High-performance Matrix Computations

Prof. **Paolo Bientinesi**

pauldj@aices.rwth-aachen.de

GEMM

- Let $A \in \mathcal{R}^{m \times k}, B \in \mathcal{R}^{k \times n}, C \in \mathcal{R}^{m \times n}$.
- Consider $C := C + AB$.
- Mathematically, $\forall i \; \forall j \; c_{ij} := c_{ij} + \sum_h a_{ih} b_{hj}$
- Implementation?
  For this course, it does not matter how different algorithms are generated.
  Morale of the story: the higher the level of BLAS, the better.

Triple loops, no BLAS

i,j,h

```
for i = 1 : m,
  for j = 1 : n,
    for h = 1 : k,
```
$$c_{i,j} := c_{i,j} + a_{i,h} b_{h,j}$$

j,i,h

```
for j = 1 : n,
  for i = 1 : m,
    for h = 1 : k,
```
$$c_{i,j} := c_{i,j} + a_{i,h} b_{h,j}$$

i,h,j

```
for i = 1 : m,
  for h = 1 : k,
    for j = 1 : n,
```
$$c_{i,j} := c_{i,j} + a_{i,h} b_{h,j}$$

h,i,j

```
for h = 1 : k,
  for i = 1 : m,
    for j = 1 : n,
```
$$c_{i,j} := c_{i,j} + a_{i,h} b_{h,j}$$

j,h,i

```
for j = 1 : n,
  for h = 1 : k,
    for i = 1 : m,
```
$$c_{i,j} := c_{i,j} + a_{i,h} b_{h,j}$$

h,j,i

```
for h = 1 : k,
  for j = 1 : n,
    for i = 1 : m,
```
$$c_{i,j} := c_{i,j} + a_{i,h} b_{h,j}$$

DOT: BLAS 1

Notation:
a^i = i-th row of A,
a_j = j-th column of A

GEMM as a sequence of inner products (DOTs).

$\forall_i \forall_j \; c_{ij} := c_{ij} + a^i b_j$

$$C := C + AB = C + \left[\begin{array}{c} a^1 \\ \hline \vdots \\ \hline a^m \end{array}\right] [b_1 | \dots | b_n] = \left[\begin{array}{c|c|c} c_{11} + a^1 b_1 & \dots & c_{1n} + a^1 b_n \\ \hline \vdots & \ddots & \\ \hline c_{m1} + a^m b_1 & & c_{mn} + a^m b_n \end{array}\right].$$

```
for i = 1 : m,
  for j = 1 : n,
     c_{i,j} := dot(c_{i,j}, a^i, b_j)
```

```
for j = 1 : n,
  for i = 1 : m,
     c_{i,j} := dot(c_{i,j}, a^i, b_j)
```

GER: Blas 2

GEMM as a sum of outer products (GERs)

$C := C + \sum_h a_h b^h$

$$C := C + AB = C + [a_1 | \ldots | a_k] \begin{bmatrix} b^1 \\ \hline \vdots \\ \hline b^k \end{bmatrix} = C + a_1 b^1 + a_2 b^2 + \cdots + a_k b^k.$$

for h = 1 : k,
    $C := \texttt{ger}(C, a_h, b^h)$

GEMV: BLAS 2

GEMM as a sequence of GEMVs.

$\forall_i \; c^i := c^i + a^i B$

$$C := C + AB = \left[\begin{array}{c} c^1 \\ \hline \vdots \\ \hline c^m \end{array}\right] + \left[\begin{array}{c} a^1 \\ \hline \vdots \\ \hline a^m \end{array}\right] B = \left[\begin{array}{c} c^1 + a^1 B \\ \hline \vdots \\ \hline c^m + a^m B \end{array}\right].$$

for i = 1 : m,
$\quad c^i := \mathtt{gemb}(c^i, a^i, B)$

GEMV: BLAS 2

GEMM as a sequence of GEMVs.

$\forall_j \; c_j := c_j + Ab_j$

$$C := C + AB = [c_1|c_2|\ldots|c_n] + A[b_1|b_2|\ldots|b_n] = [c_1 + Ab_1|c_2 + Ab_2|\ldots|c_n + Ab_n].$$

```
for j = 1 : n,
    c_j := gemb(c_j, A, b_j)
```

GEMM: BLAS 3

$$C := \texttt{gemm}(C, A, B)$$