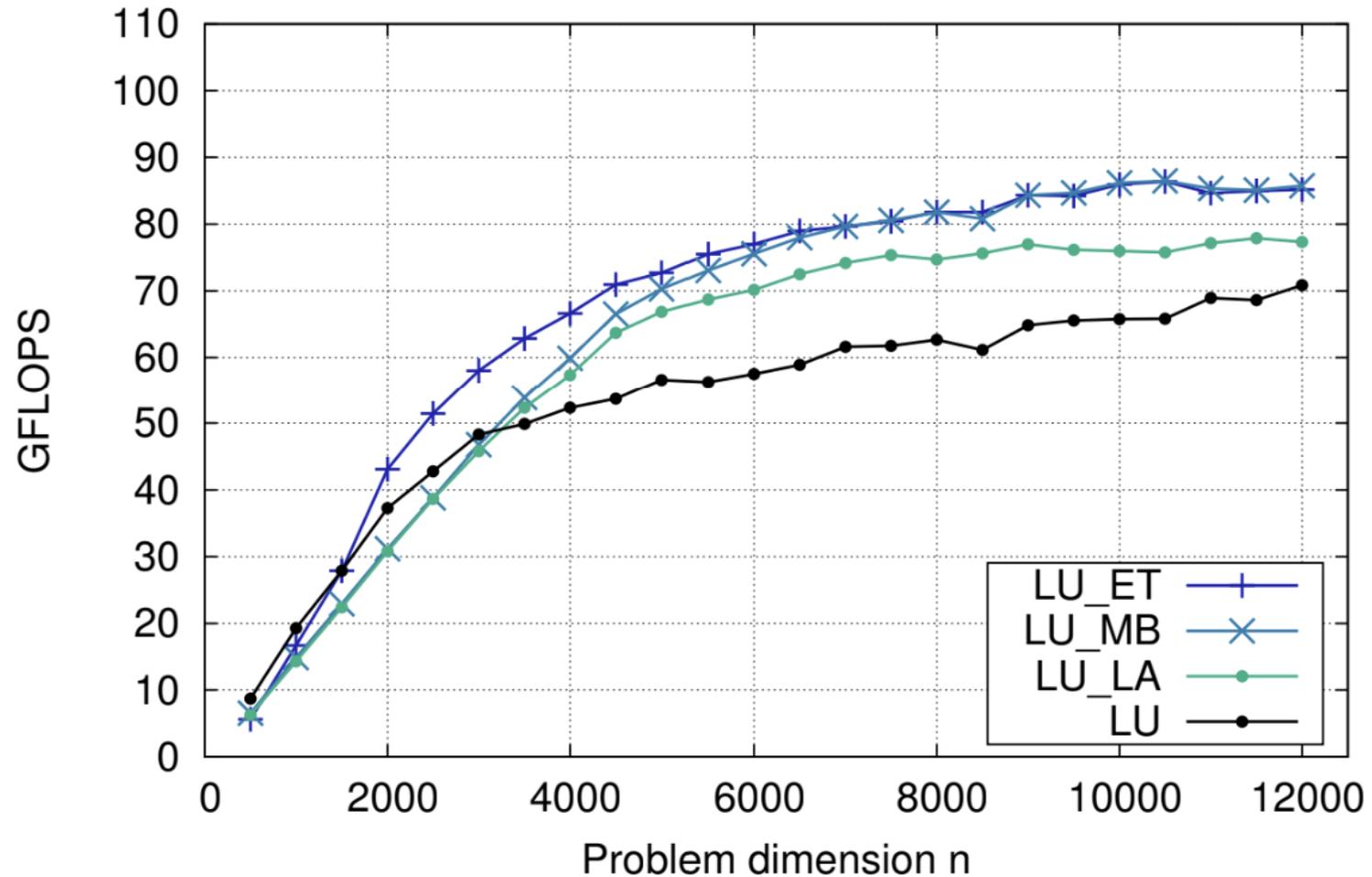


Alternativa: maleabilidad + early termination

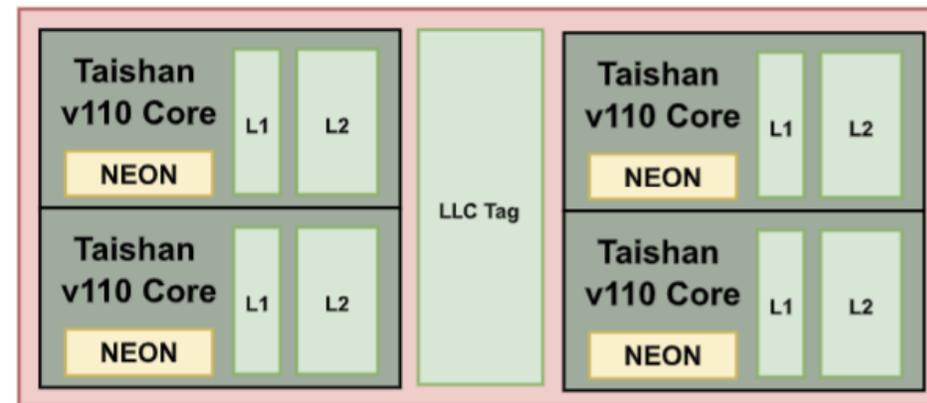


Resultados

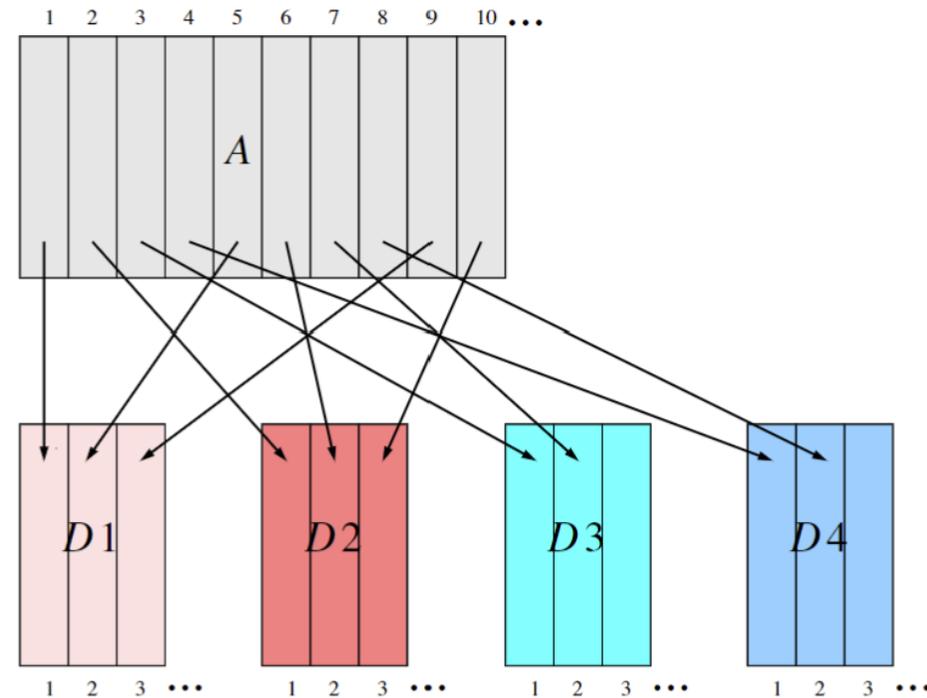
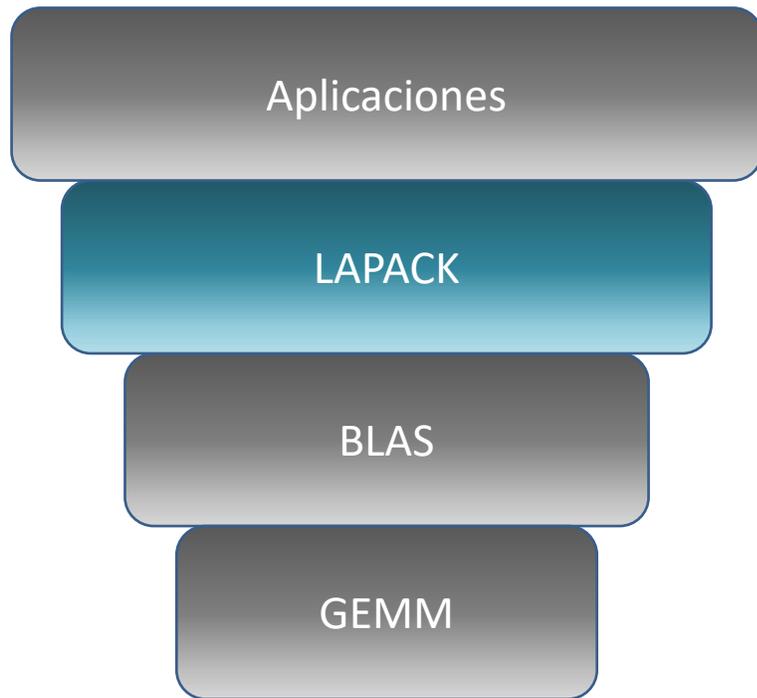
LU on Intel Xeon E5-2603 v3



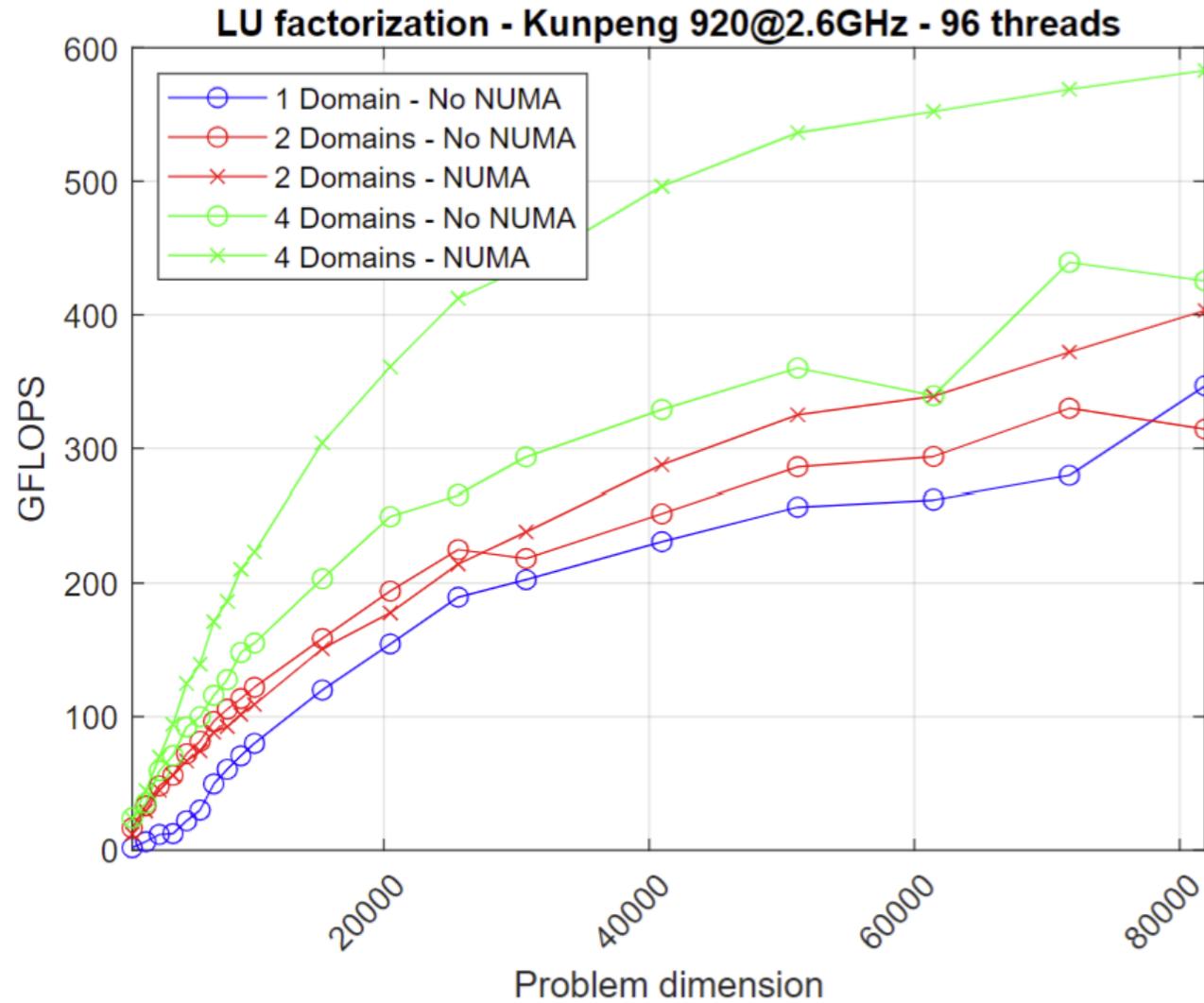
Optimizaciones en LAPACK: arquitecturas NUMA



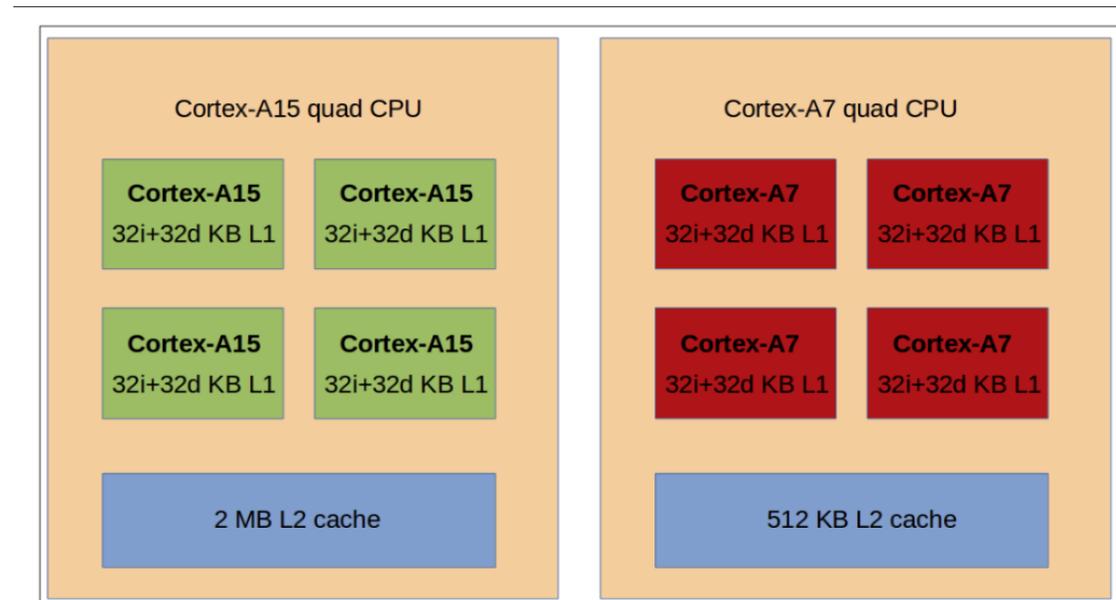
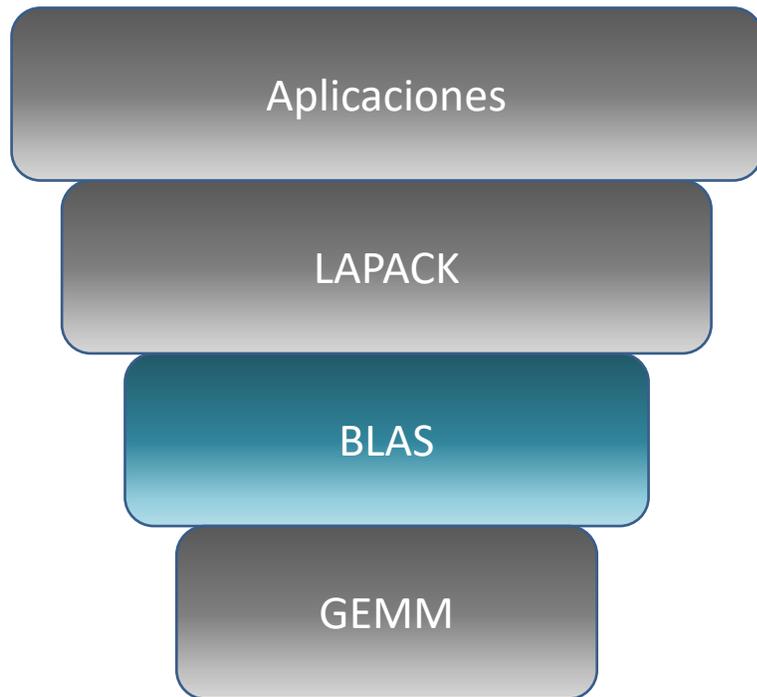
Optimizaciones en LAPACK: arquitecturas NUMA



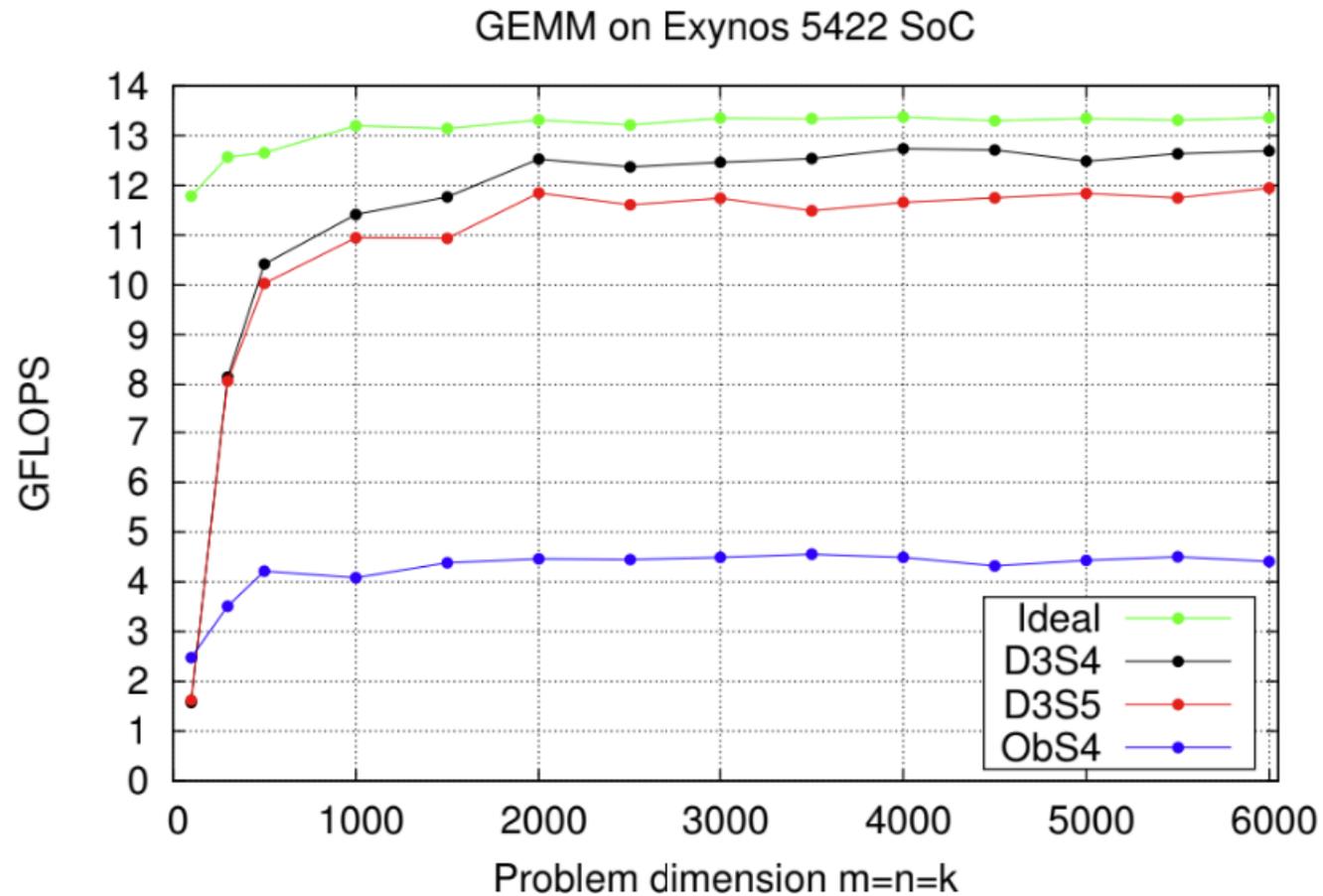
Resultados



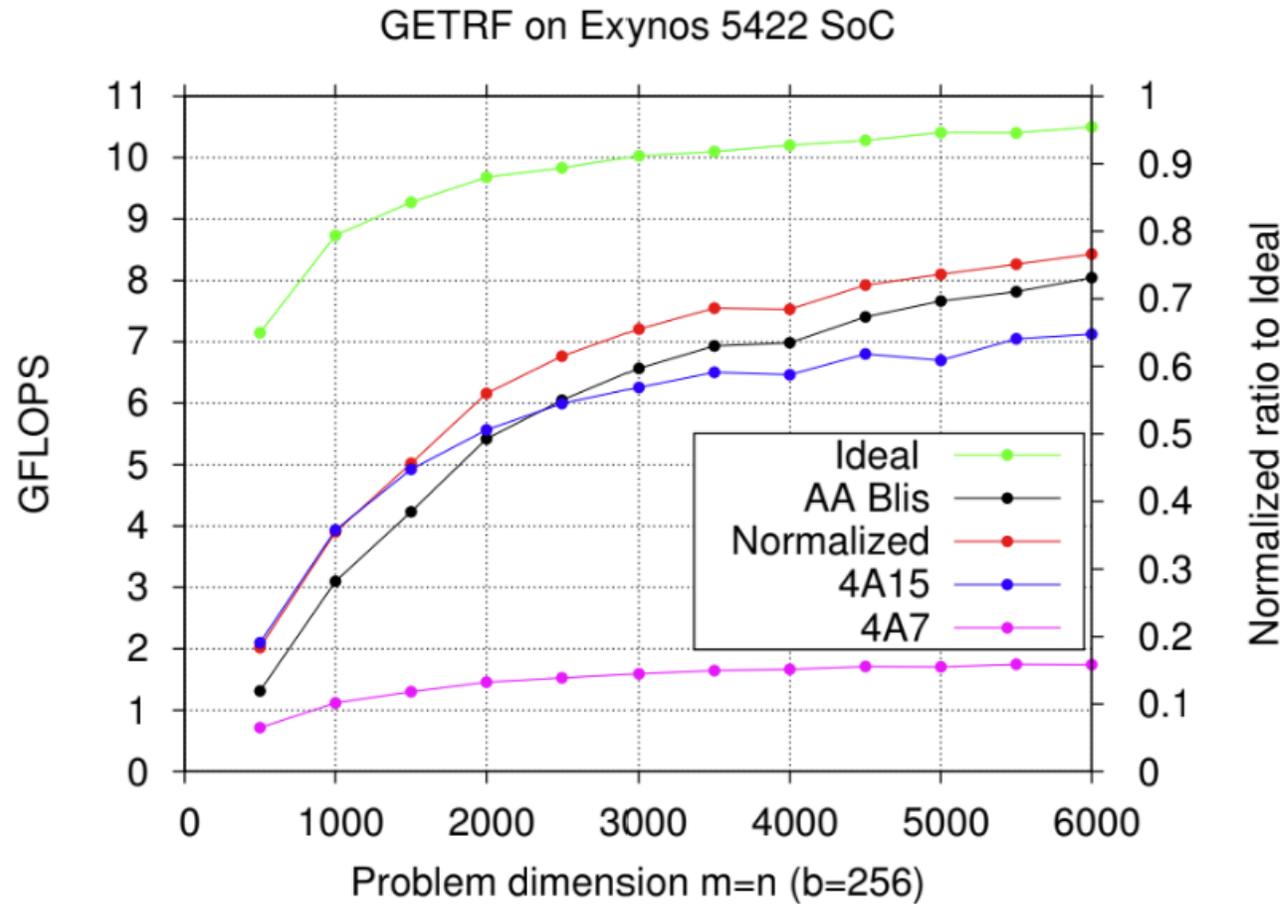
Optimizaciones en BLAS: procesadores asimétricos



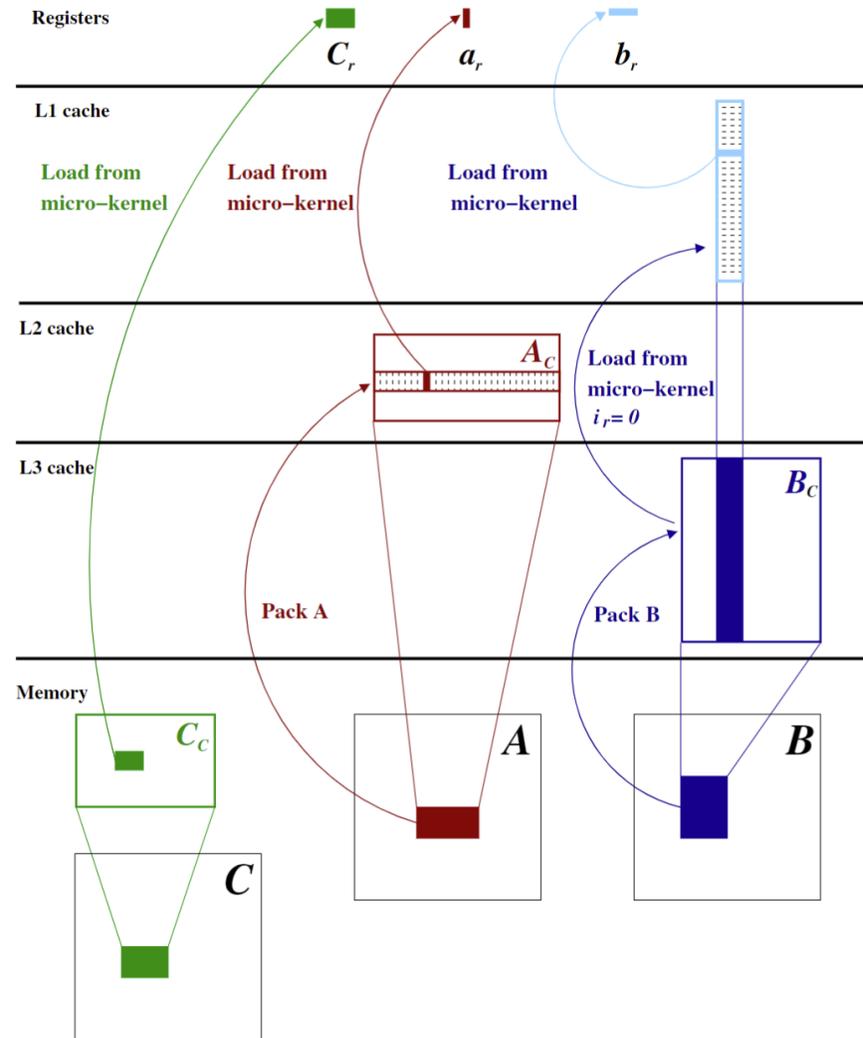
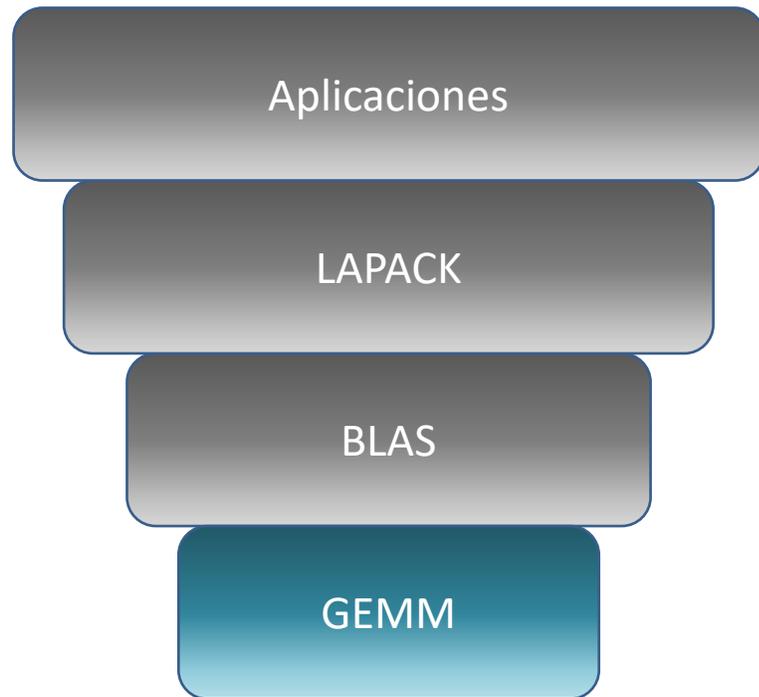
Resultados



Resultados

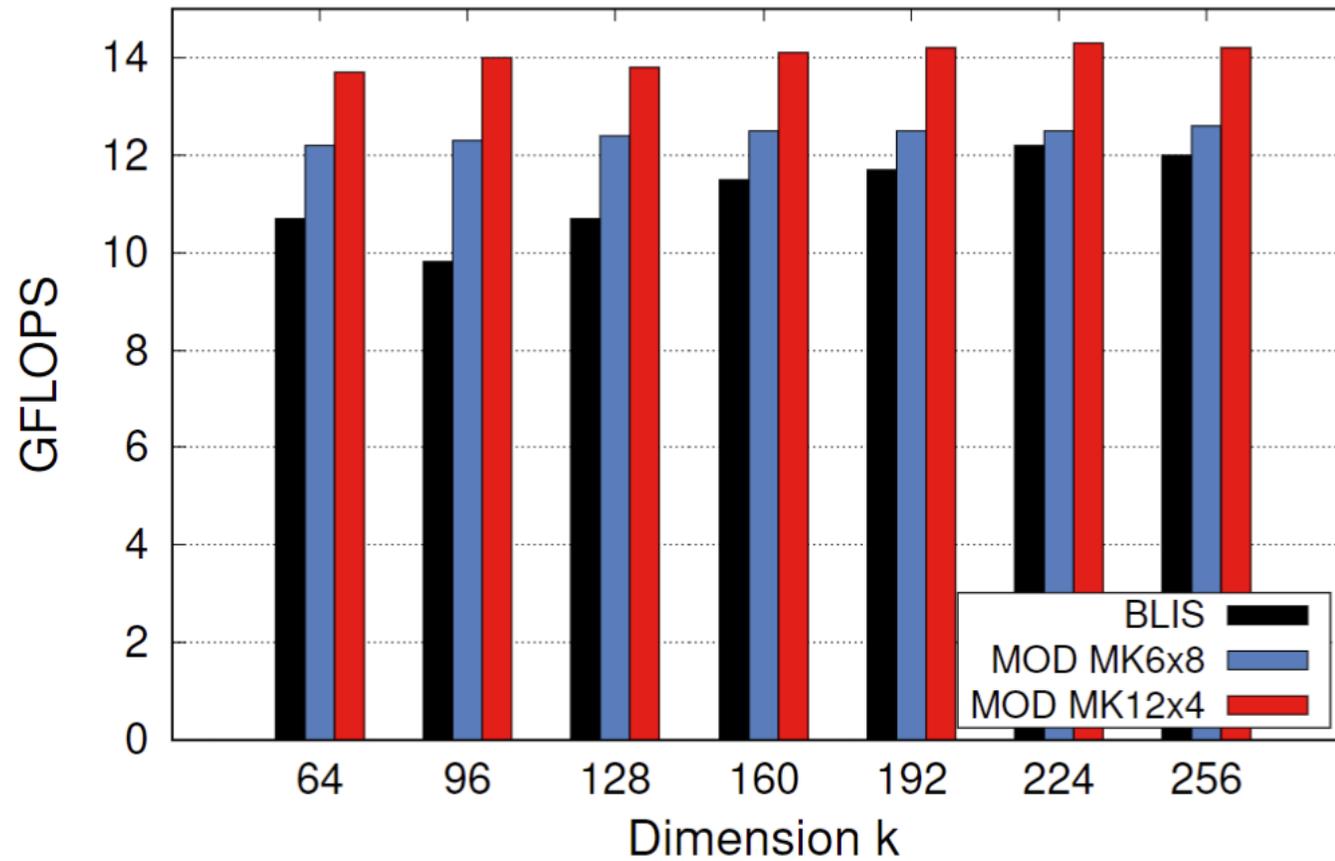


Optimización de GEMM: parámetros de cache y micro-kernels

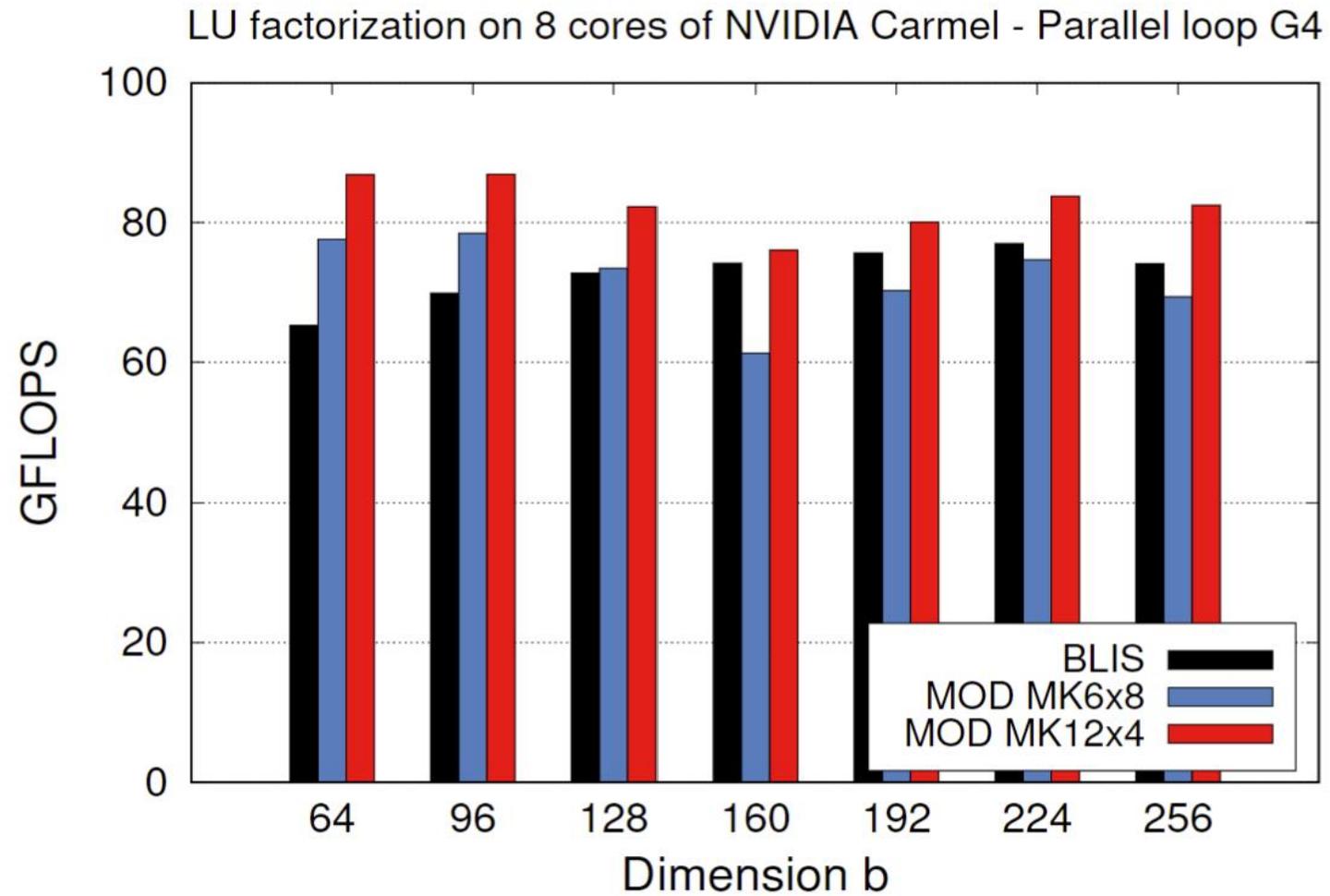


Resultados

GEMM on 1 core of NVIDIA Carmel



Resultados



Conclusiones

- Álgebra lineal utilizada en muy diversos campos
- Optimizaciones posibles a distintos niveles
- Optimizaciones en función del algoritmo, hardware y forma de los operandos
- Trabajo futuro: nuevas arquitecturas, nuevos algoritmos, consumo energético

Muchas gracias por vuestra atención.
¿Alguna pregunta?