Cache Behavior Modelling for Codes Involving Banded Matrices

Computer Architecture Group.
Departament of Electronics and Systems.
University of A Coruña.

Authors:
Diego Andrade
Basilio B. Fraguela
Ramón Doallo

Outline

- Introduction.
- The **Probabilistic Miss Equations (PME)** model.
- The PME model extension for codes involving **non-uniform** banded matrices.
- Experimental results.
- Conclusions.

Introduction (I)

• There is a big gap between the processor and the memory speed.

• Memory hierarchies try to cushion this gap.

Introduction (and II)

Methods to study the cache behavior:

- Trace driven simulation.
- Hardware counters.
- Analytical models.
  - The Probabilistic Miss Equations (PME) Model.

The PME model (I)

```
DO IZ=1, NZ, LZ
  ...
  DO I1=1, N1, L1
    ...
    DO I0=1, N0, L0
      A(fA1(IA1)), ..., fAdA(IAdA))
      ...
    END DO
    ...
  END DO
  ...
END DO
```

```
DO IZ=1, NZ, LZ
  ...
  DO I1=1, N1, L1
    if (B(fB1(IB1)), ..., fBdB(IBdB)))
    ...
    DO I0=1, N0, L0
      ...
      A(fA1(B(fB1(IB1))), ..., fAdA(IAdA))
      if (C(fC1(IC1)), ..., fCdC(ICdC)))
      ...
      A(fA1(IA1)), ..., fAdA(IAdA))
      ...
    END DO
    ...
  END DO
  ...
END DO
```

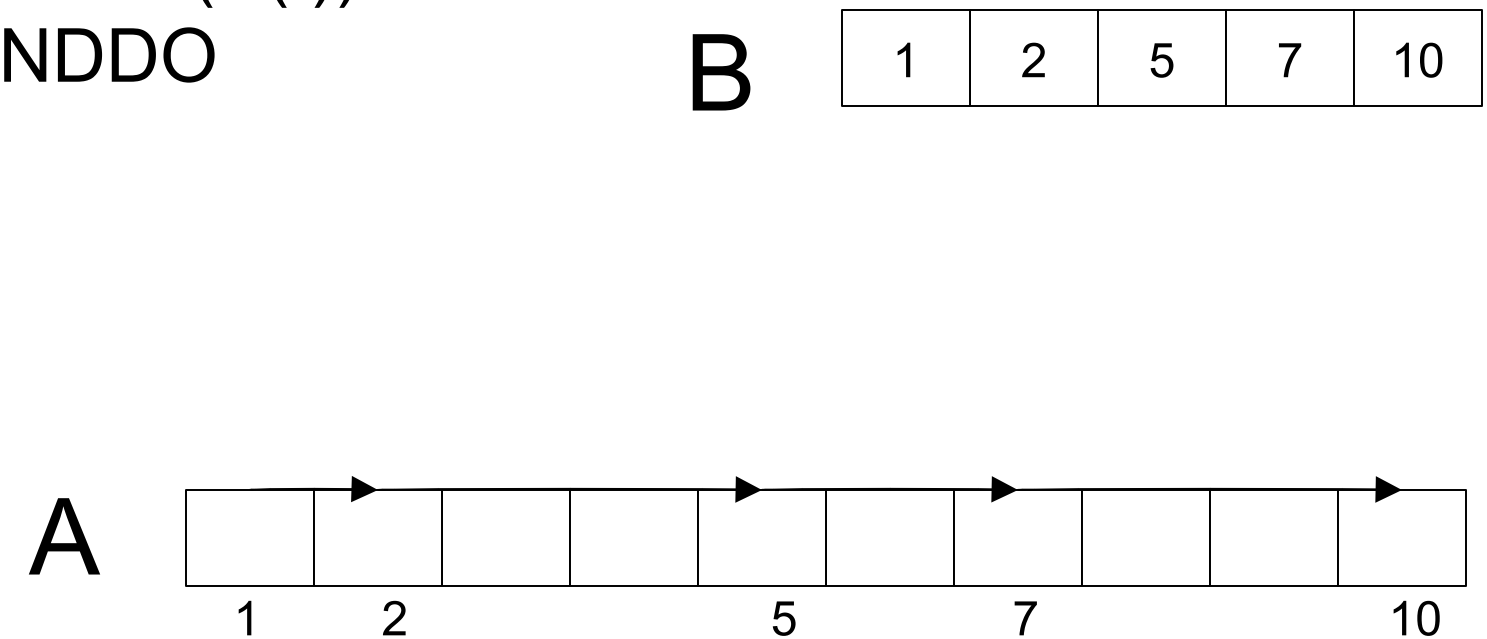
D. Andrade, B.B. Fraguela, R. Doallo. Precise Automatable Analytical Modeling of the Cache Behavior of Codes with Indirections. **Accepted for publication in ACM Transactions on Architecture and Code Optimizations.**

B.B. Fraguela, R. Doallo, E.L. Zapata. Probabilistic Miss Equations: Evaluating Memory Hierarchy Performance. **IEEE Transactions on Computers, 2003**.

D. Andrade, B.B. Fraguela, R. Doallo. Analytical Modeling of Codes with Arbitrary Data-dependent Conditional Structures. **Journal of Systems Architecture, 2006.**

The PME Model (II)

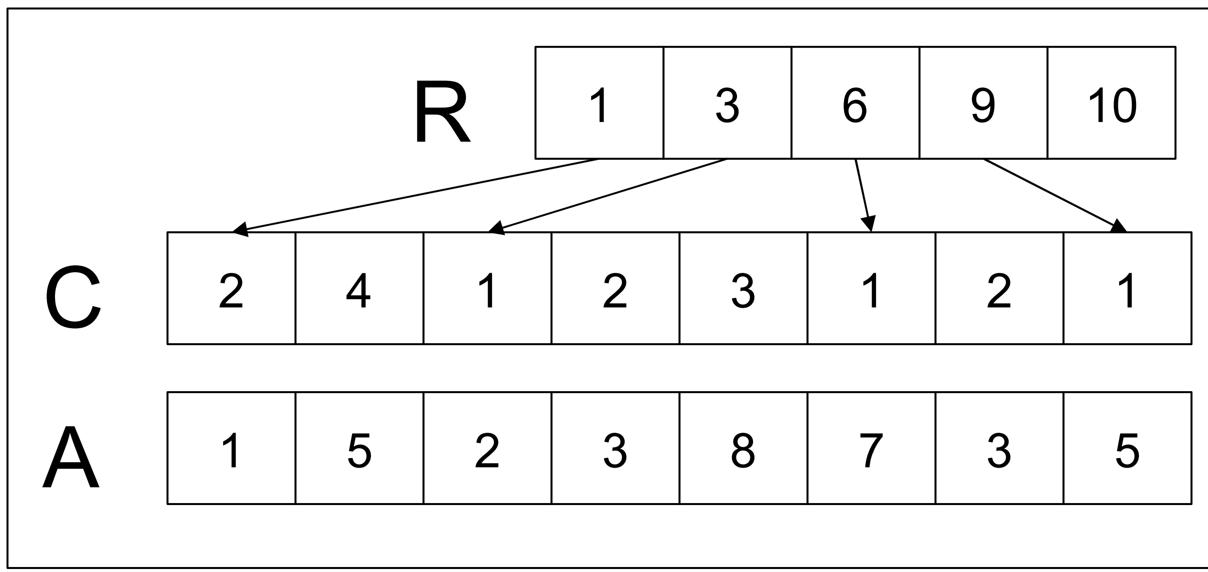
```
DO I=1,5
      A(B(I))
ENDDO
```



The PME Model (III)

Sparse Computation : Compressed Row Storage (**CRS**) Method.

0	1	0	5
2	3	8	0
7	3	0	0
5	0	0	0



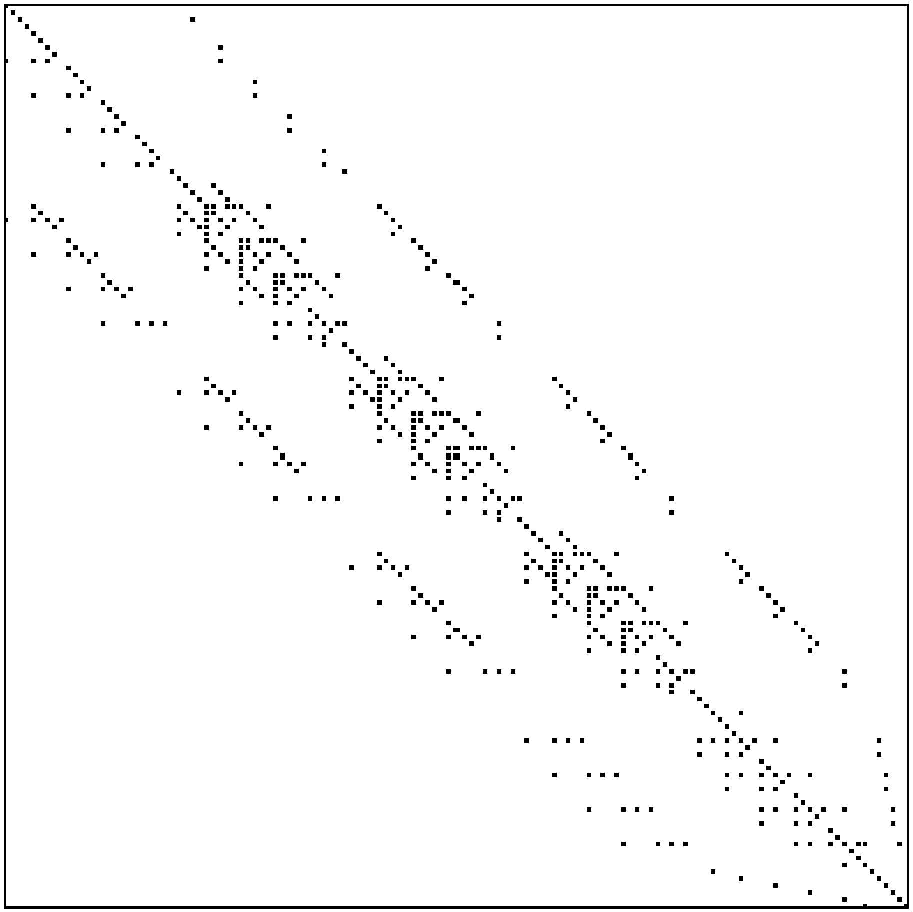
The PME model (IV)

Sparse Matrix-Vector Product

```
DO I=1,M
    REG=0
    DO J=R(I),R(I+1)-1
        REG=REG + A(J) * X(C(J))
    ENDDO
    D(I)=REG
ENDDO
```

PME model (V)

Input data



The PME model (and VI)

Sparse Matrix-Vector Product

```
DO I=1,M
    REG=0
    DO J=R(I),R(I+1)-1
        REG=REG + A(J) * X(C(J))
    ENDDO
    D(I)=REG
ENDDO
```

F_{R0}(Region)

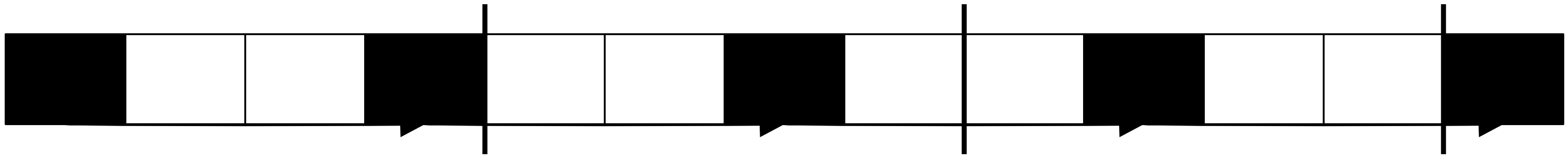
F_{R1}(Region)

A(1)
A(2)
A(3)
A(4)
A(5)
A(6)
A(7)
A(8)
A(9)
A(10)

Line cache size=4

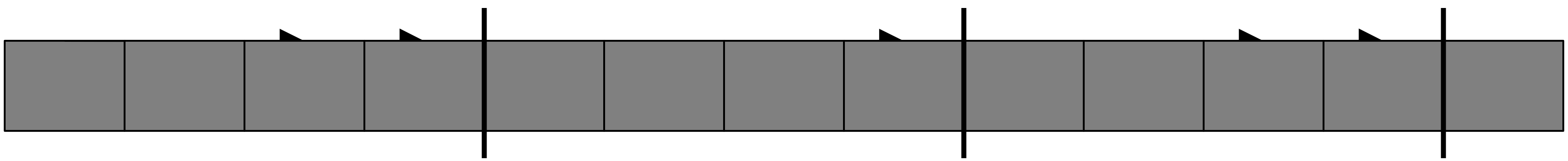
PME model extension involving banded matrices (I)

REGULAR ACCESS PATTERN



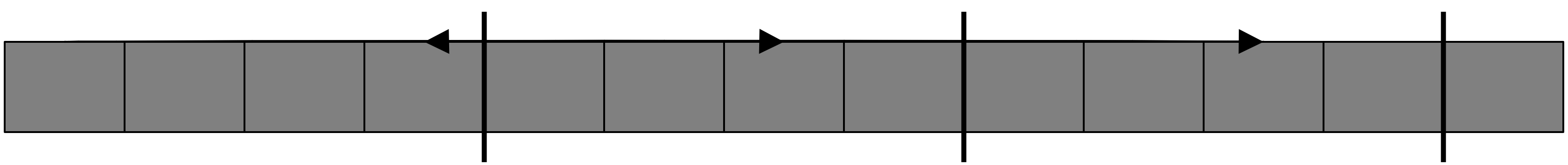
1	3	4	8	11	12

MONOTONIC IRREGULAR ACCESS PATTERN

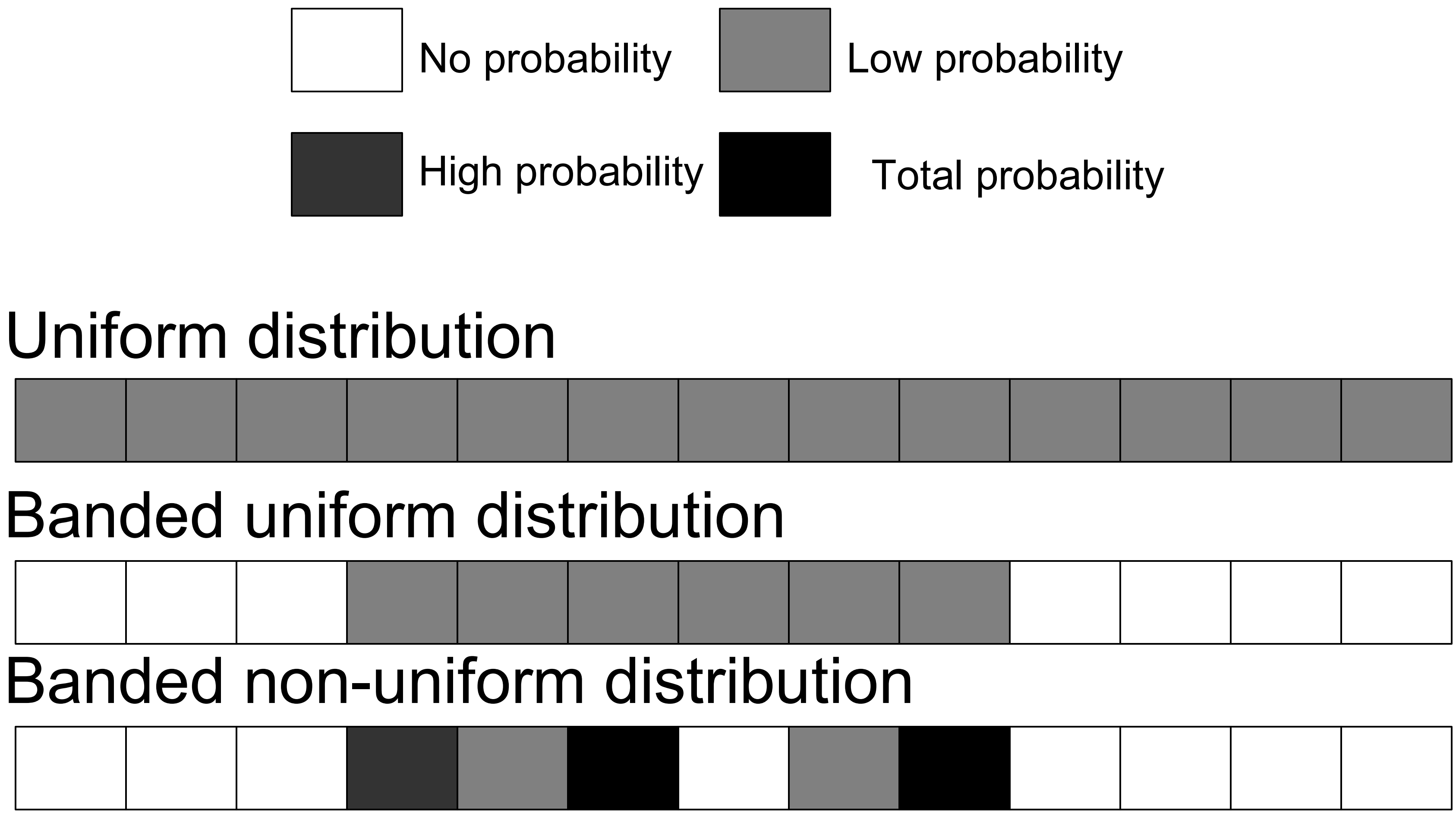


1	11	4	7

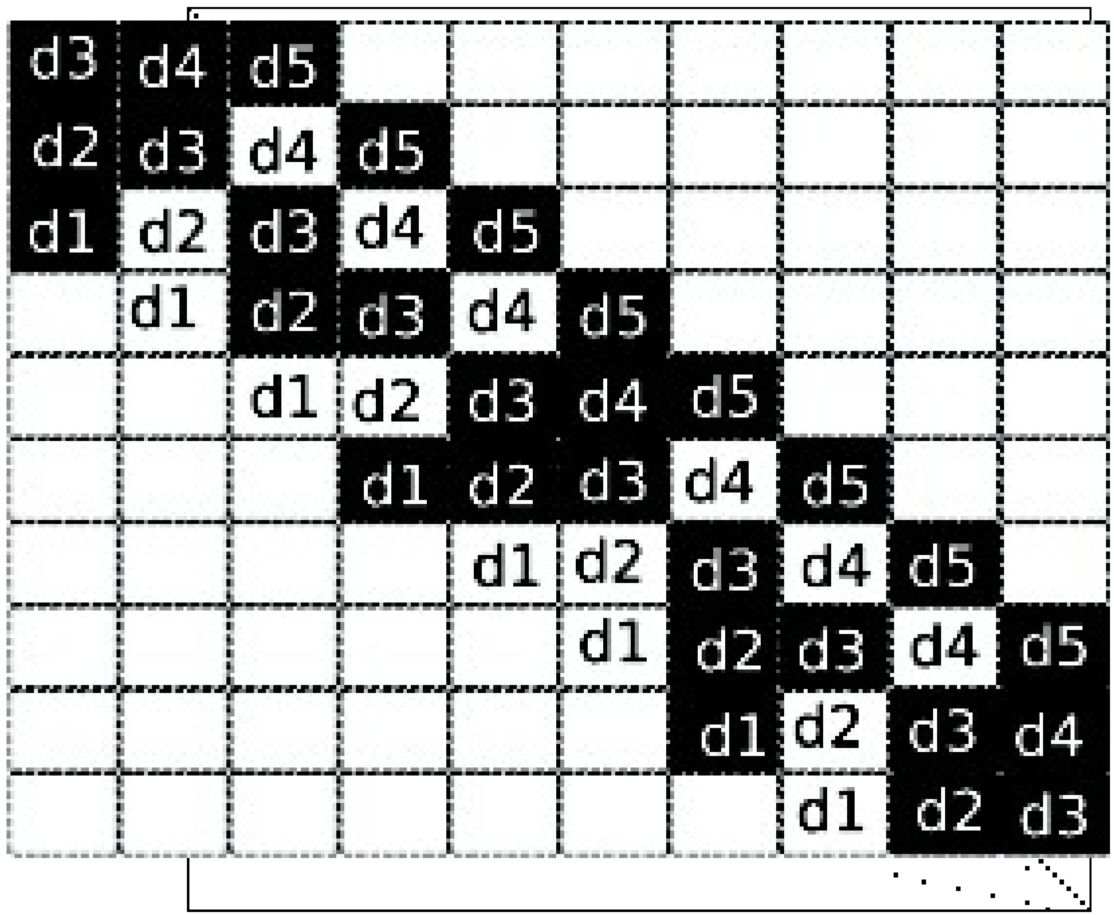
NON-MONOTONIC IRREGULAR ACCESS PATTERN



PME model extension involving banded matrices (II)



PME model extension involving banded matrices (and III)

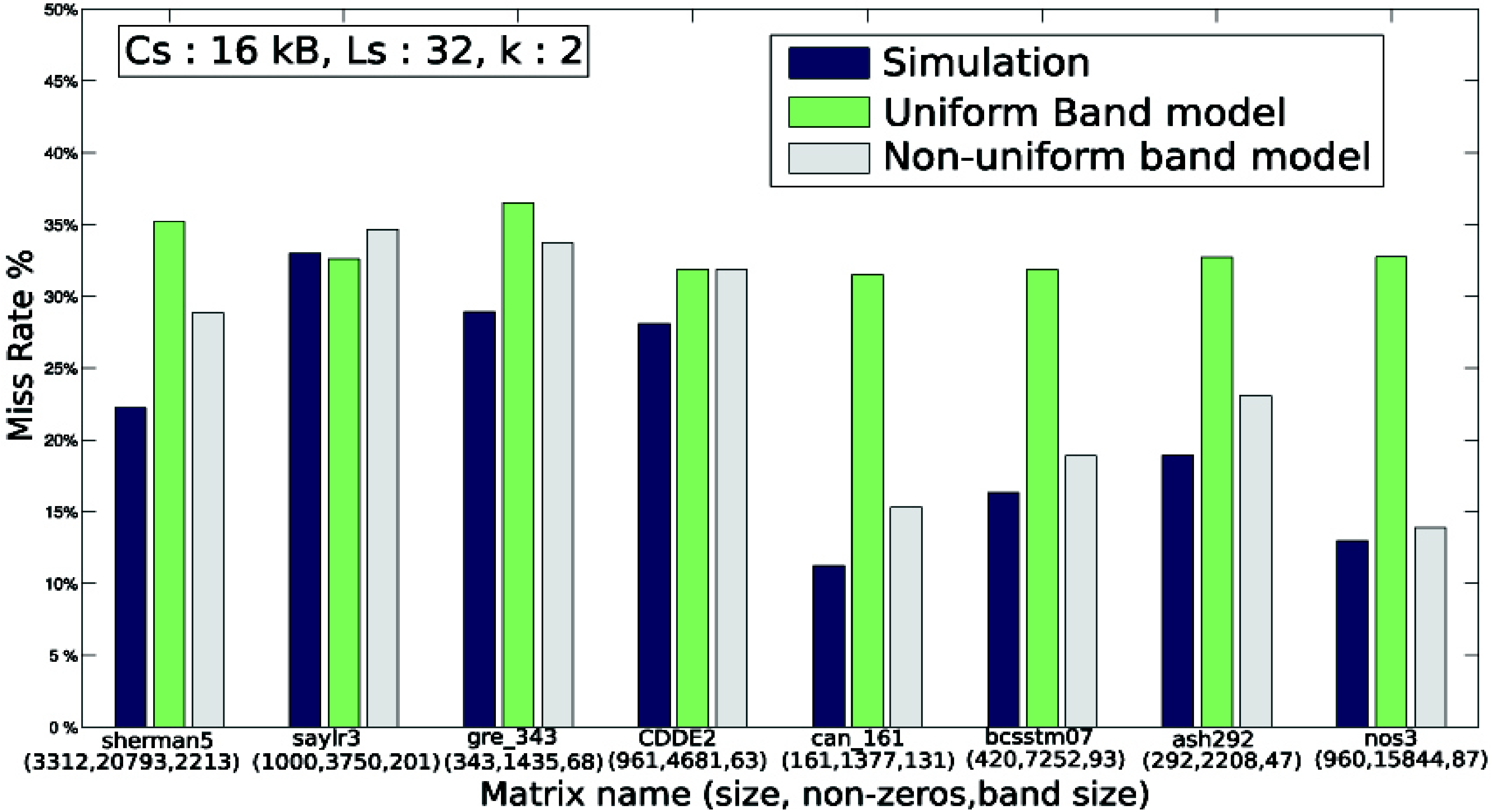


Experimental Results (I)

Code	$\overline{MR_{Sim}}$	Uniform bands model		Non-uniform bands model	
		$\overline{MR_{Mod}}$	$\overline{\Delta_{MR}}$	$\overline{MR_{Mod}}$	$\overline{\Delta_{MR}}$
SPMXV	14.00%	15.57%	1.80%	14.45%	0.70%
SPMXDMIKJ	27.66%	45.62%	26.81%	28.85%	4.19%
SPMXDMIJK	8.62%	27.48%	17.23%	10.91%	3.10%
SPMXDMJIK	7.87%	10.63%	3.23%	8.36%	0.78%
TRANSPOSE	10.31%	11.38%	3.55%	9.52%	3.23%

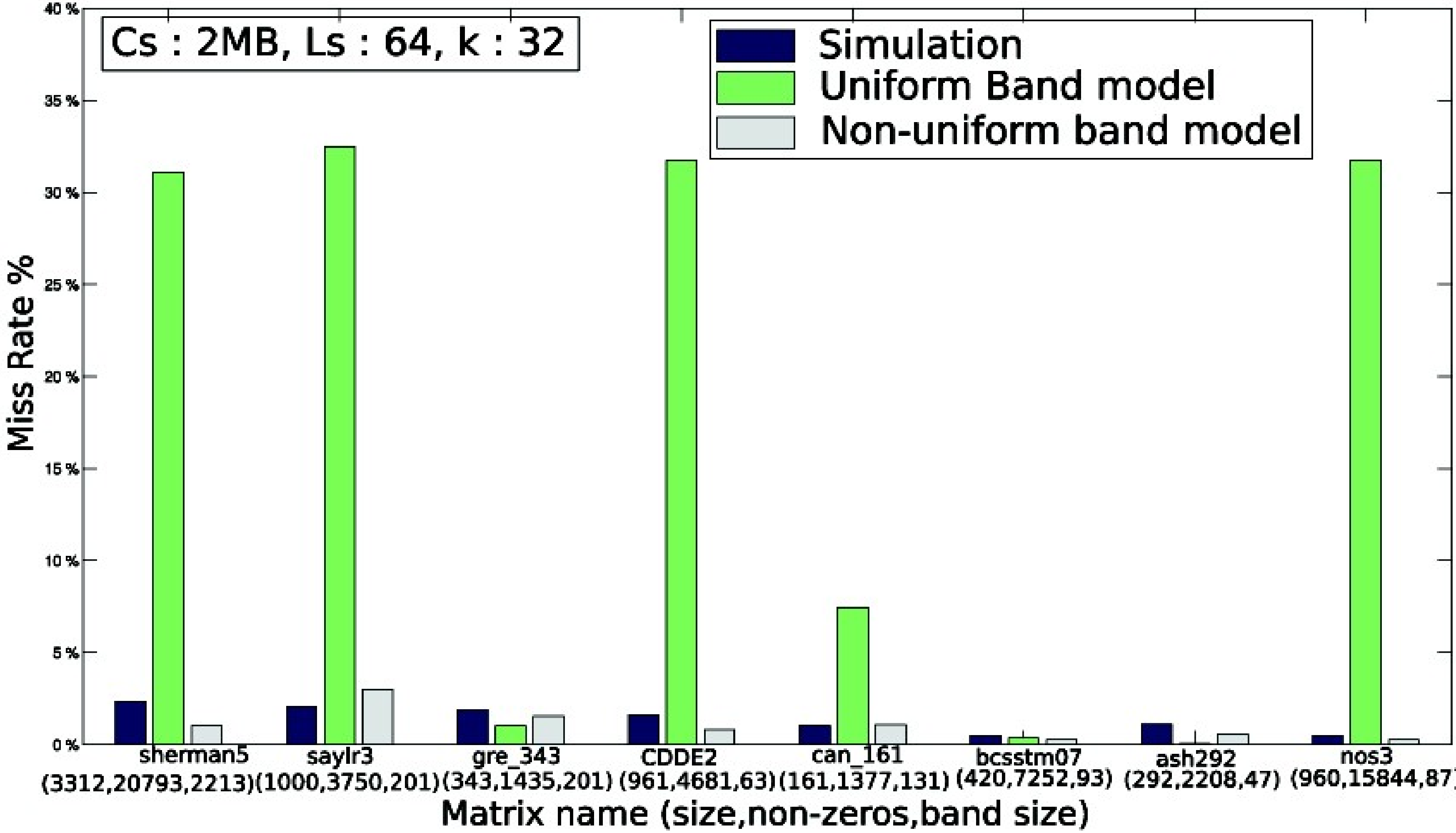
177 Matrices from Harwell-Boeing and NEP collections (52%)

Experimental Results (II)



Code : Sparse Matrix-Dense Matrix Product (IKJ)

Experimental Results (III)



Code : Sparse Matrix-Dense Matrix Product (IKJ)

Experimental Results (and IV)

Architecture	L1 Parameters (C_{s1},L_{s1},K_1,$Cost_1$)	L2 Parameters (C_{s2},L_{s2},K_2,$Cost_2$)	L3 Parameters (C_{s3},L_{s3},K_3,$Cost_3$)
Itanium 2	(16K,64,4,8)	(256K,128,8,24)	(6MB,128,24,120)
PowerPC 7447A	(32K,32,8,9)	(512K,64,8,150)	-

Choose the optimal loop nesting ordering for the **sparse matrix-dense matrix product** for both architectures for different problem sizes.

Conclusions

- PME automatable model extension for codes involving non-uniform banded matrices.
- The model accuracy is high.
- The model can be used as a guide in compiler optimization processes.