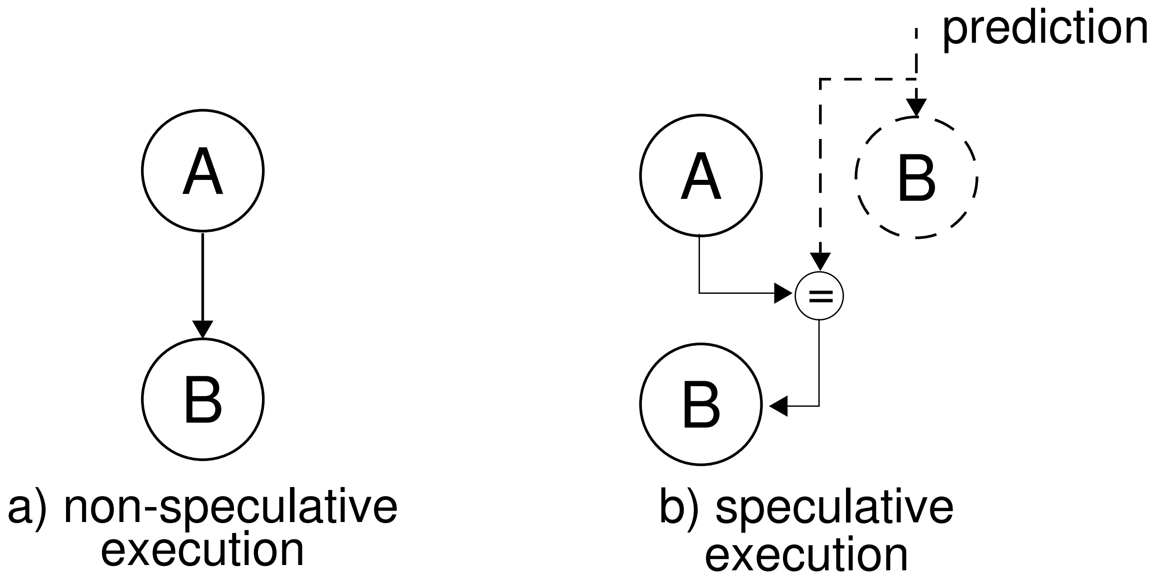
Split Last-Address Predictor

Enric Morancho, José María Llabería, Àngel Olivé
Departament d'Arquitectura de Computadors
Universitat Politècnica de Catalunya

Work supported by the Ministry of Education and Science of Spain (CICYT TIC-0429/95)

Introduction

- Processors must exploit more ILP
- To avoid restrictions imposed by true-data dependencies:
    - ❑ Prediction of the result of an operation
    - ❑ Speculative execution of dependent operations



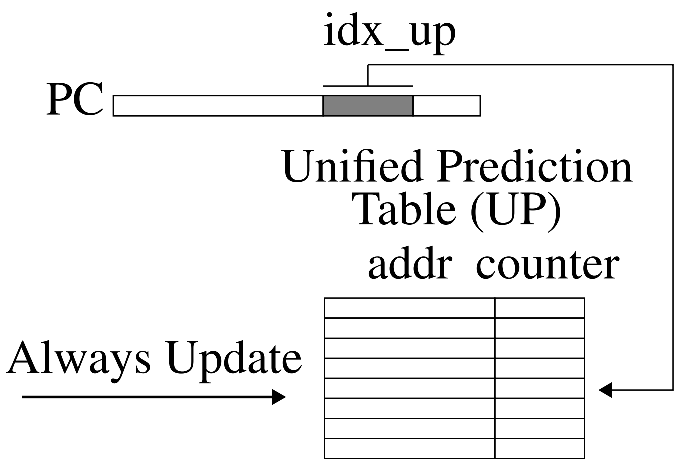
Last-address predictor

- Loads: effective addresses prediction
- Prediction = last effective address computed by the load
- Captured Predictability = $\dfrac{\text{correct predictions}}{\text{dynamic loads}} \times 100$
- Maximum capturable predictability

Benchmark	Pred. capt.
go	52.7
li	35.6
ijpeg	21.7
perl	79.2

Unified Predictor (UP)

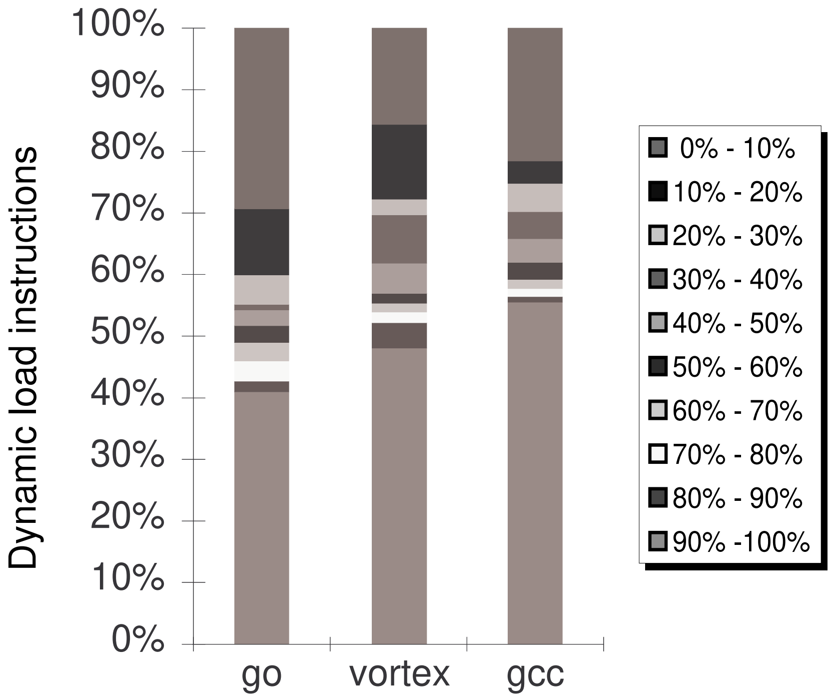
- Update policy: always update



- ❑ Counters prevent some mispredictions (accuracy)
  - ✓ Also prevent some correct predictions
- ❑ Unpredictable loads can evict predictable ones

Predictability distribution

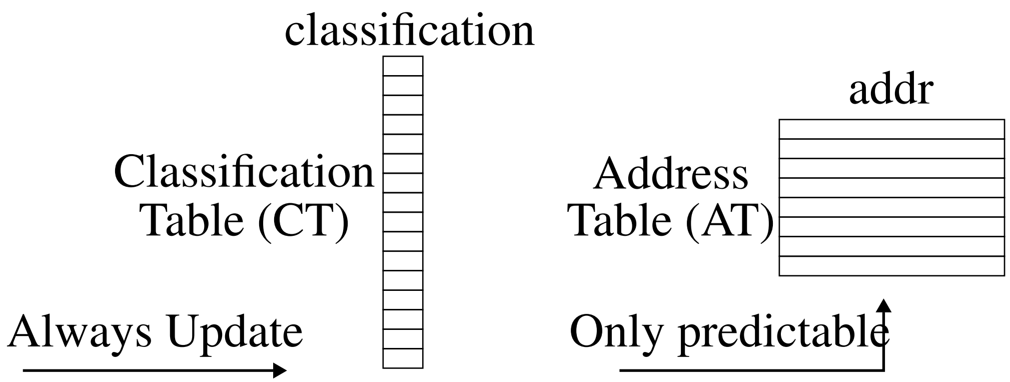
- Predictability is distributed non-uniformly



- ❑ Loads highly predictable
- ❑ Loads highly unpredictable

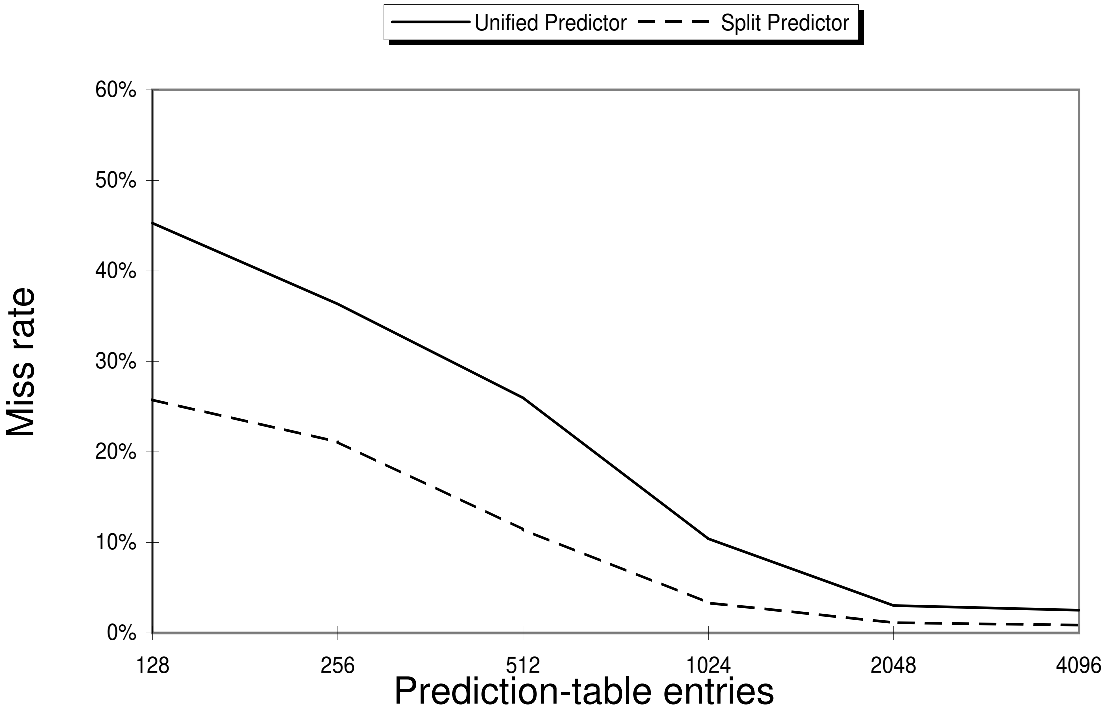
Idea of the Split Predictor (SP)

- Filtering load instructions that update prediction table
  - ❑ Only predictable loads update table
    - ✓ table size reduction
  - ❑ Loads must be classified



Miss rate

- Benchmark go, perfect filtering, LRU replacement policy

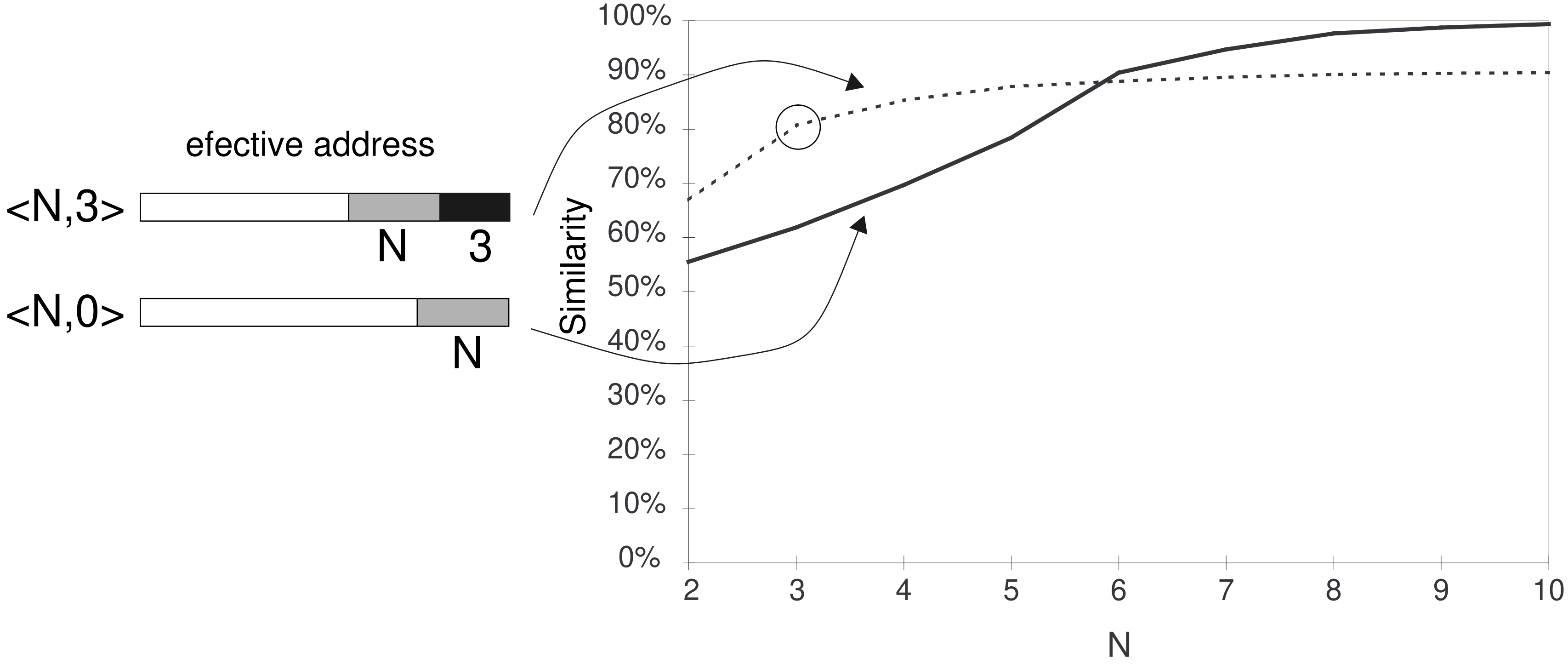


❑Miss-rate reduction

❑Similar miss rate: UP 1024 entries
SP 512 entries

Classifying mechanism <N,k>

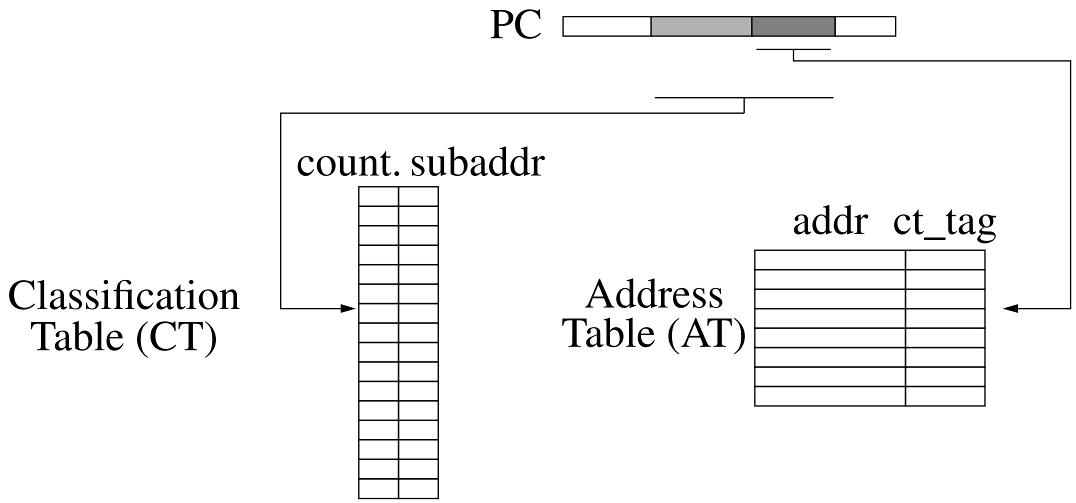
- Uses N bits discarding the k low-order bits (<N,k>)



❑ Similarity of <3,3> is 80%

Split Predictor implementation

- CT classifies load instructions using mechanism <3,3>
- SP places in AT only predictable load instructions

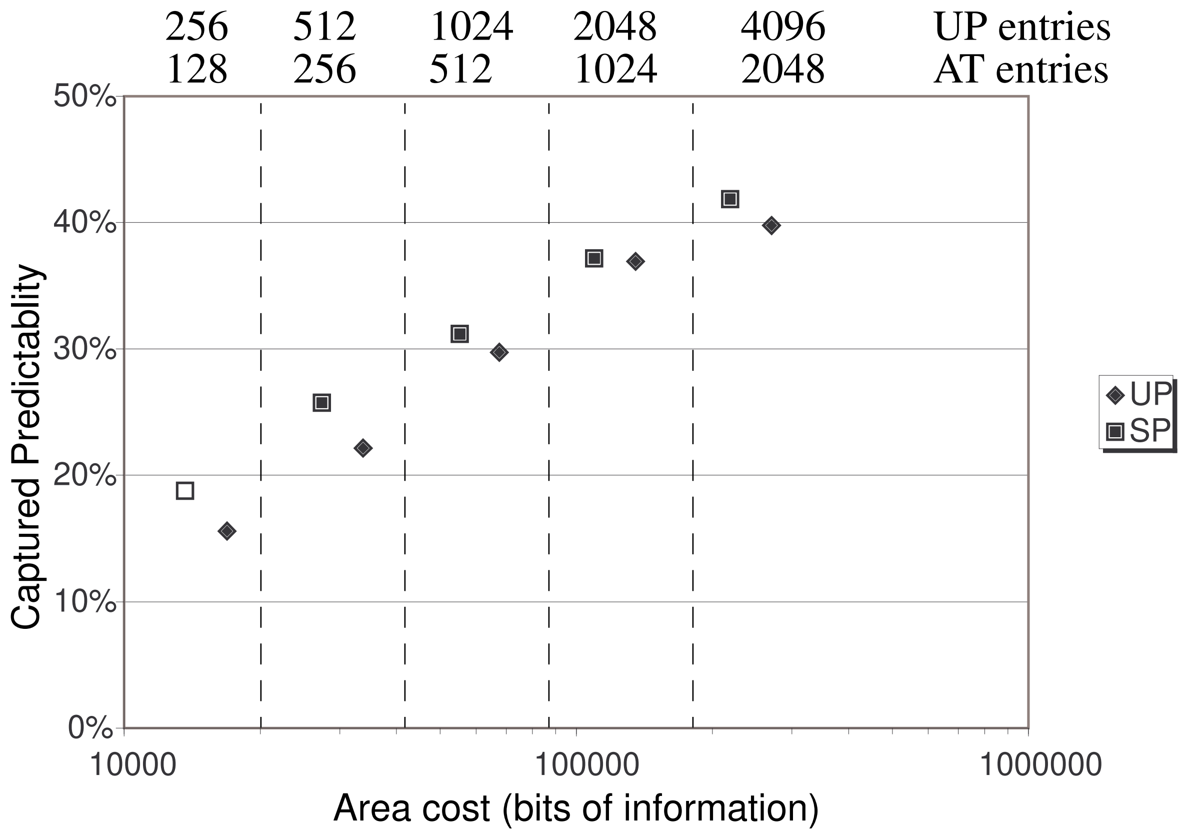


❑ Direct mapping

❑ Ratio CT entries and AT entries

Captured predictability

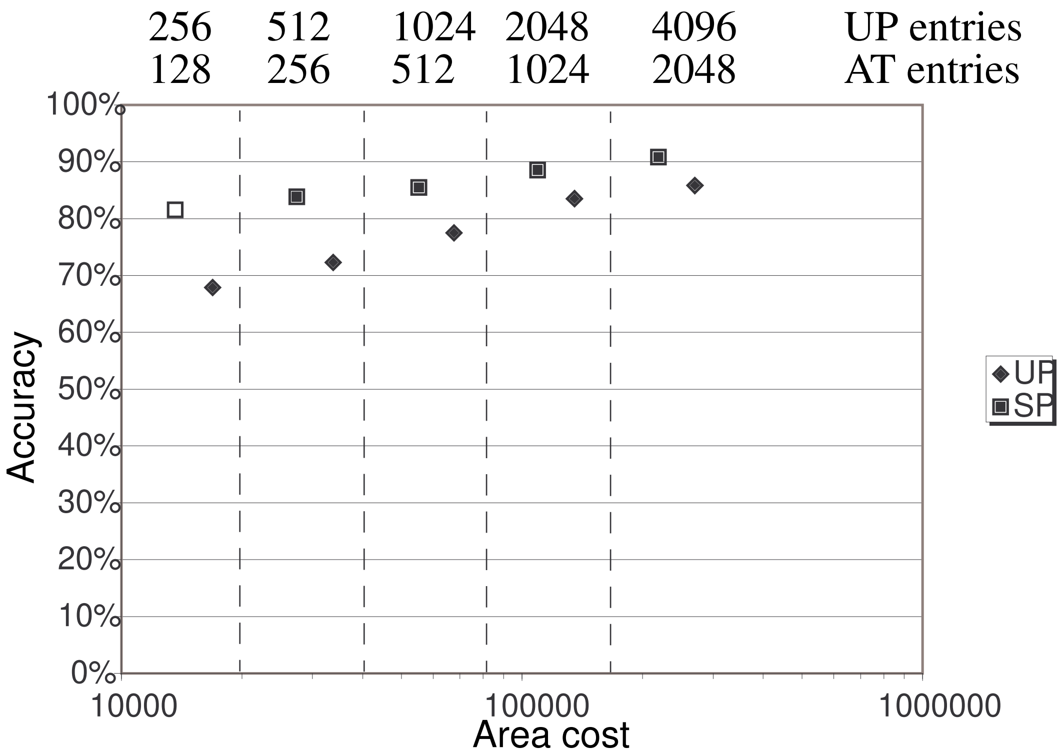
- Benchmark go, both predictors, ratio $\frac{CTentries}{ATentries} = 8$



Accuracy

- Mispredictions can produce a performance penalty

- $$\text{Accuracy} = \frac{\text{correct predictions}}{\text{predictions performed}} \times 100$$



Conclusions

- Effective addresses are predictable
- Predictability is not distributed uniformly
- Updates considering predictability
- With 19% area-cost reduction, the SP maintains captured predictability and increases accuracy of the UP
- The same idea can be applied to value prediction and stride-based prediction