# Bungee Jumps: Accelerating Indirect Branches Through Hardware/Software Co-Design

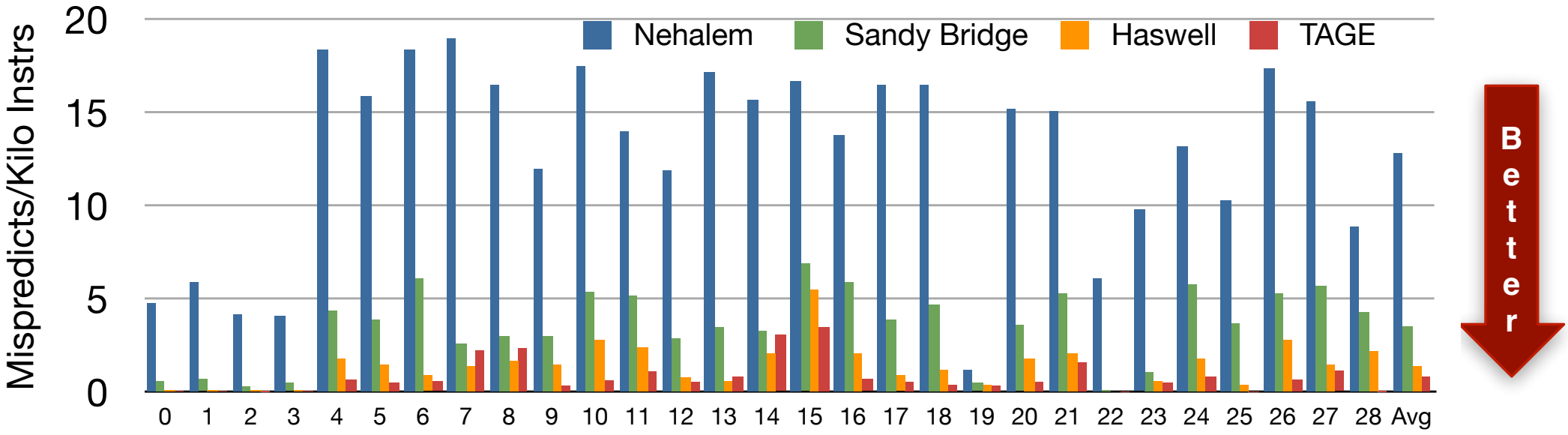
**Daniel S. McFarlin**

**Craig Zilles**

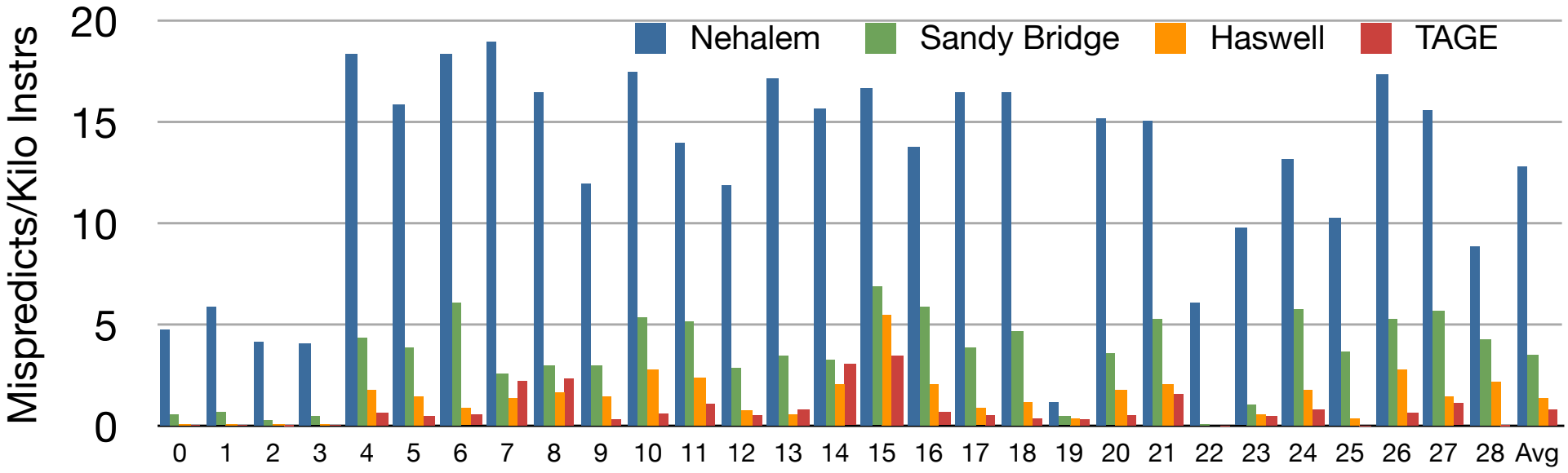
Carnegie Mellon
Electrical & Computer ENGINEERING

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

# Indirect Branches Are Increasingly Predictable

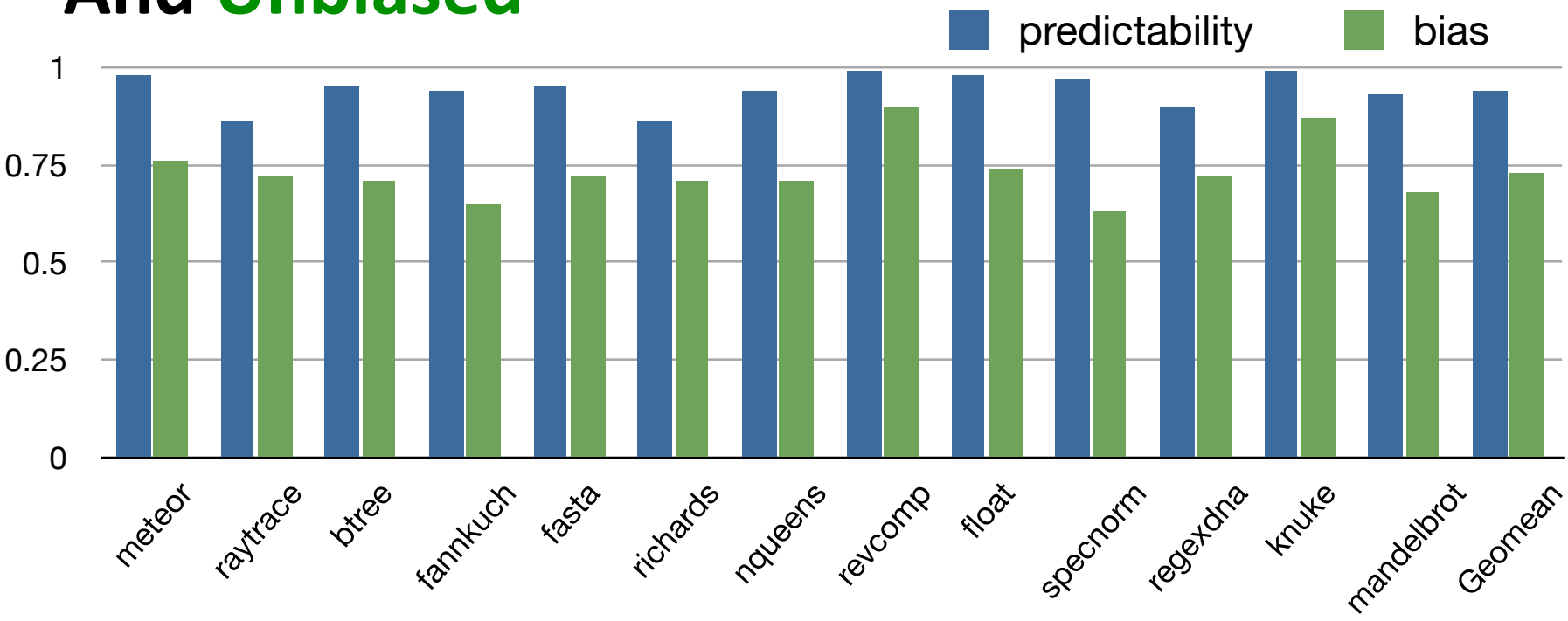


# Indirect Branches Are Increasingly Predictable

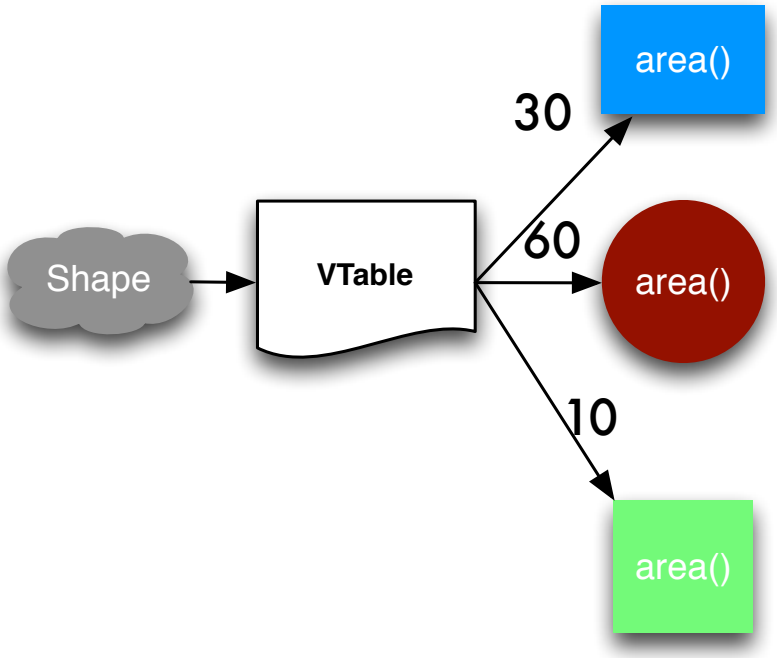



Better

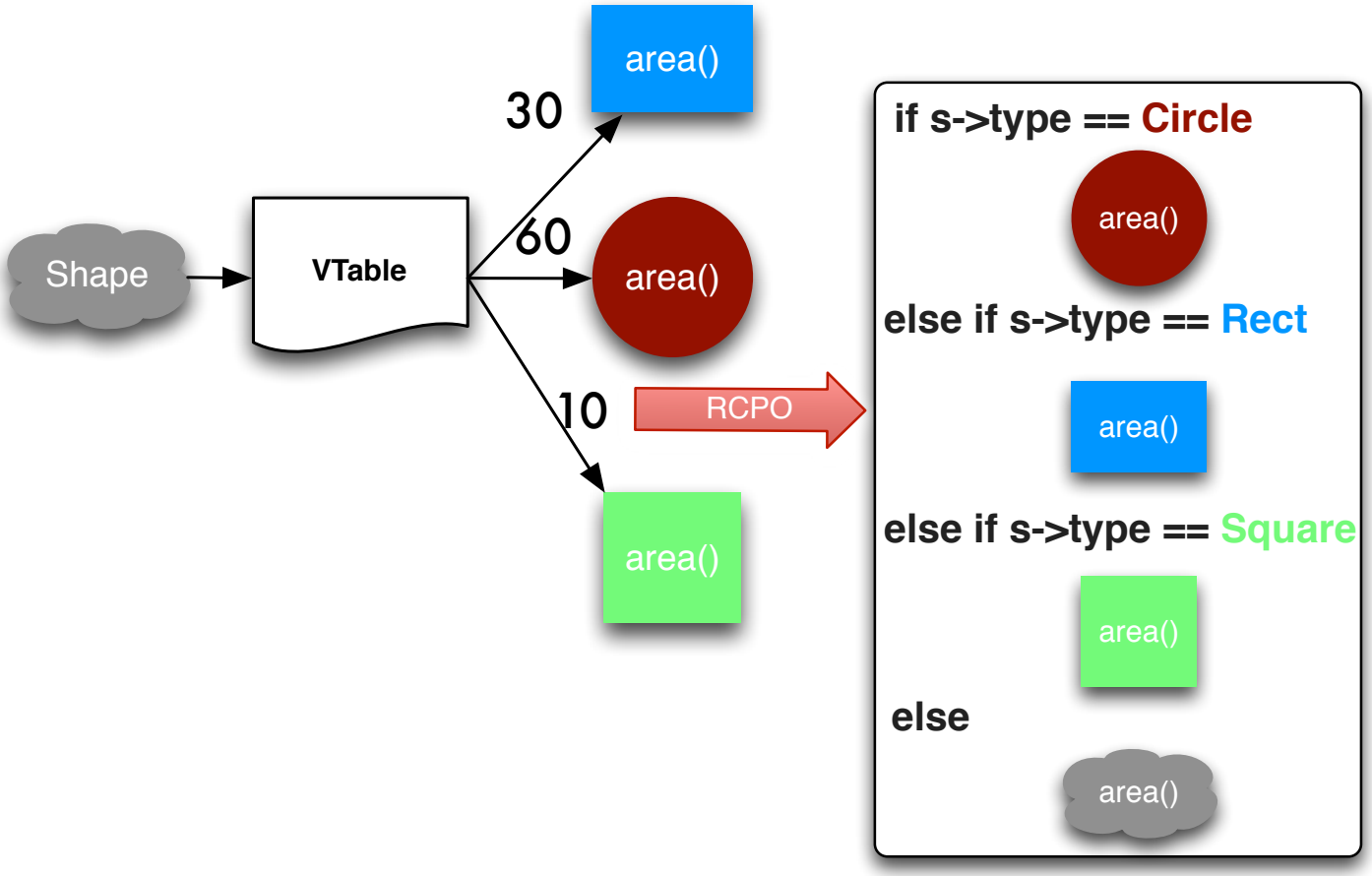
## And Unbiased



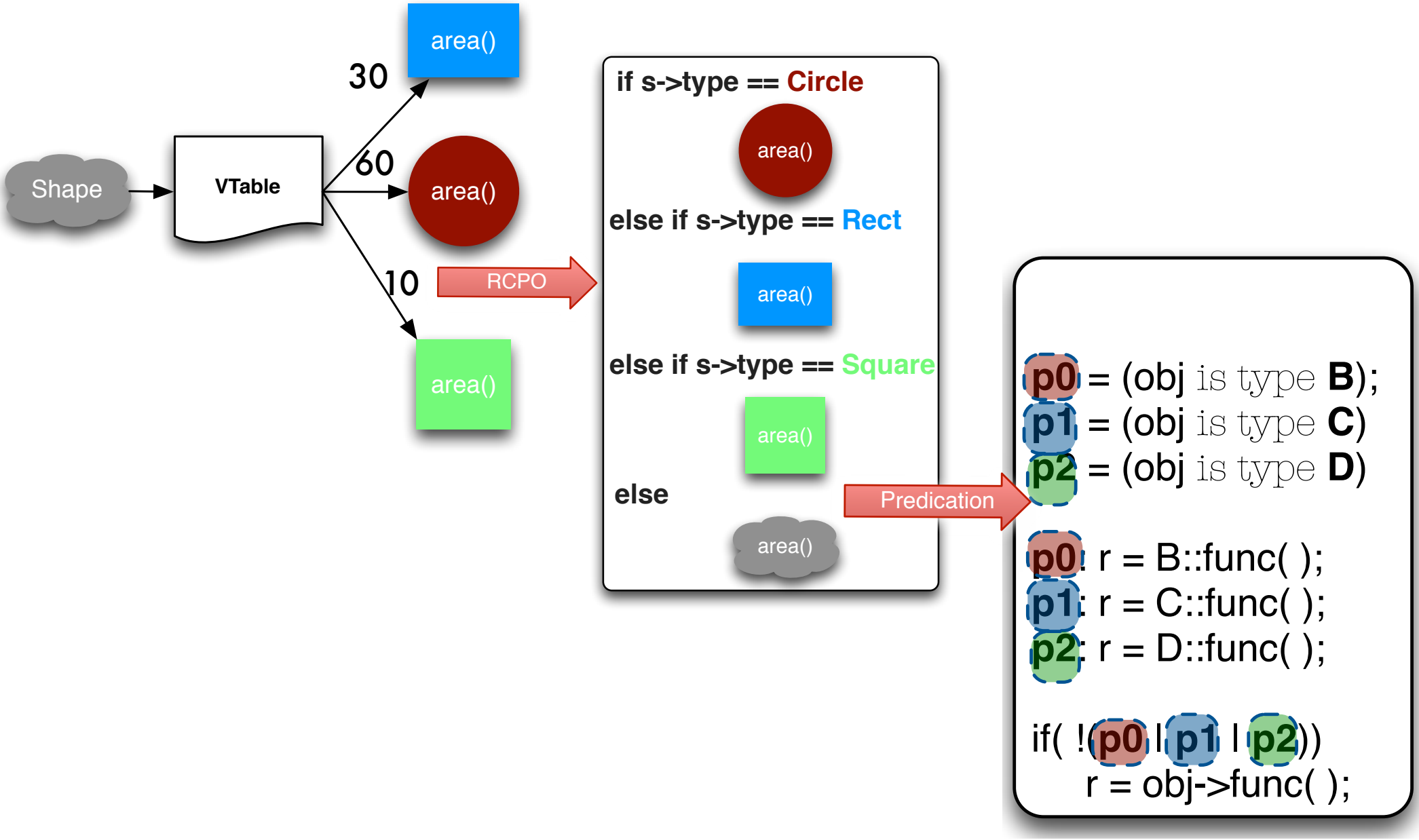
# In-Order Machines Specialize Based on Branch Bias or Eliminate Branch Prediction Altogether



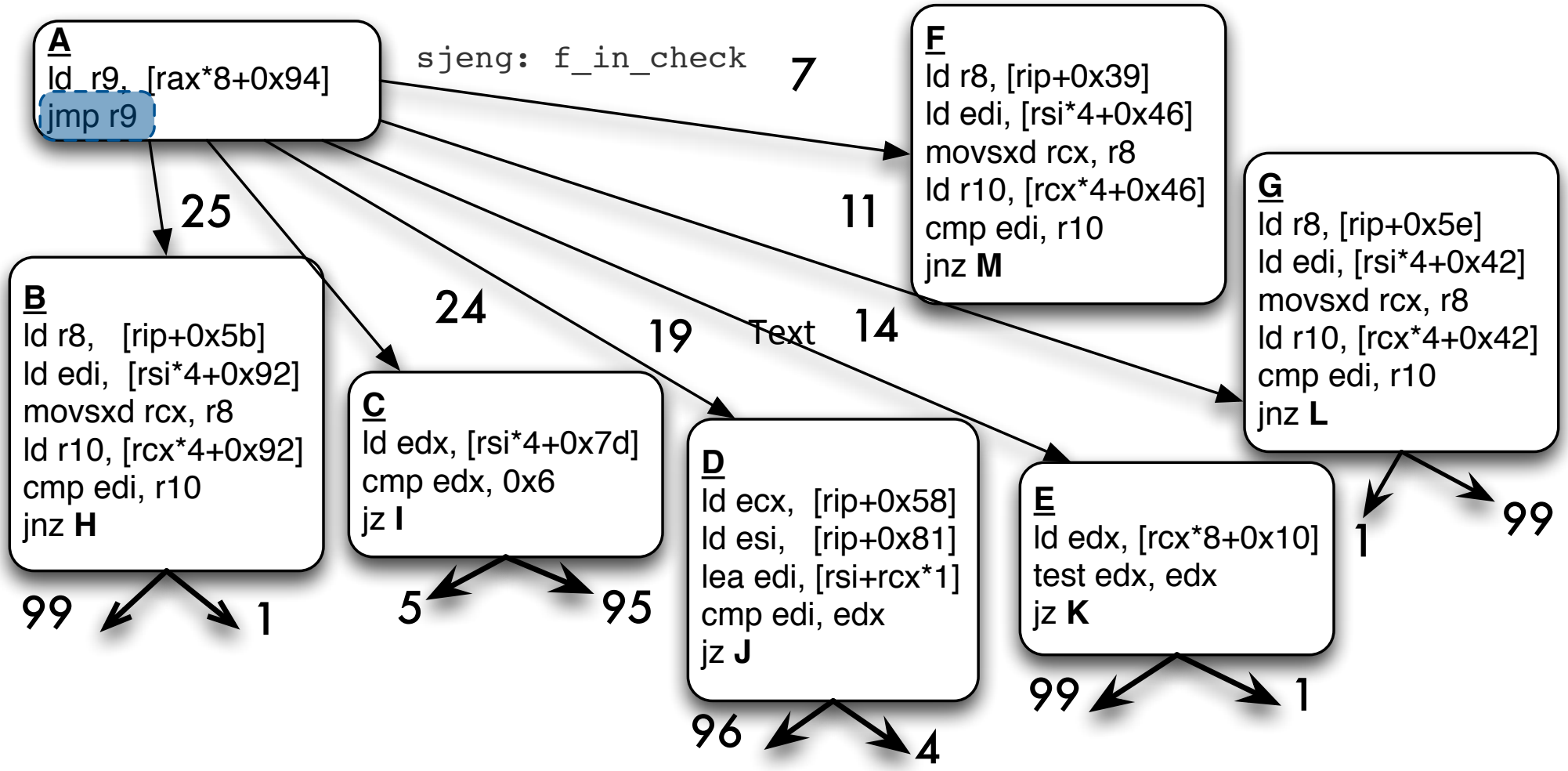
# In-Order Machines Specialize Based on Branch Bias or Eliminate Branch Prediction Altogether



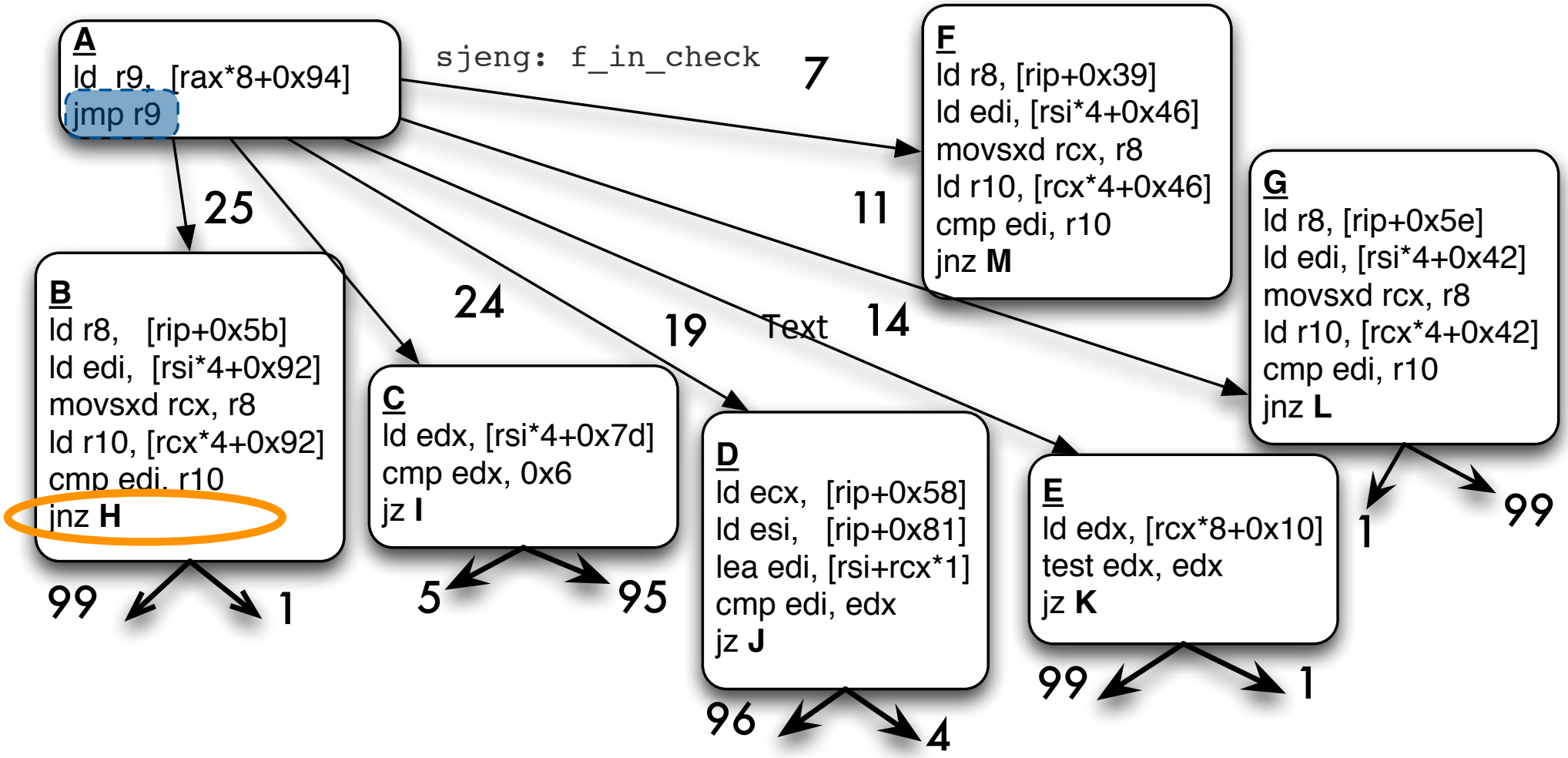
# In-Order Machines Specialize Based on Branch Bias or Eliminate Branch Prediction Altogether



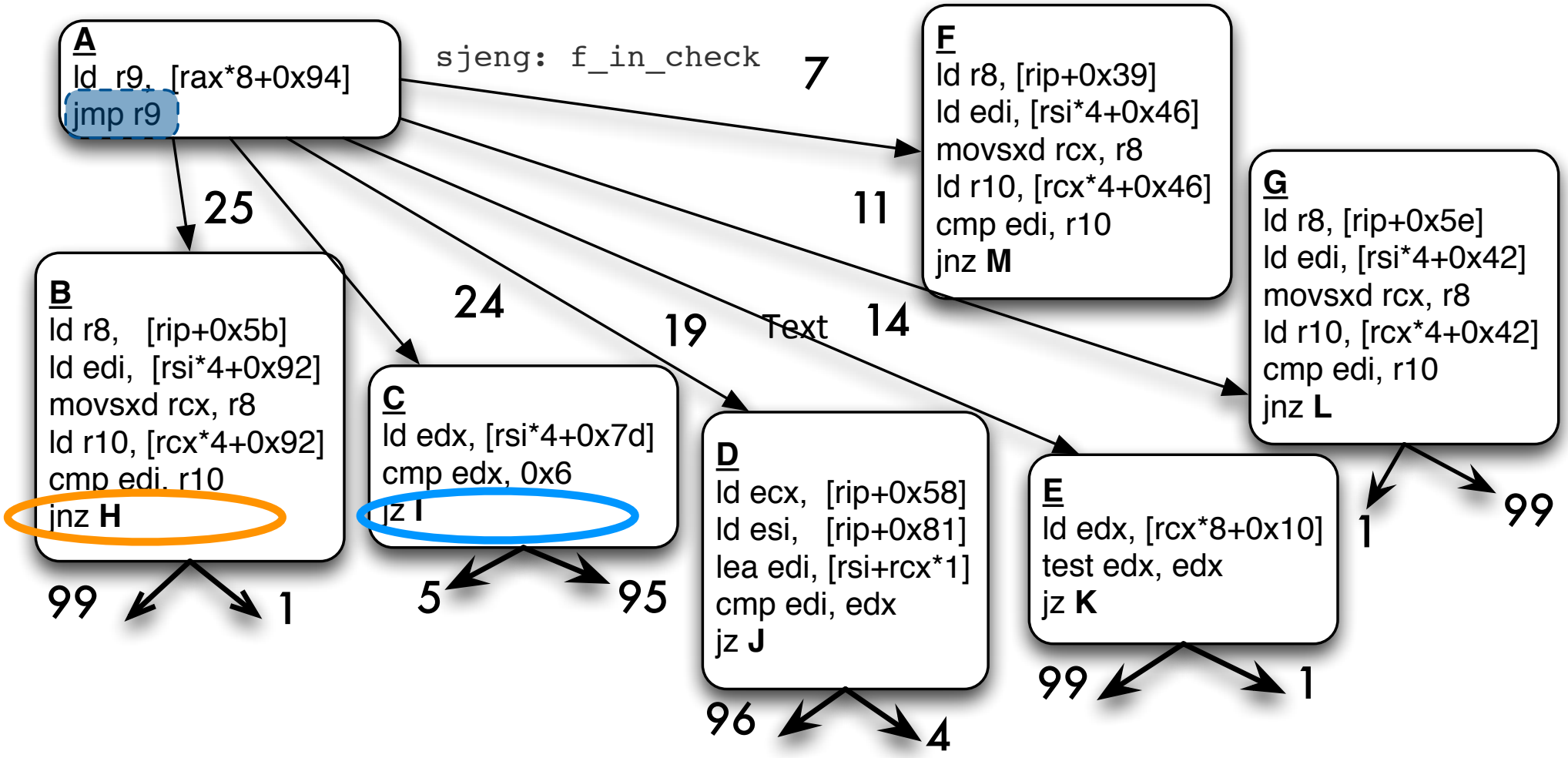
# Challenge: Non-Reconvergence & Large Number of Targets



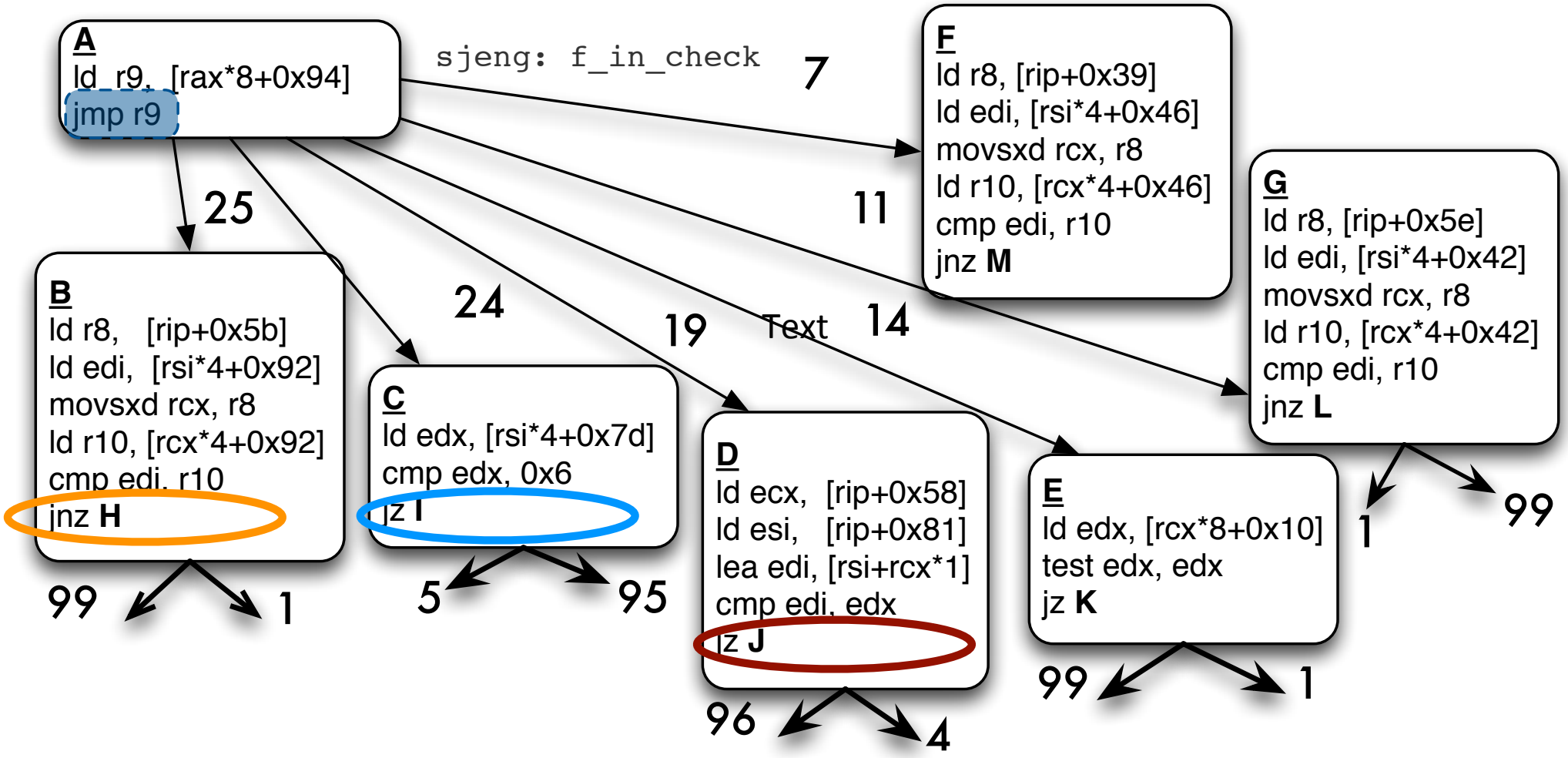
# Challenge: Non-Reconvergence & Large Number of Targets



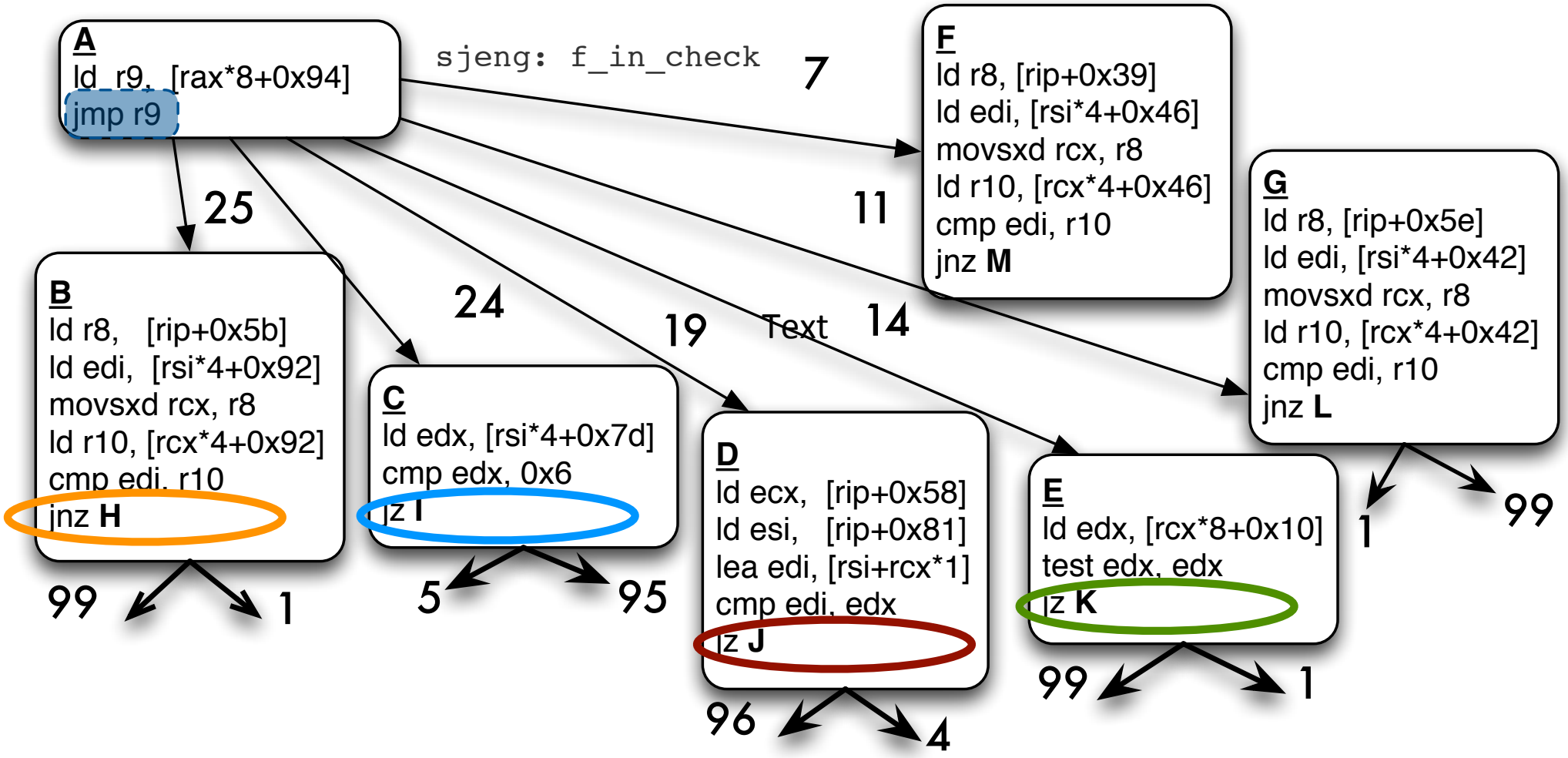
# Challenge: Non-Reconvergence & Large Number of Targets



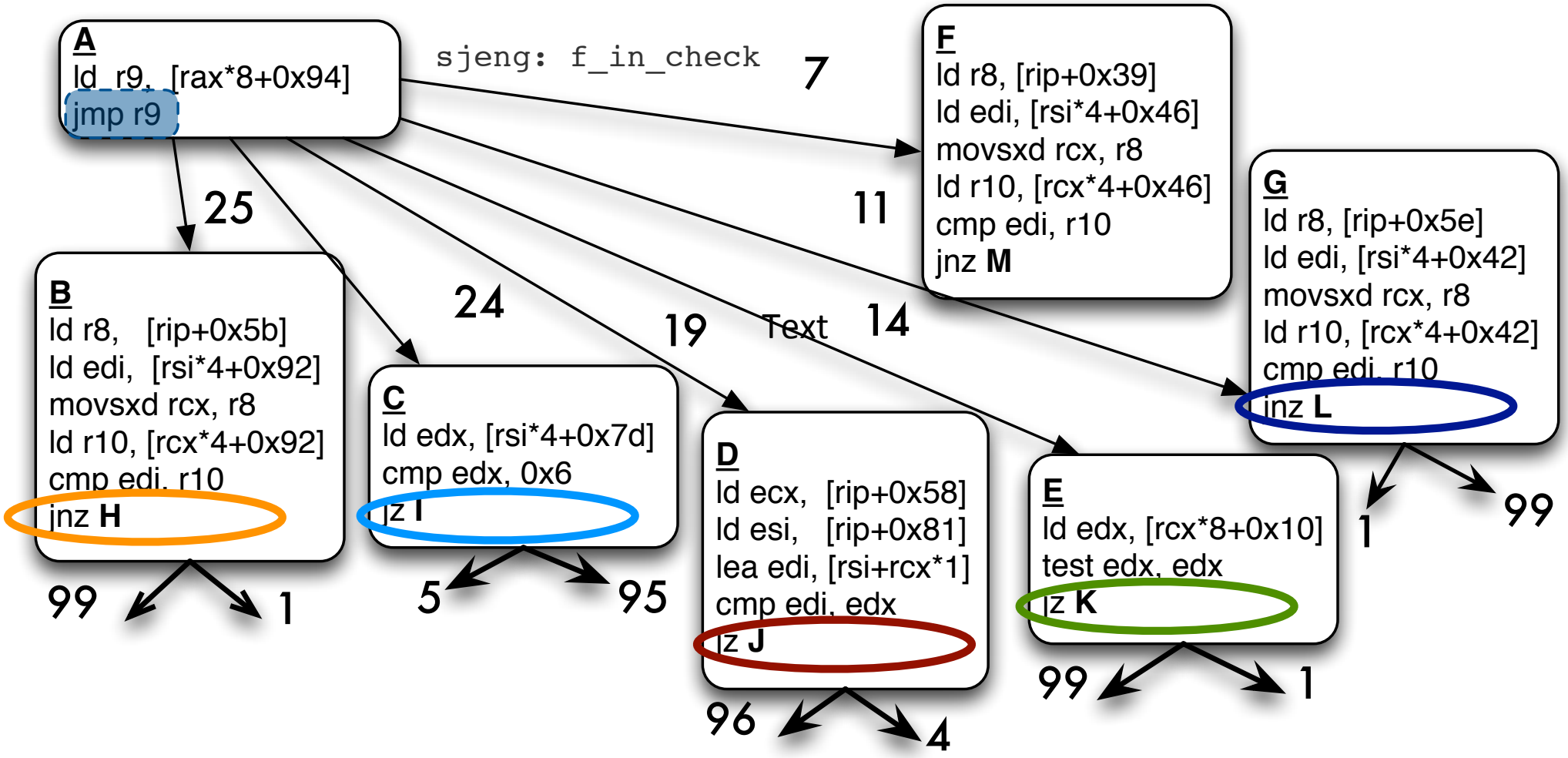
# Challenge: Non-Reconvergence & Large Number of Targets



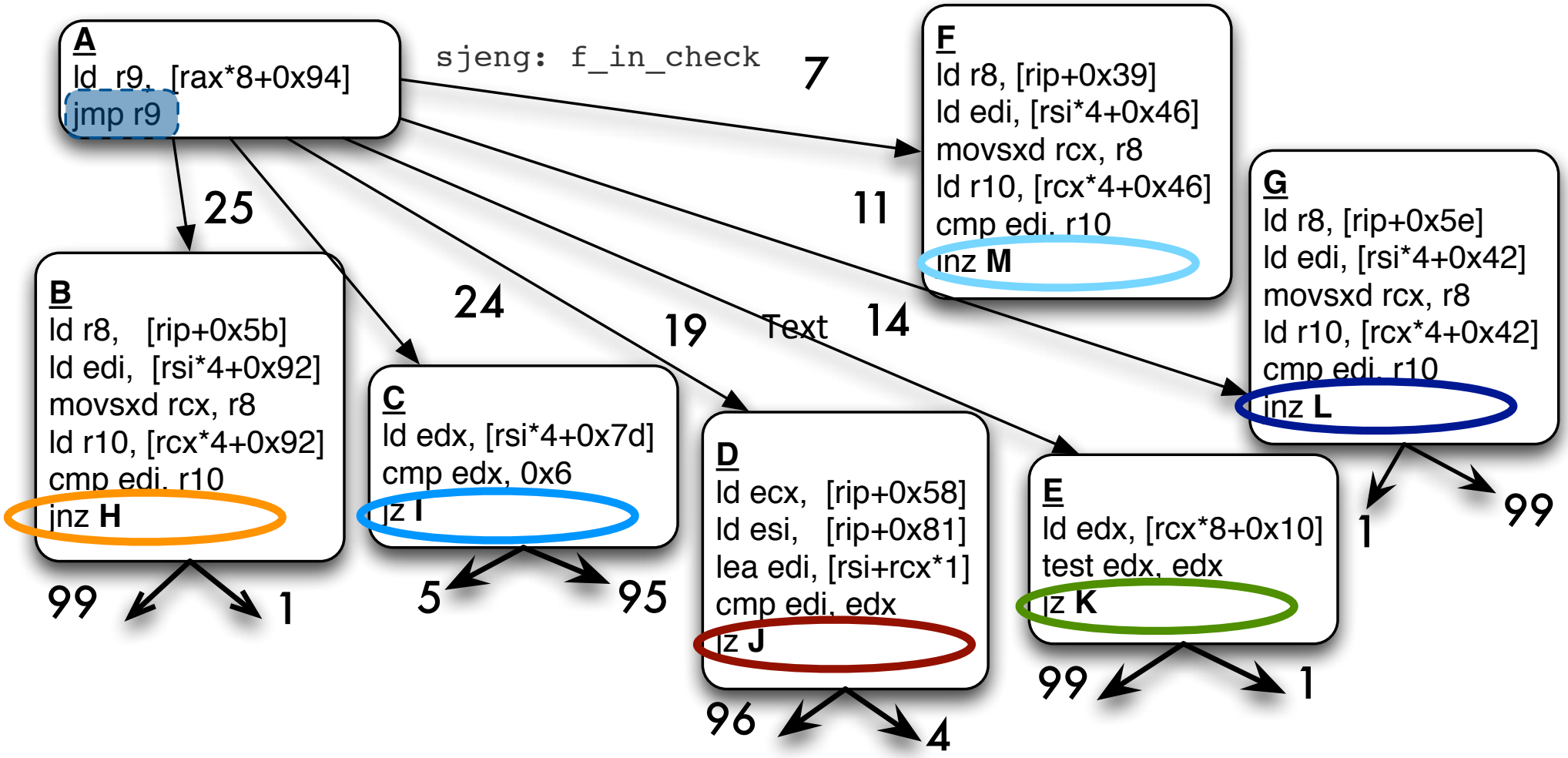
# Challenge: Non-Reconvergence & Large Number of Targets



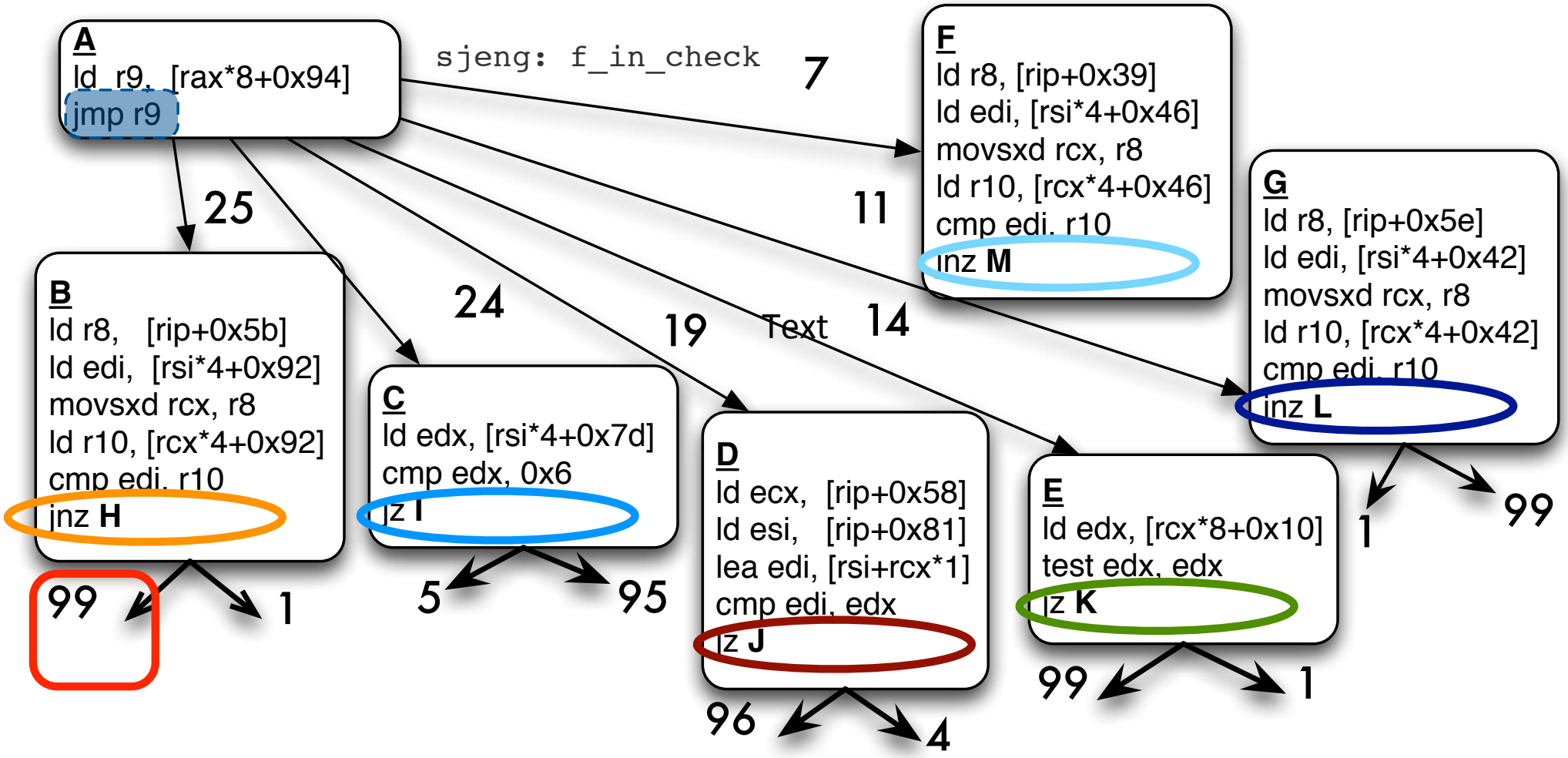
# Challenge: Non-Reconvergence & Large Number of Targets



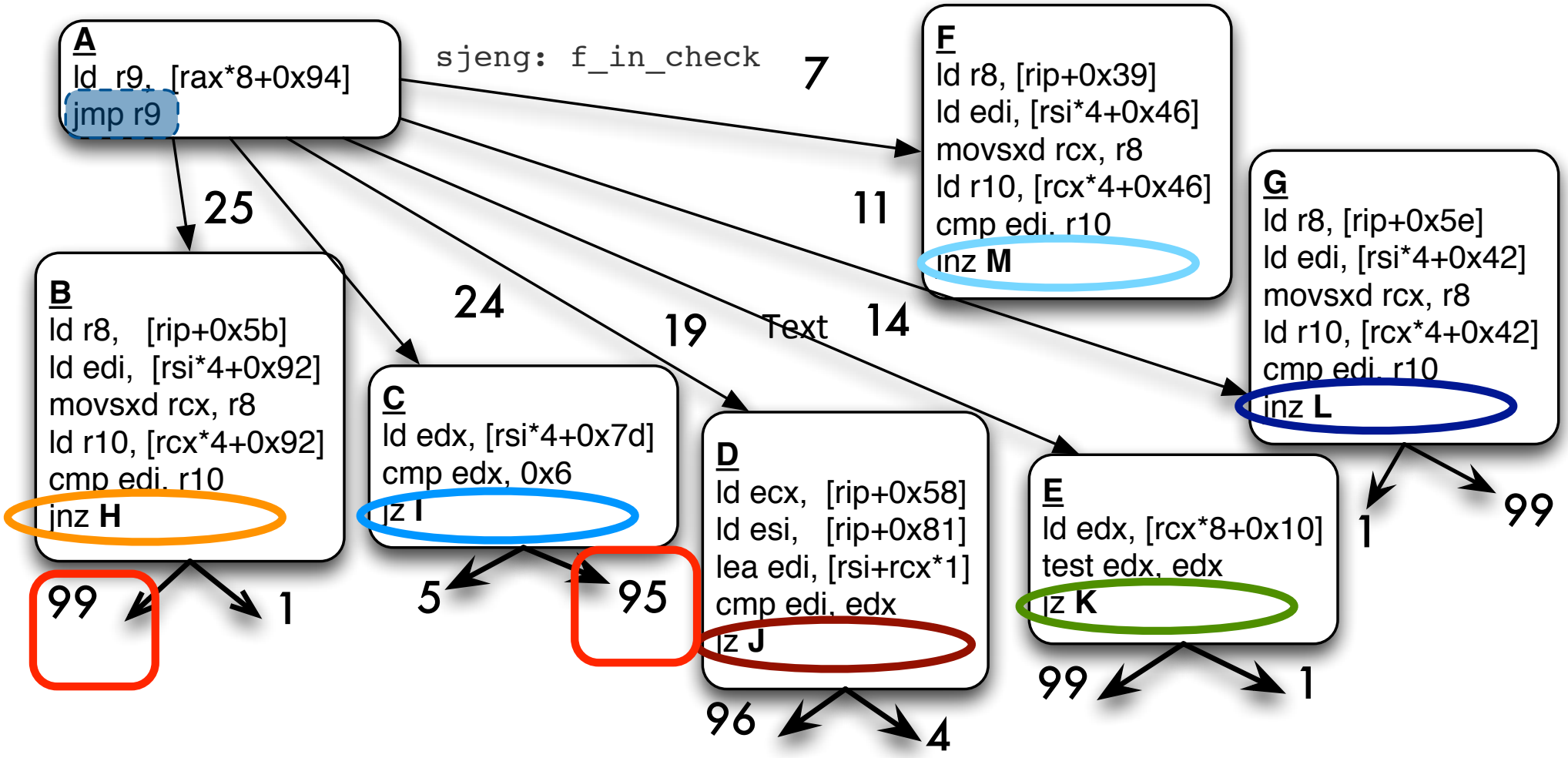
# Challenge: Non-Reconvergence & Large Number of Targets



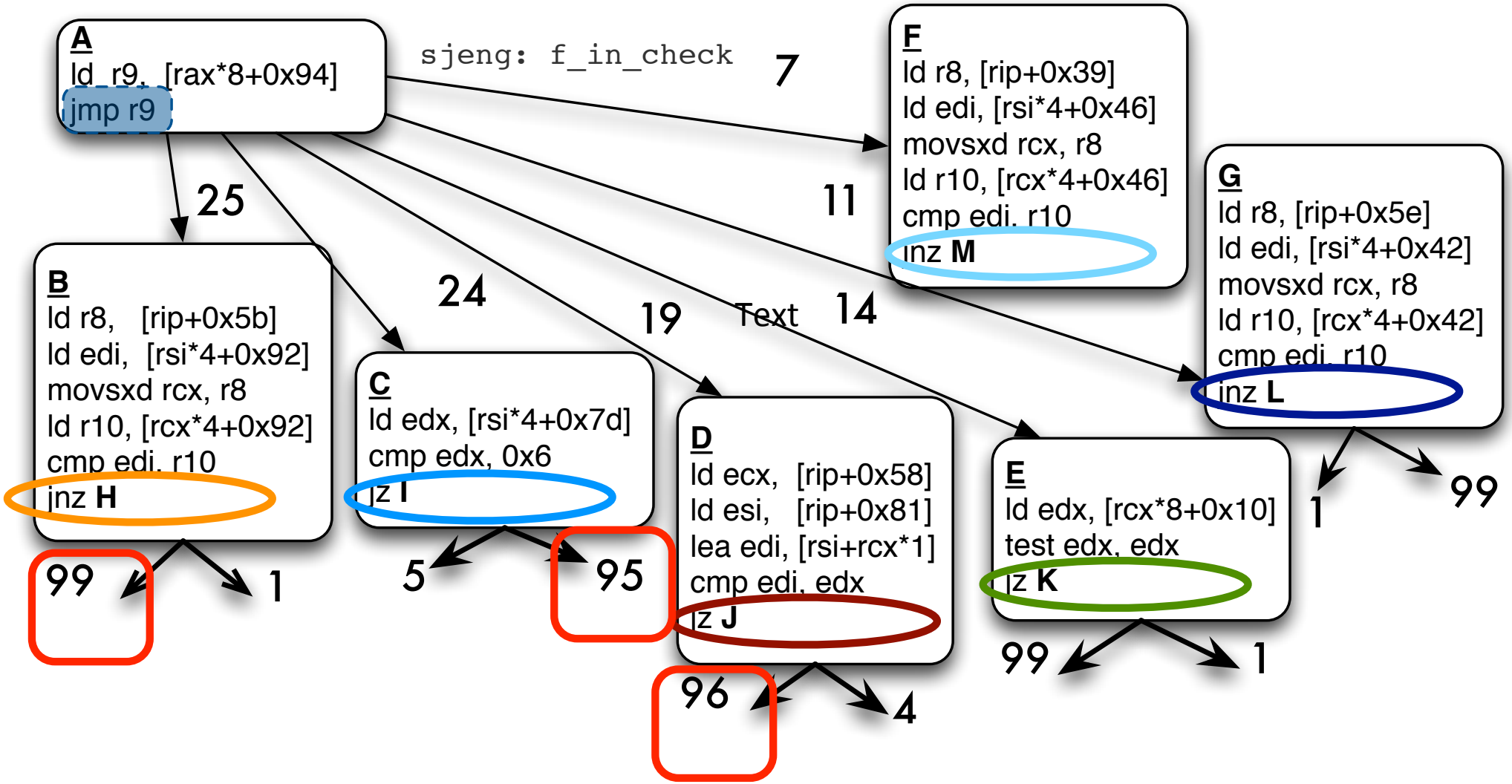
# Challenge: Non-Reconvergence & Large Number of Targets



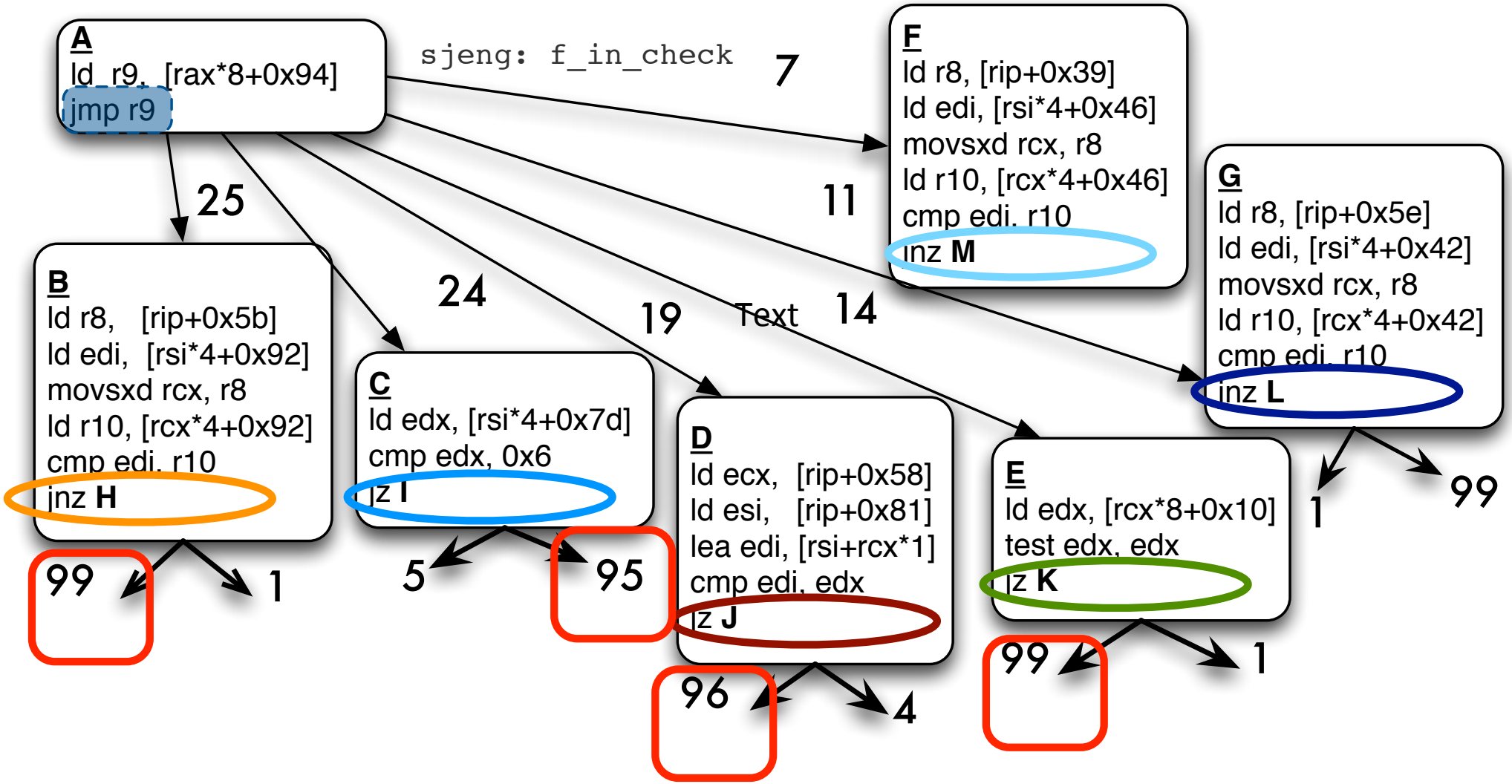
# Challenge: Non-Reconvergence & Large Number of Targets



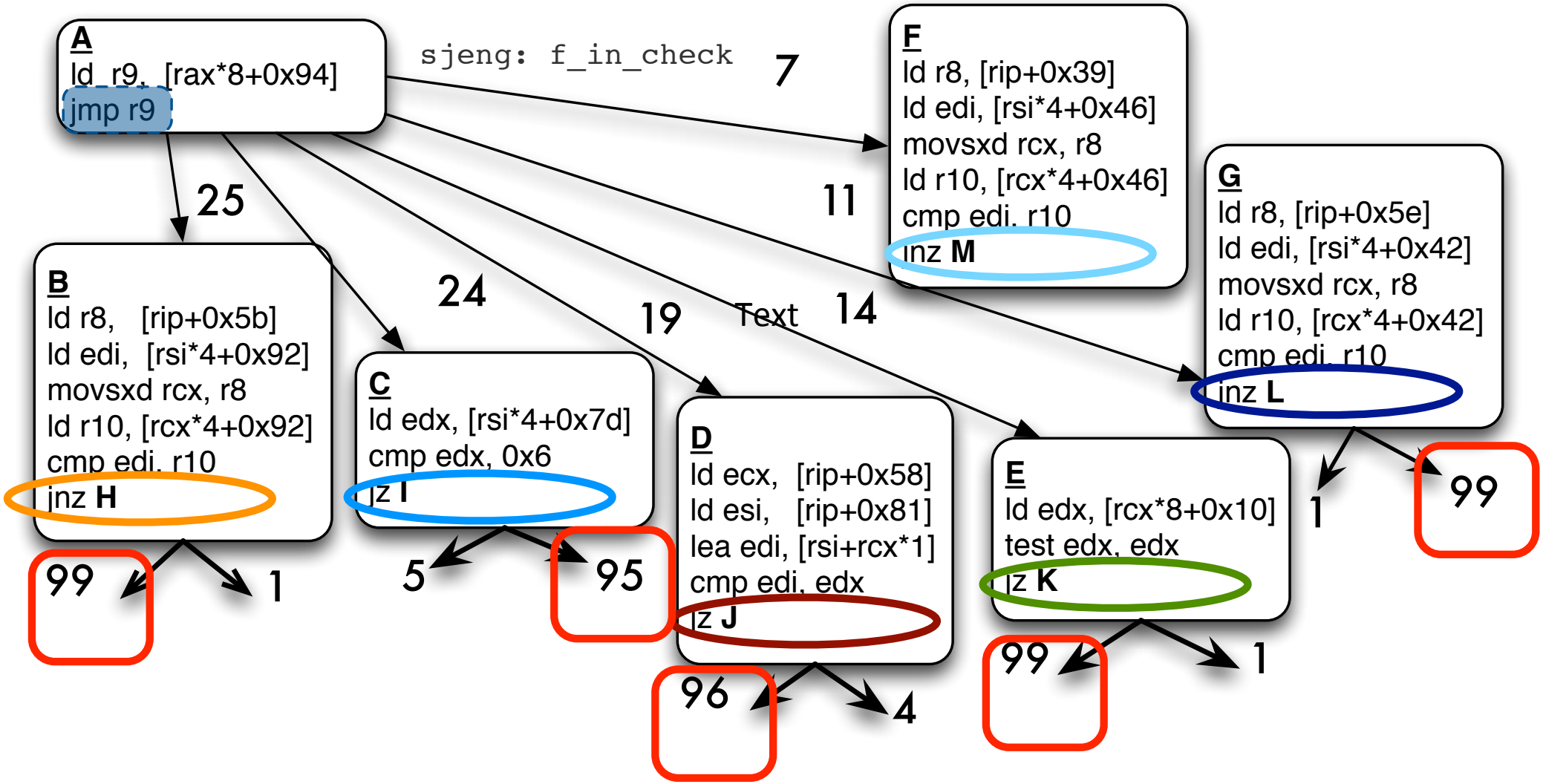
# Challenge: Non-Reconvergence & Large Number of Targets



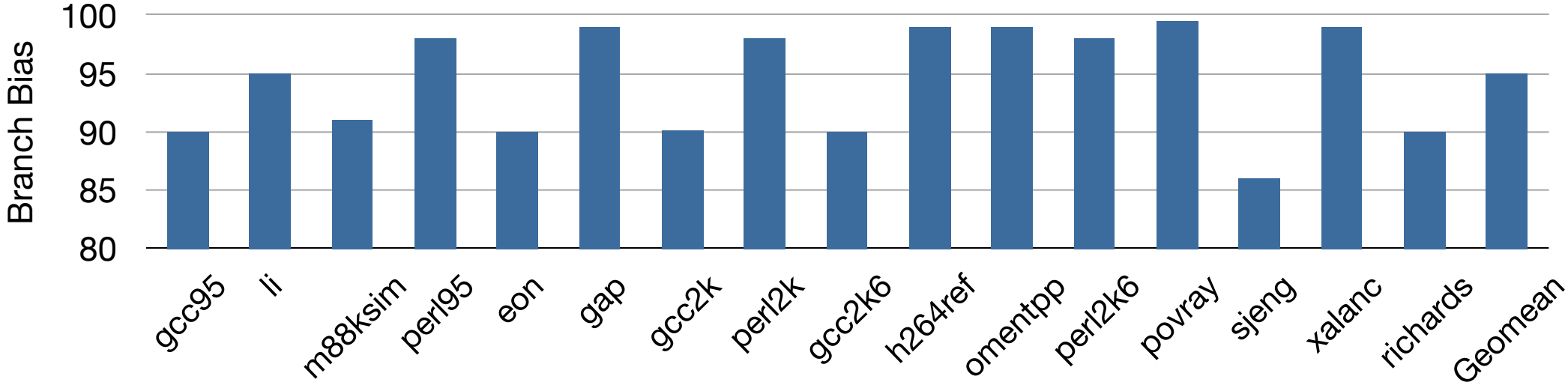
# Challenge: Non-Reconvergence & Large Number of Targets



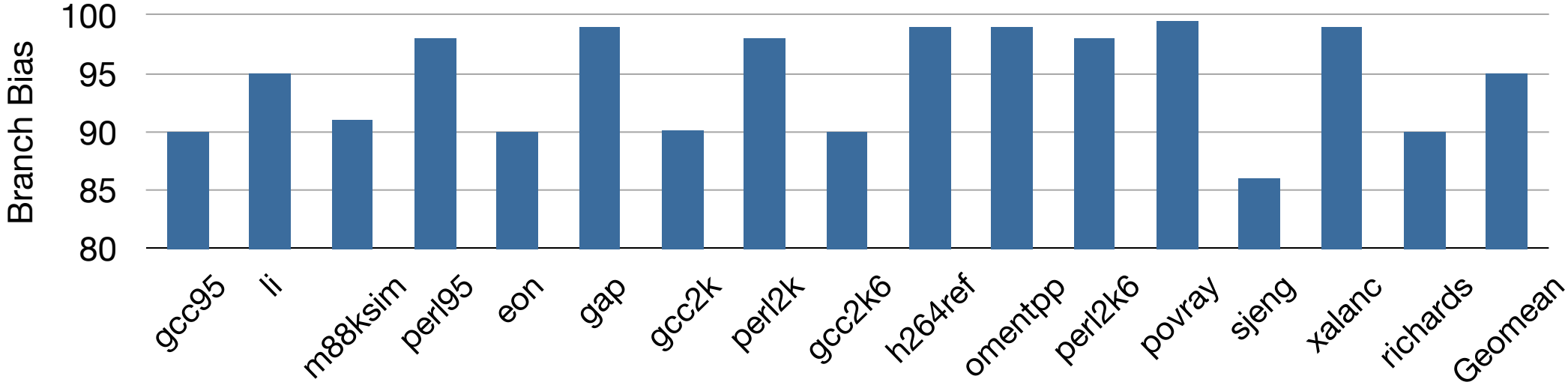
# Challenge: Non-Reconvergence & Large Number of Targets

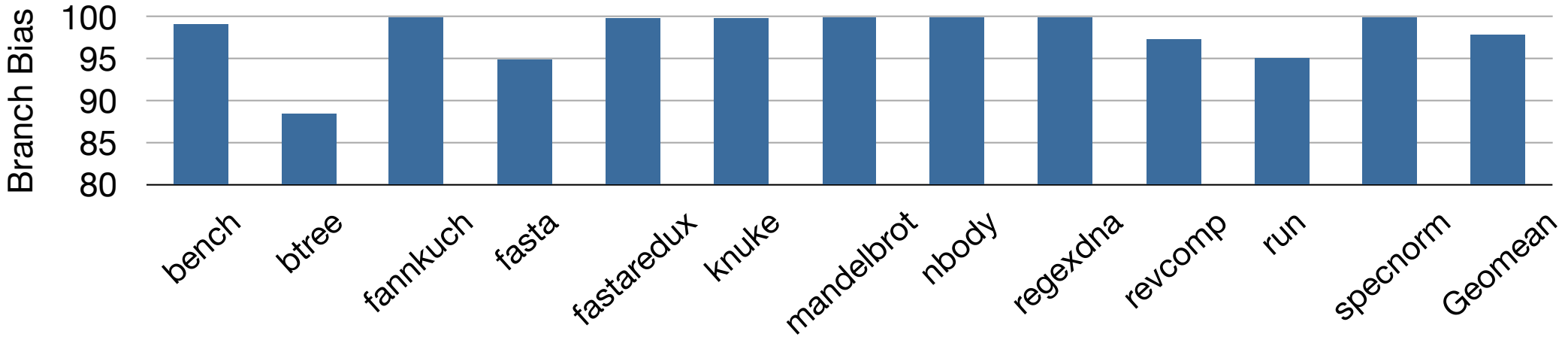


# Missed Optimization Opportunity: Next Branch Bias

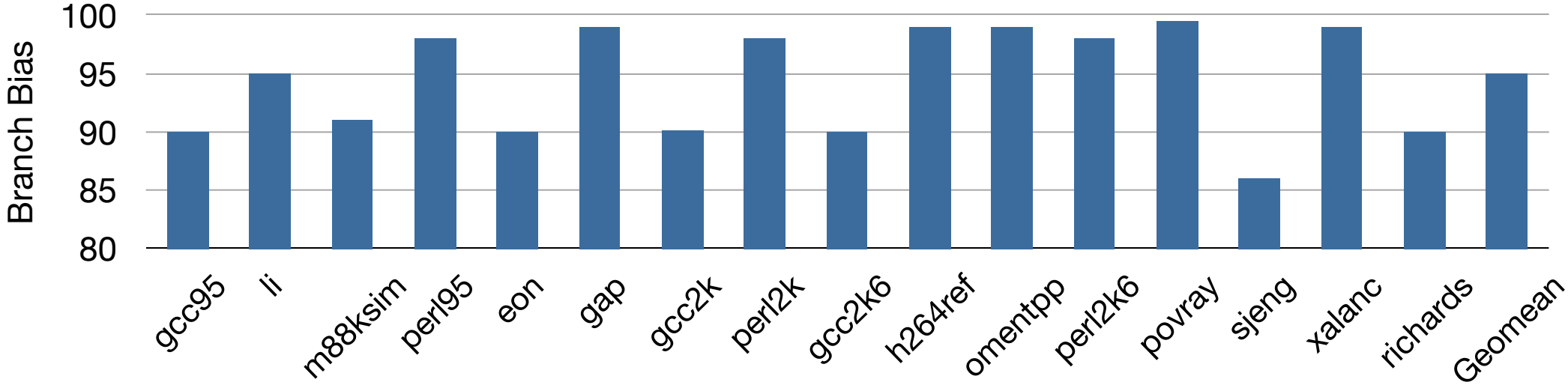


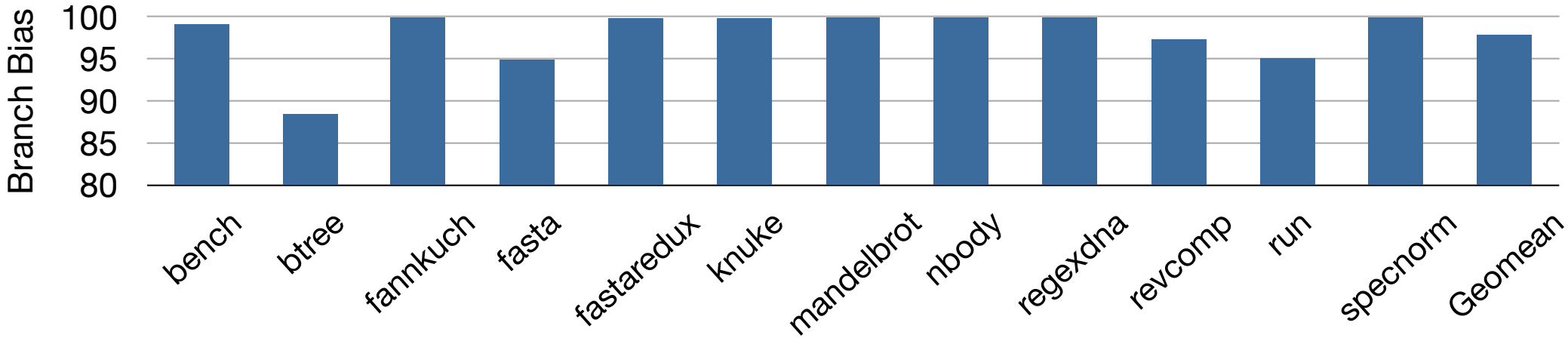
# Missed Optimization Opportunity: Next Branch Bias

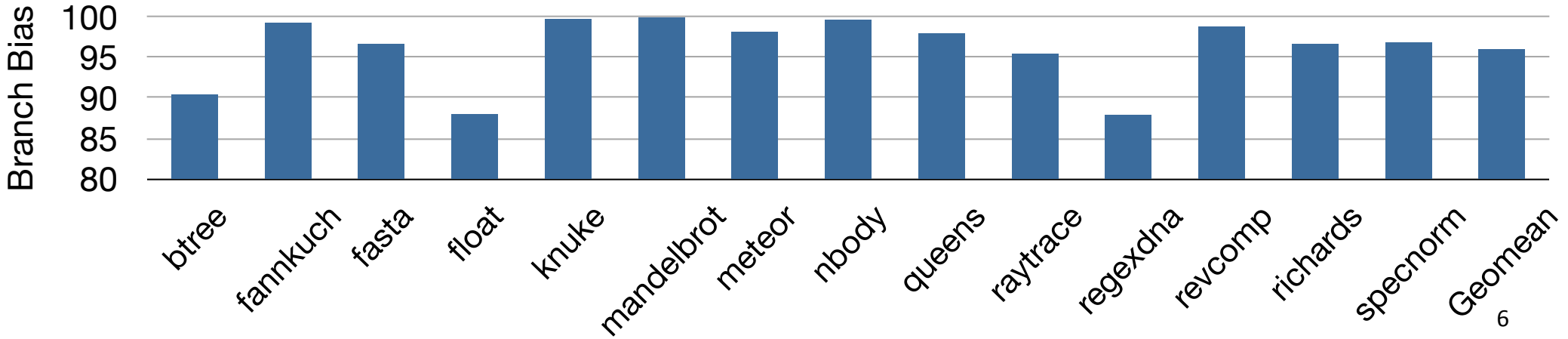


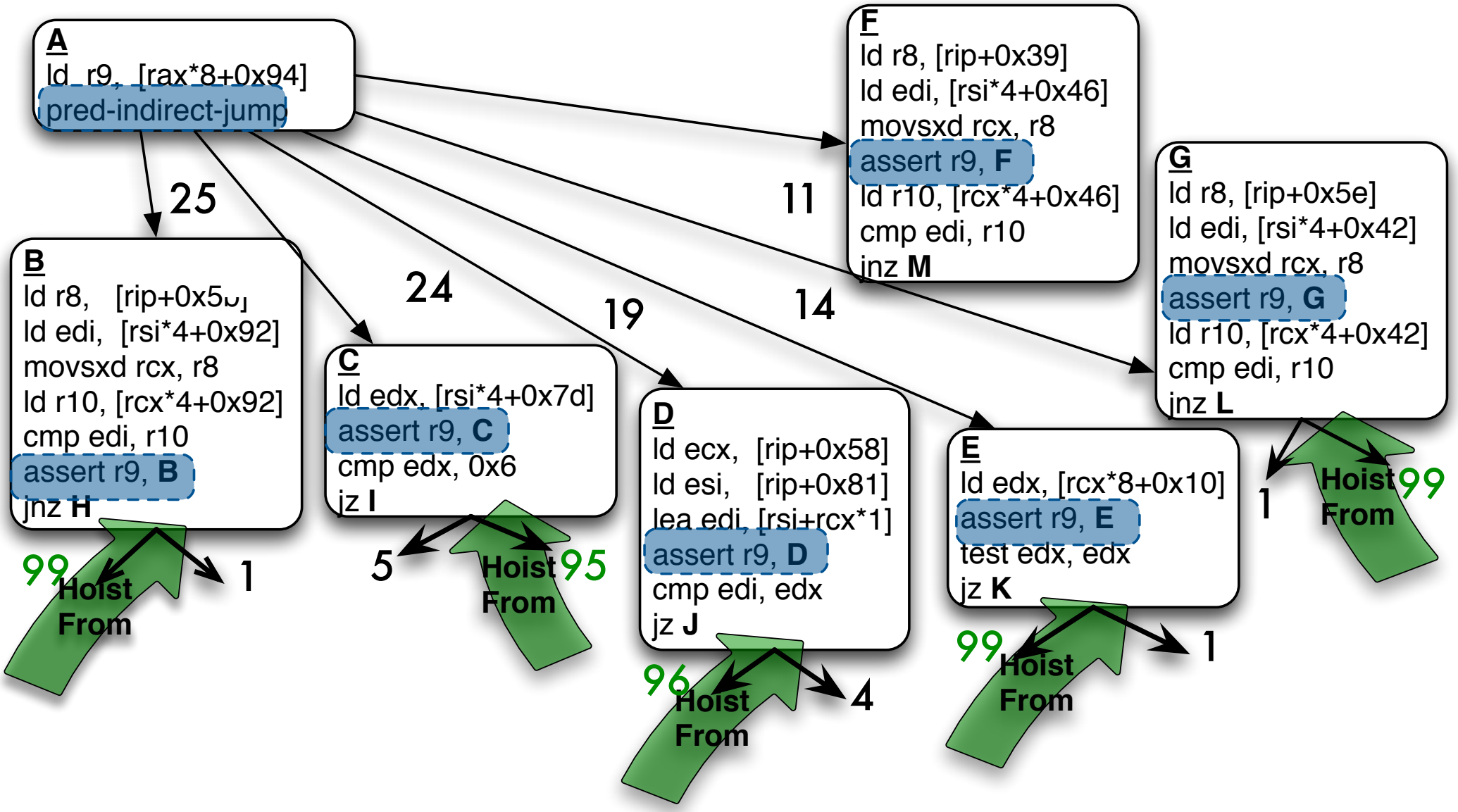
# Missed Optimization Opportunity: Next Branch Bias



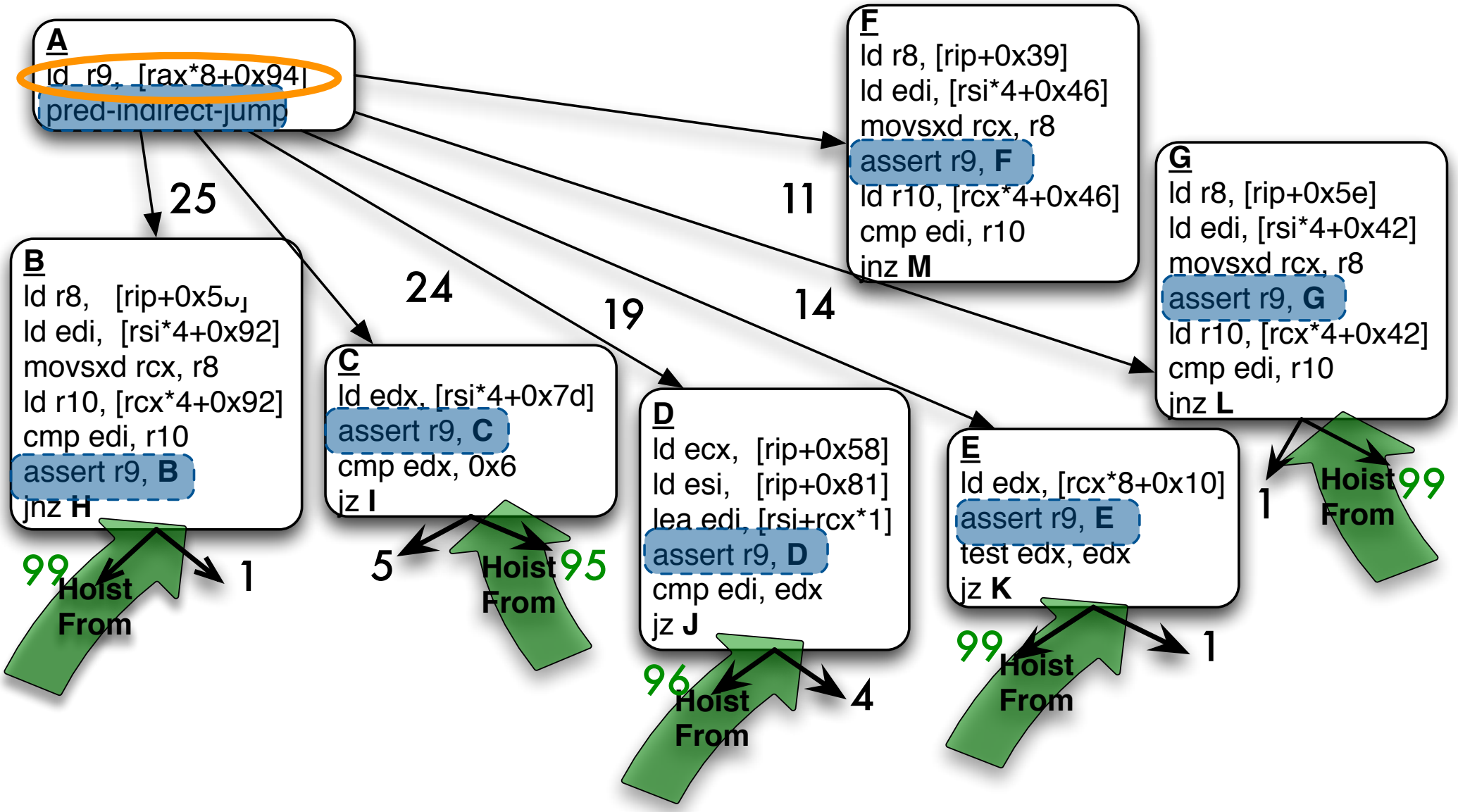



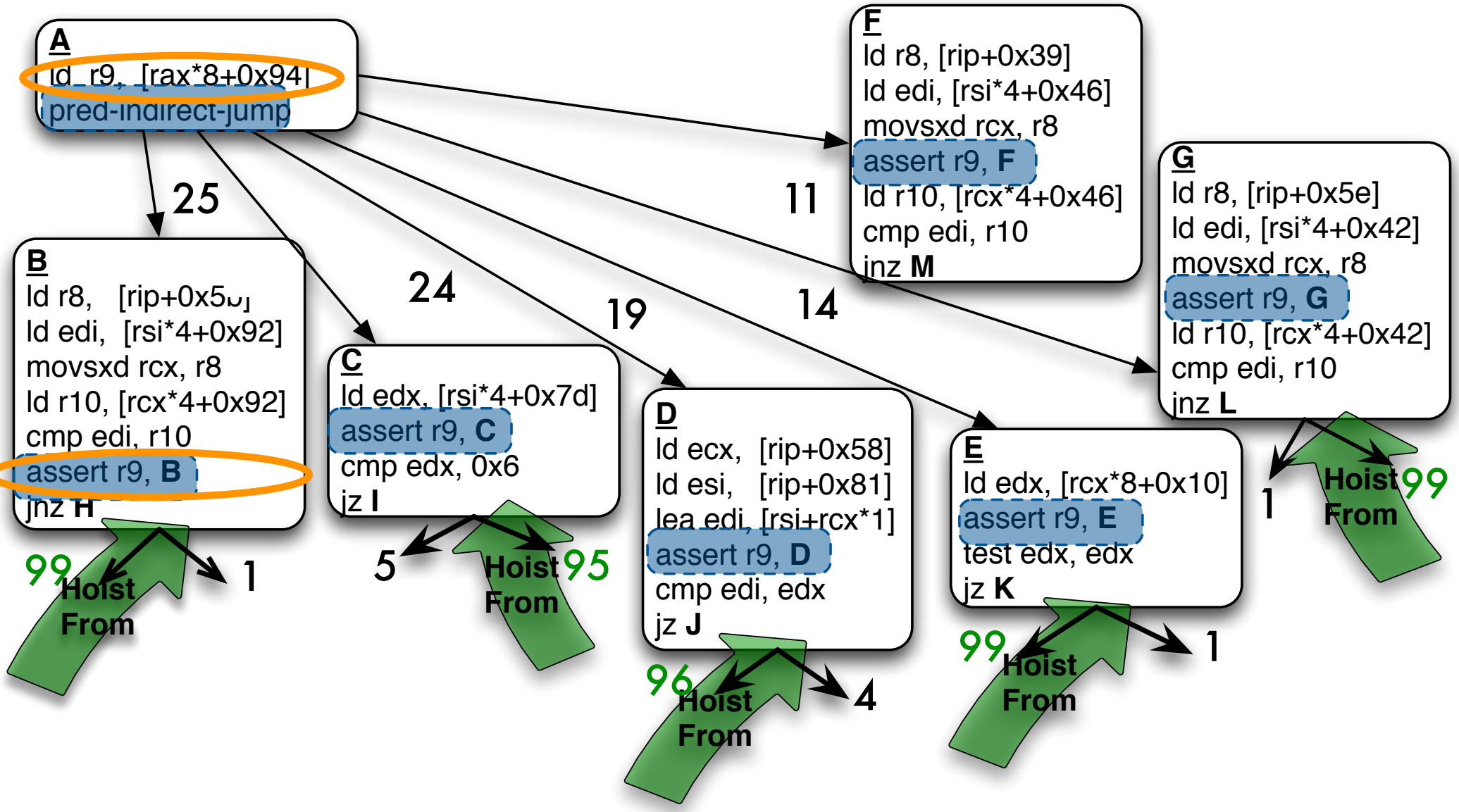
# Exploiting Predictability: Benefit from Next Branch Bias



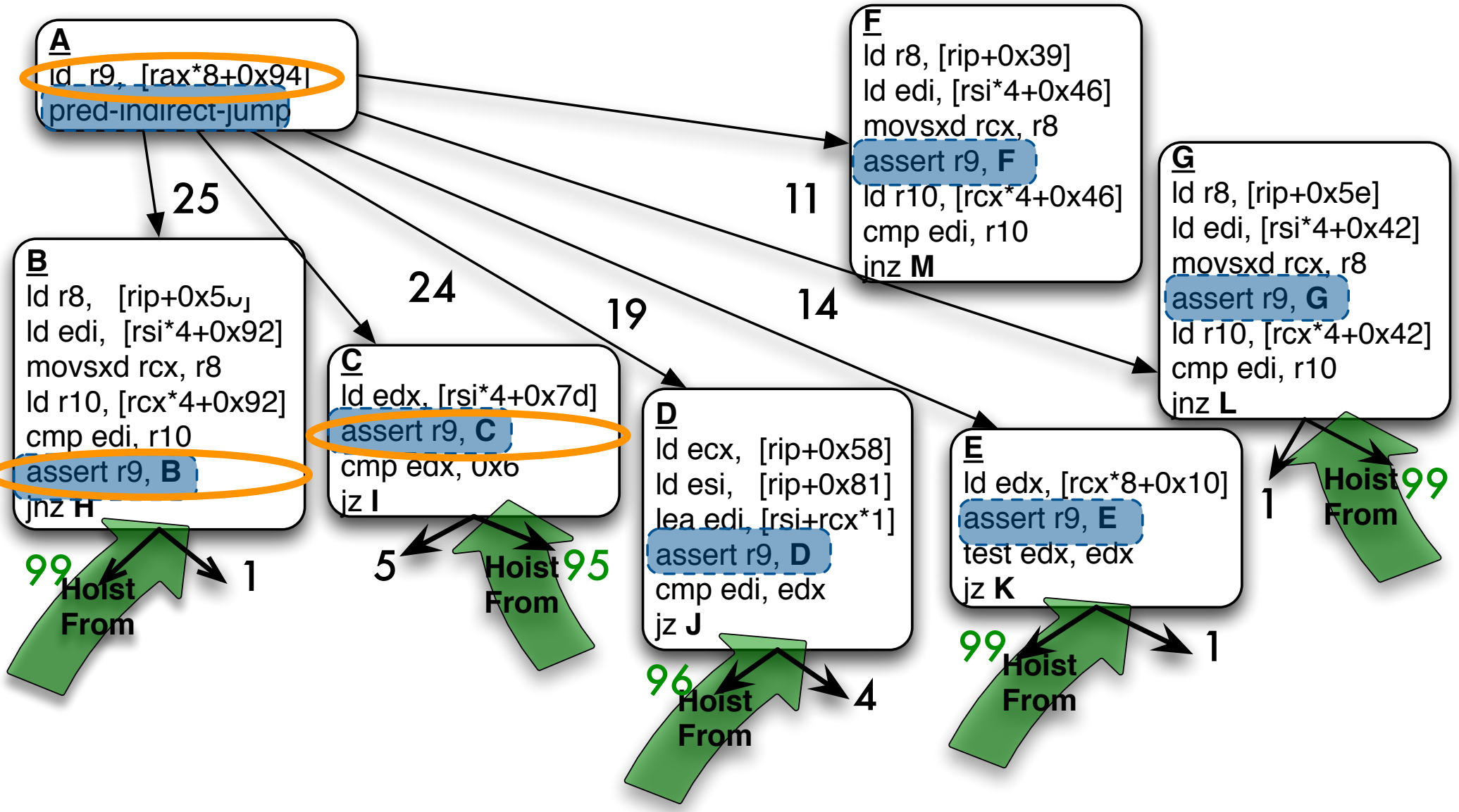
# Exploiting Predictability: Benefit from Next Branch Bias



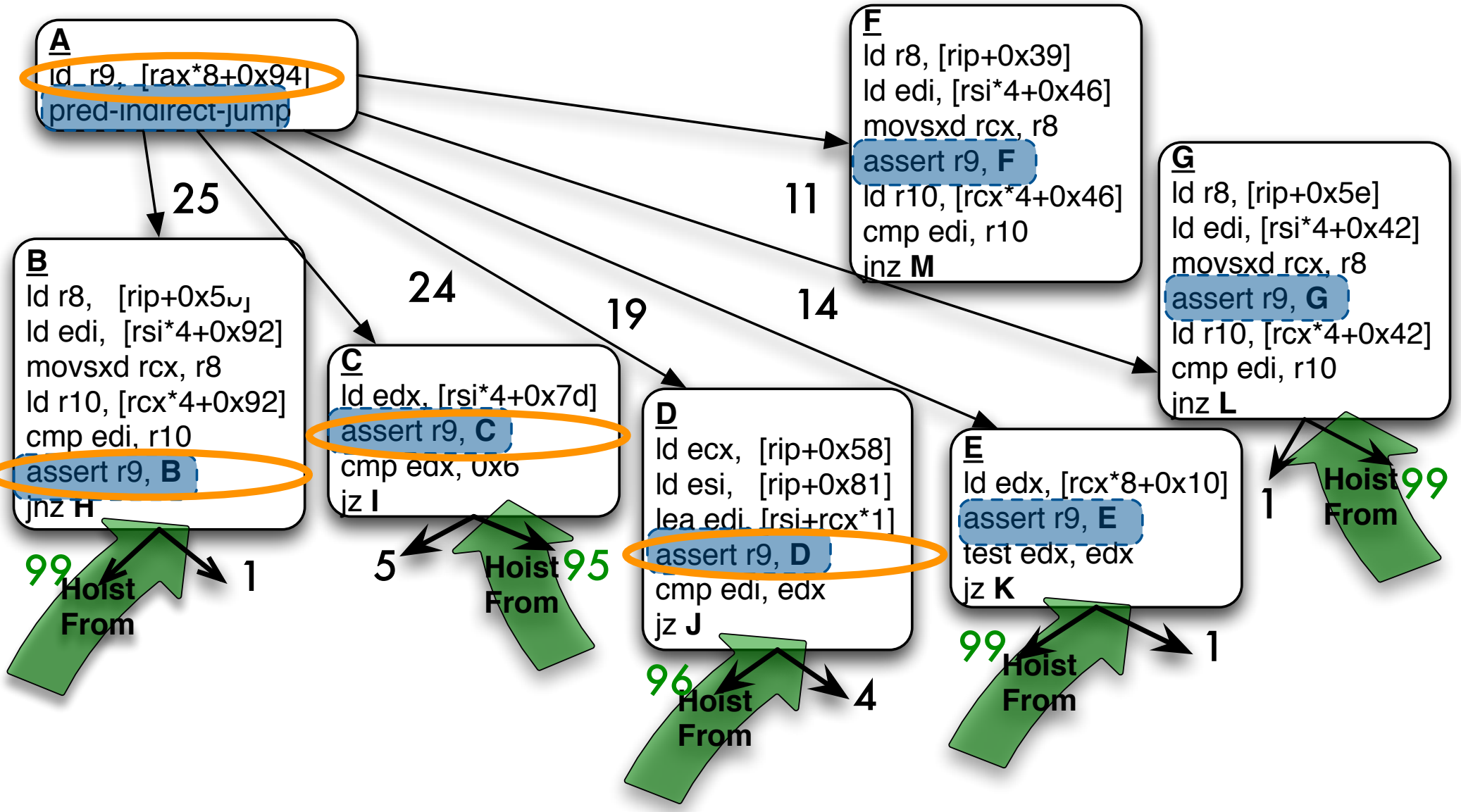
# Exploiting Predictability: Benefit from Next Branch Bias



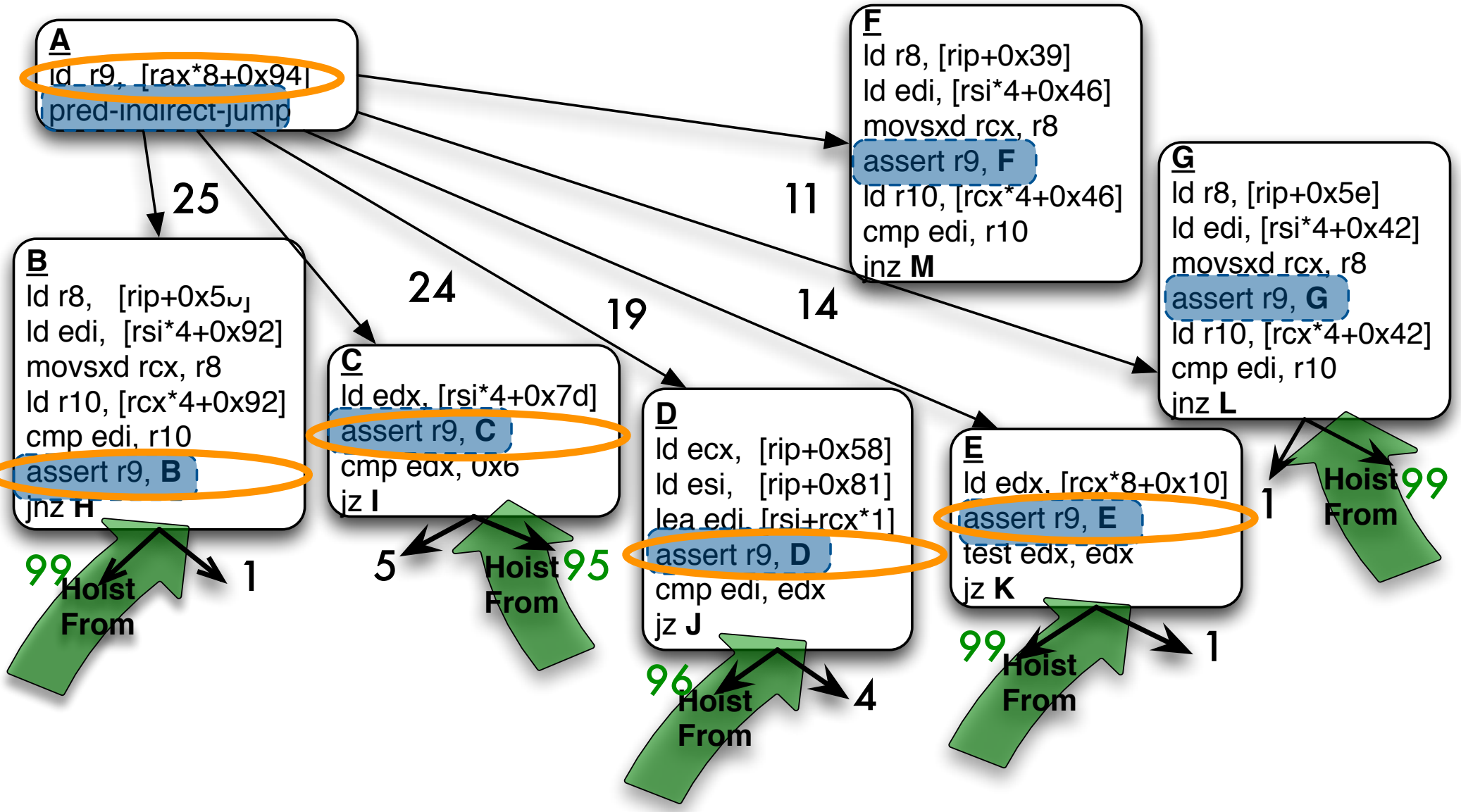
# Exploiting Predictability: Benefit from Next Branch Bias



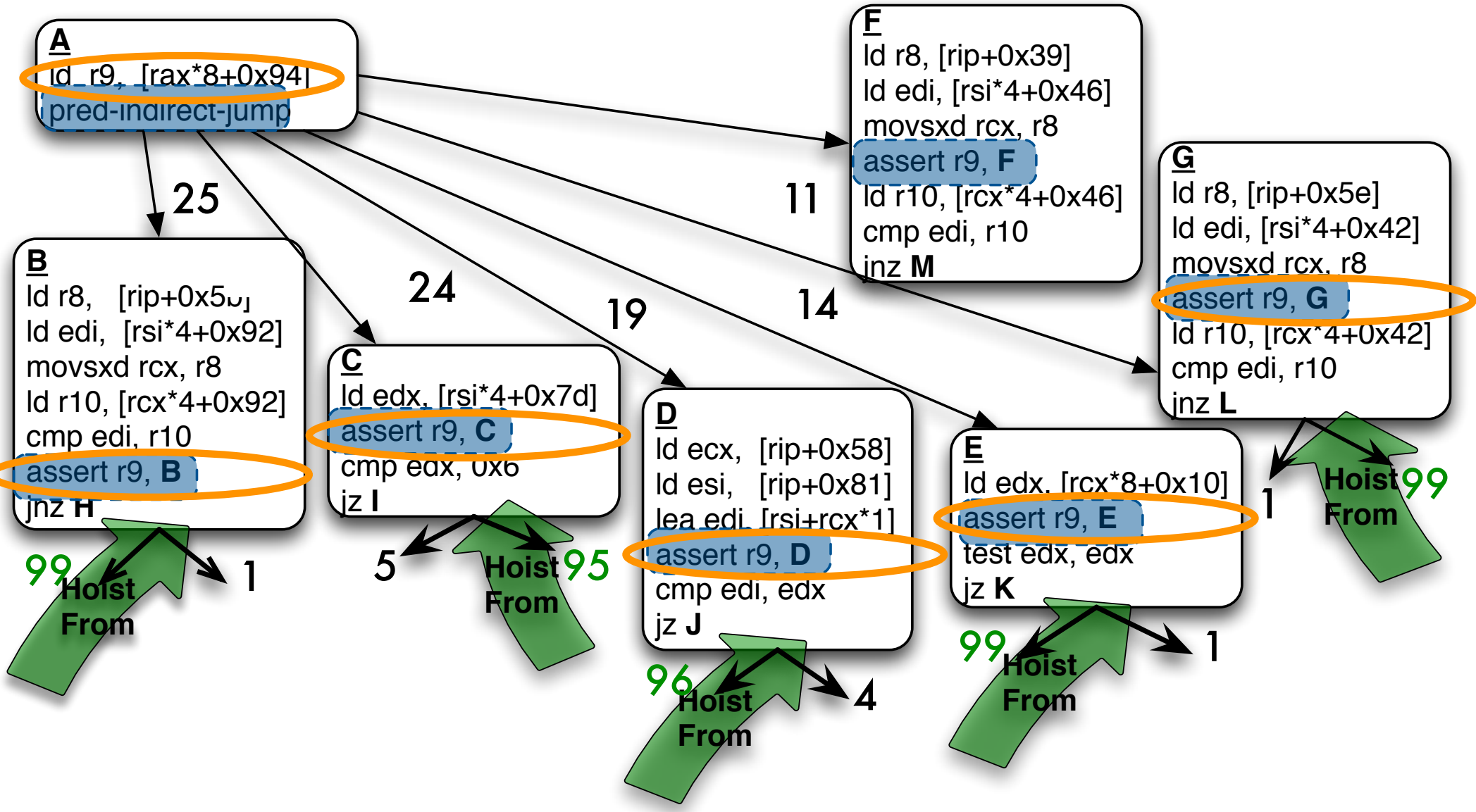
# Exploiting Predictability: Benefit from Next Branch Bias



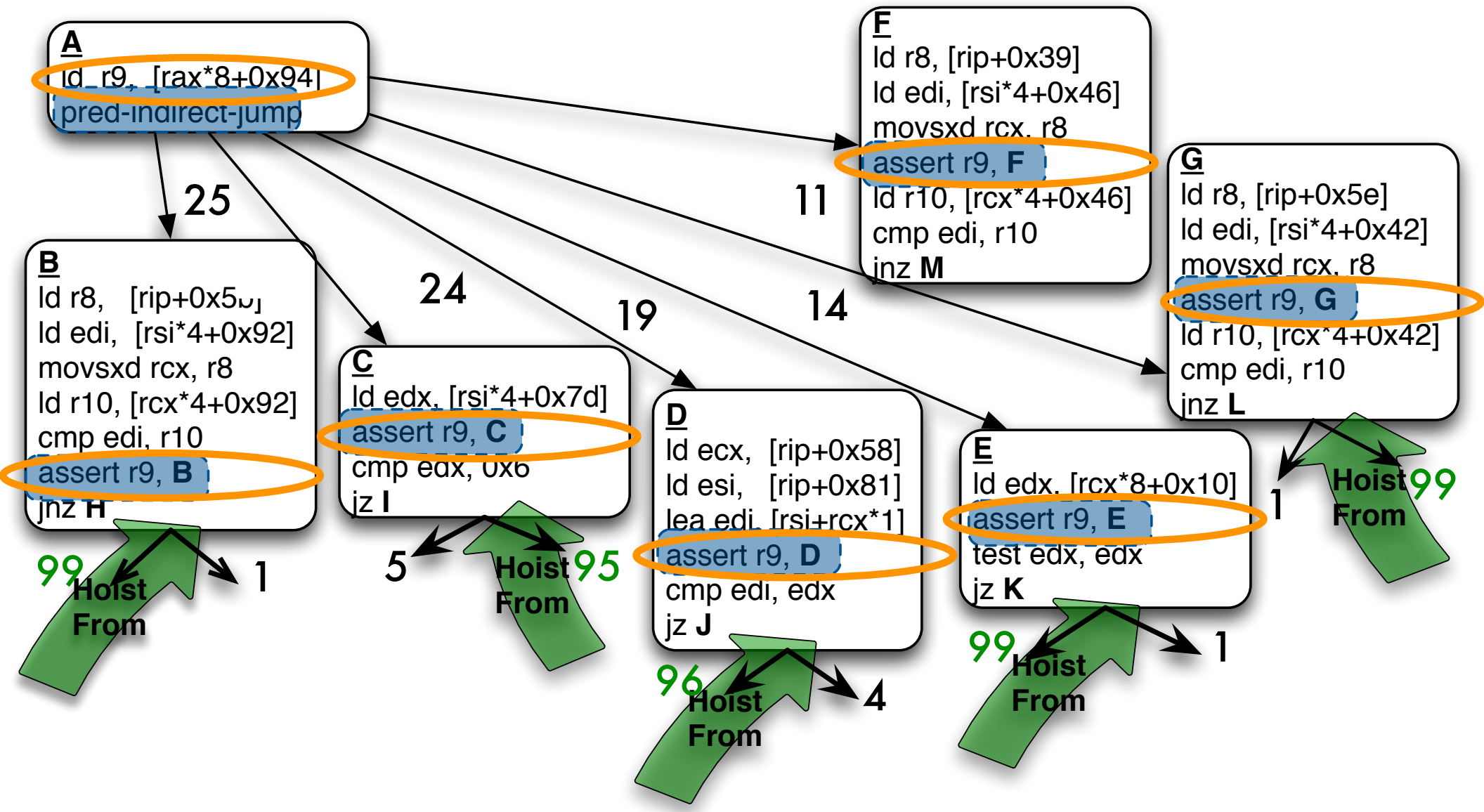
# Exploiting Predictability: Benefit from Next Branch Bias



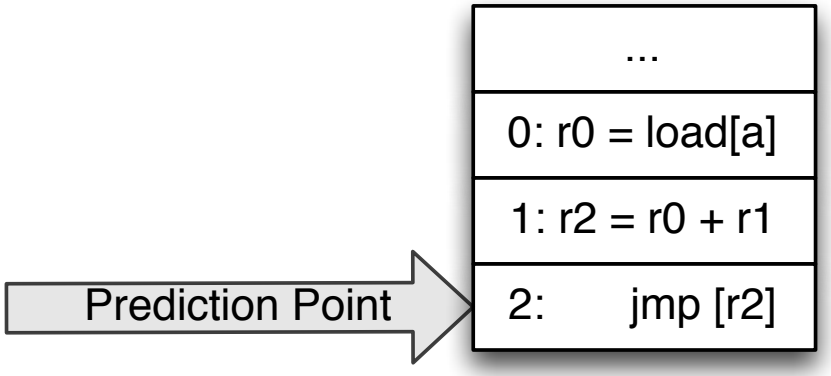
# Exploiting Predictability: Benefit from Next Branch Bias



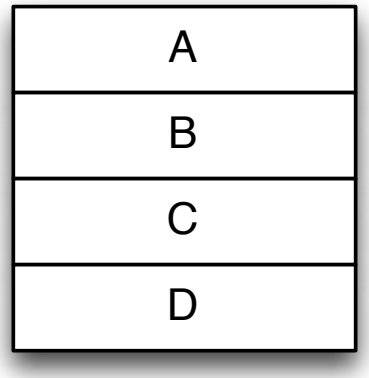
# Exploiting Predictability: Benefit from Next Branch Bias



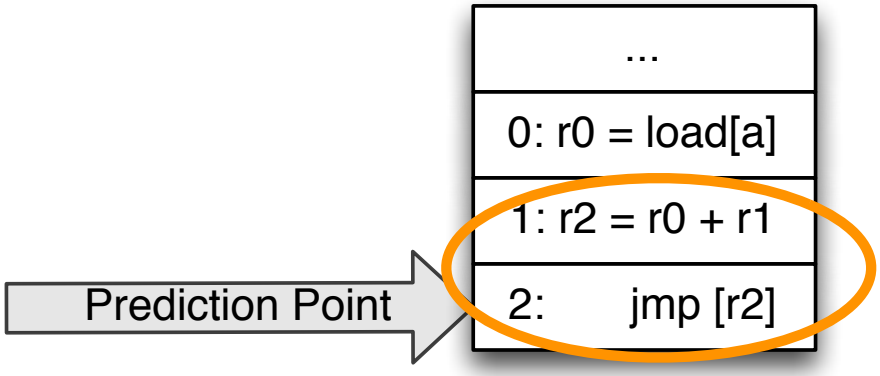
# Challenge: Invalid Predicted Target Address




Valid Targets: {A, G}
Predict A

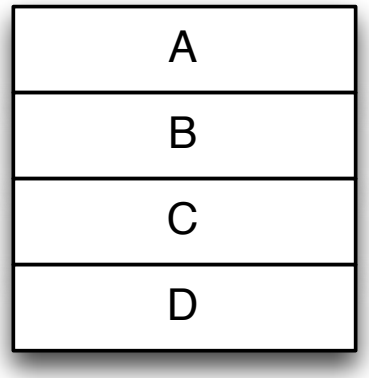


# Challenge: Invalid Predicted Target Address

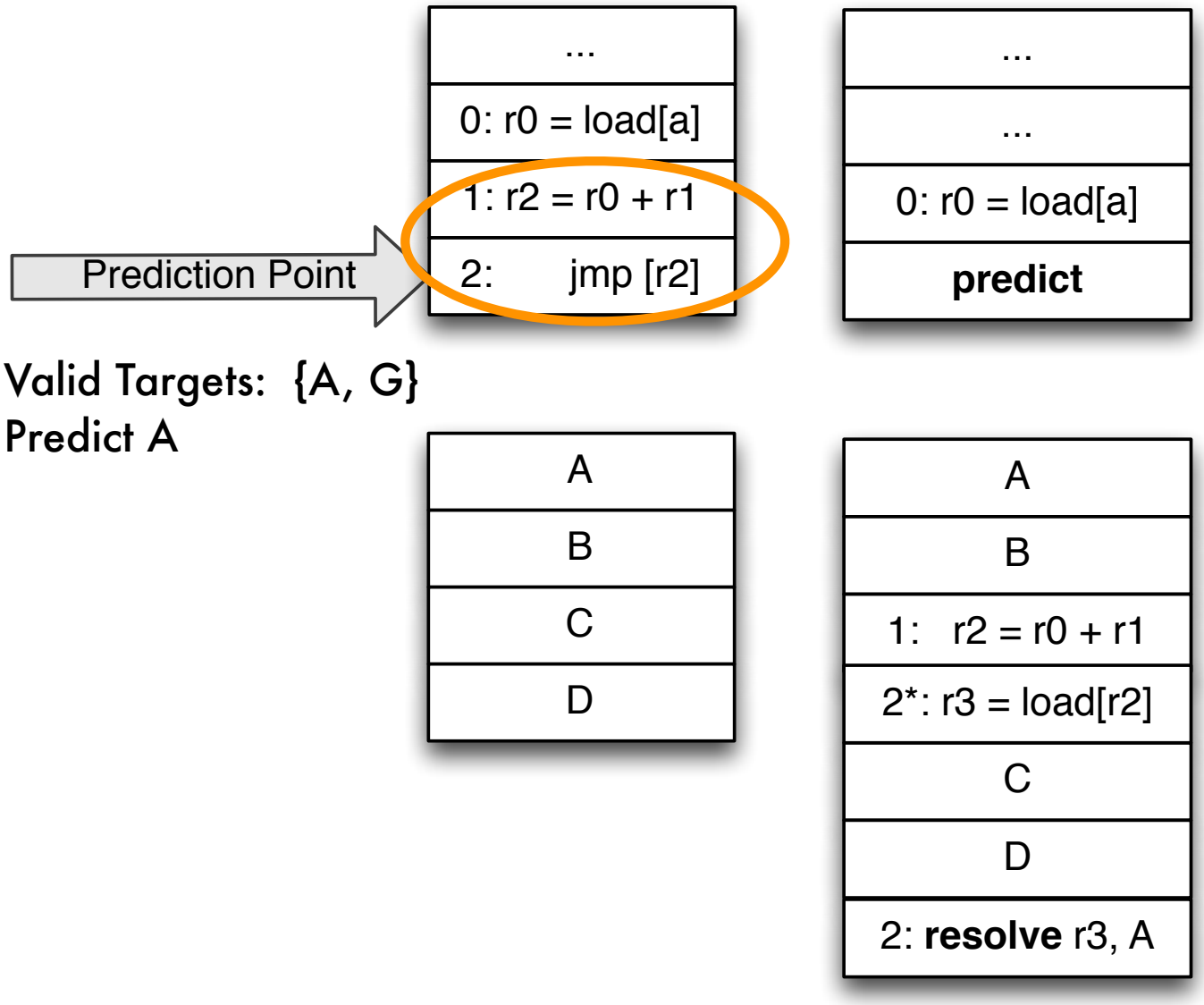



Valid Targets: {A, G}

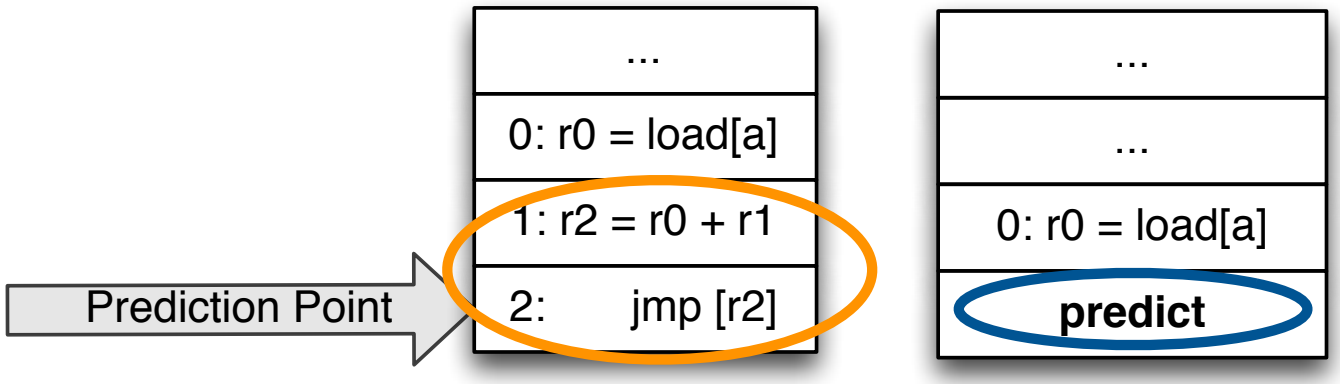
Predict A



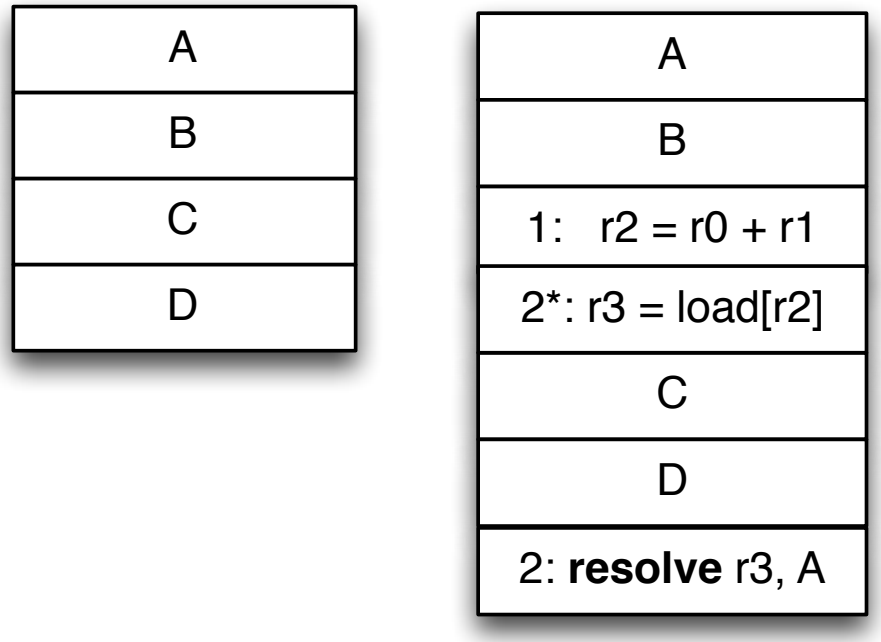
# Challenge: Invalid Predicted Target Address



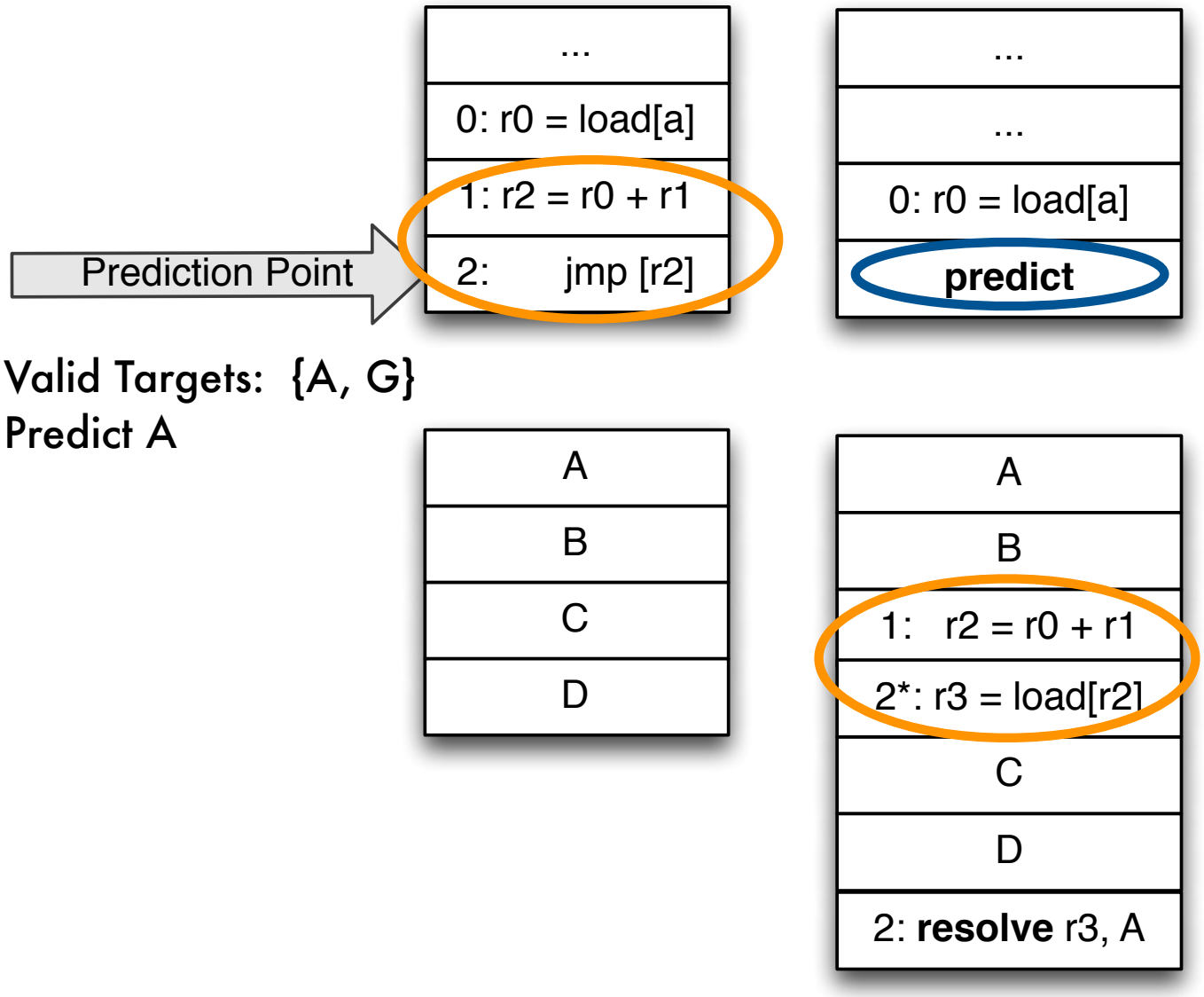
# Challenge: Invalid Predicted Target Address



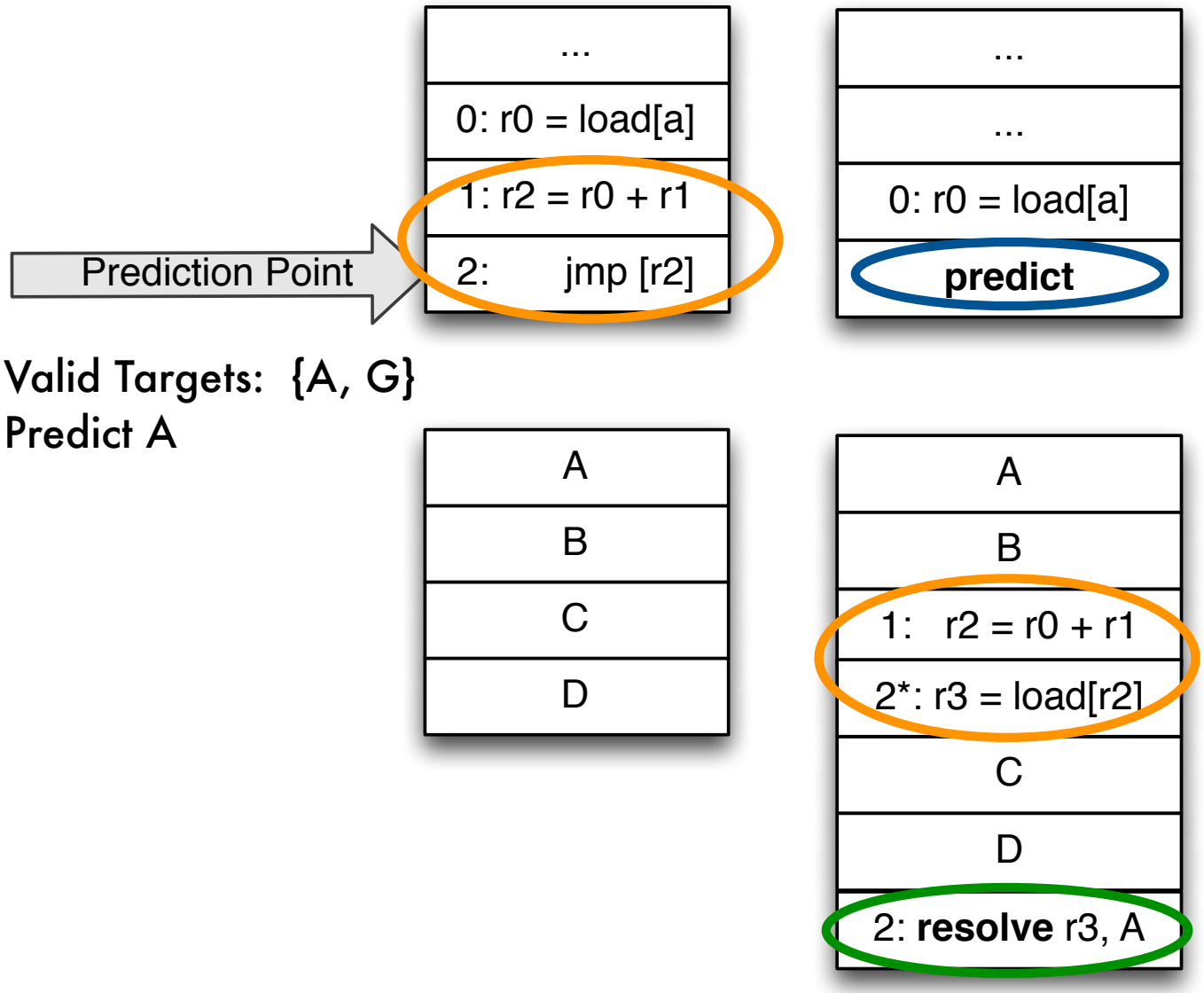

Valid Targets: {A, G}
Predict A



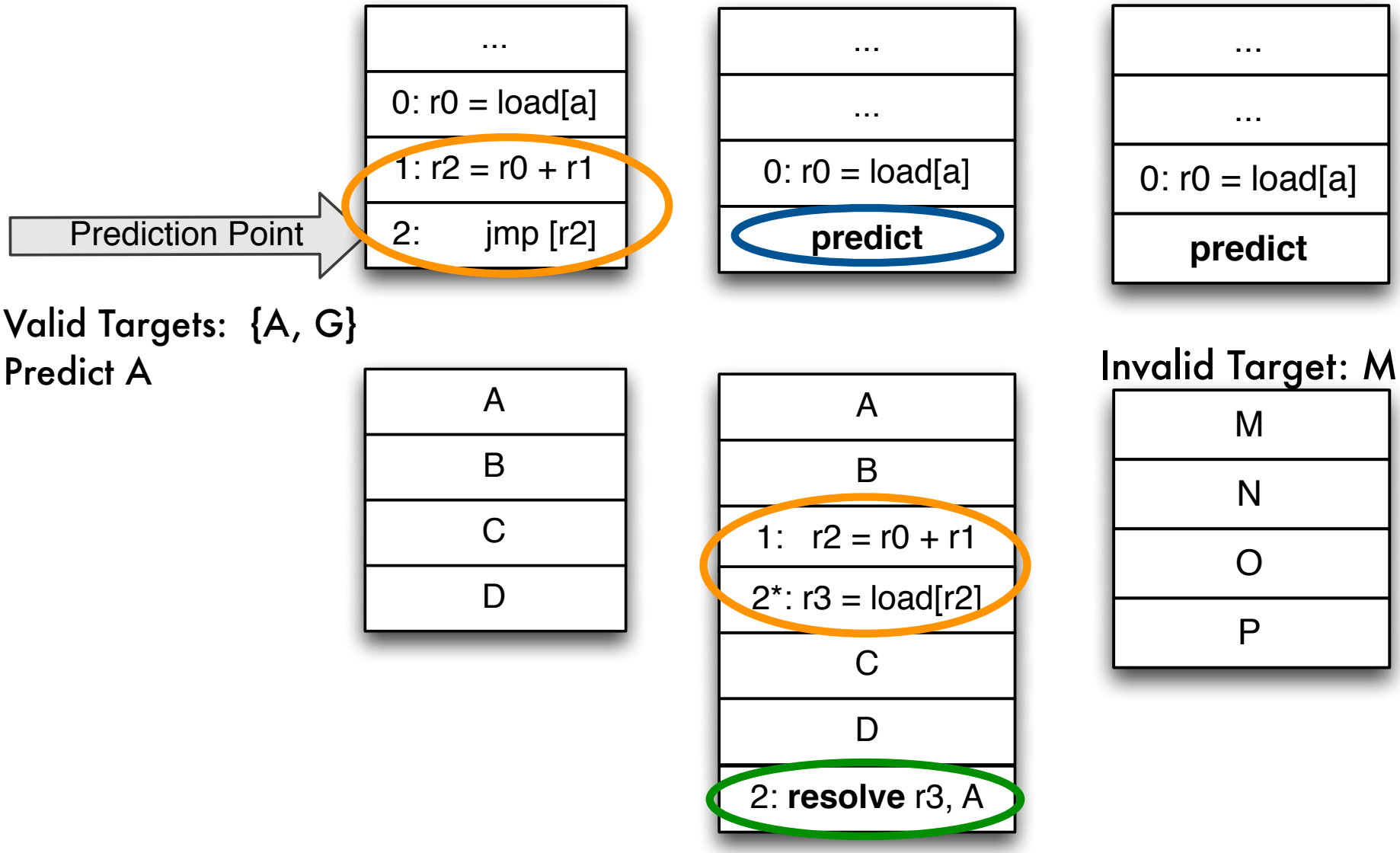
# Challenge: Invalid Predicted Target Address



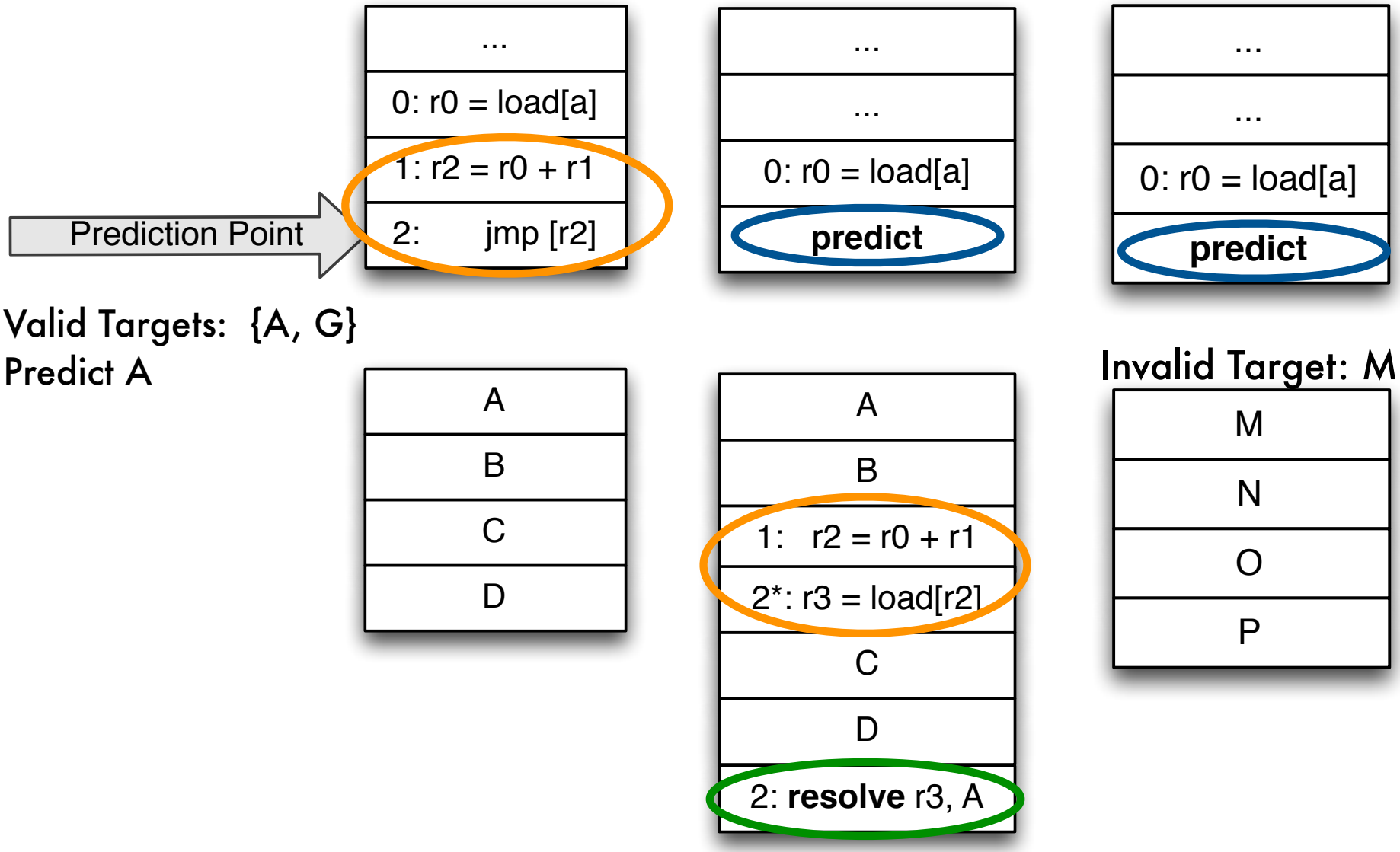
# Challenge: Invalid Predicted Target Address



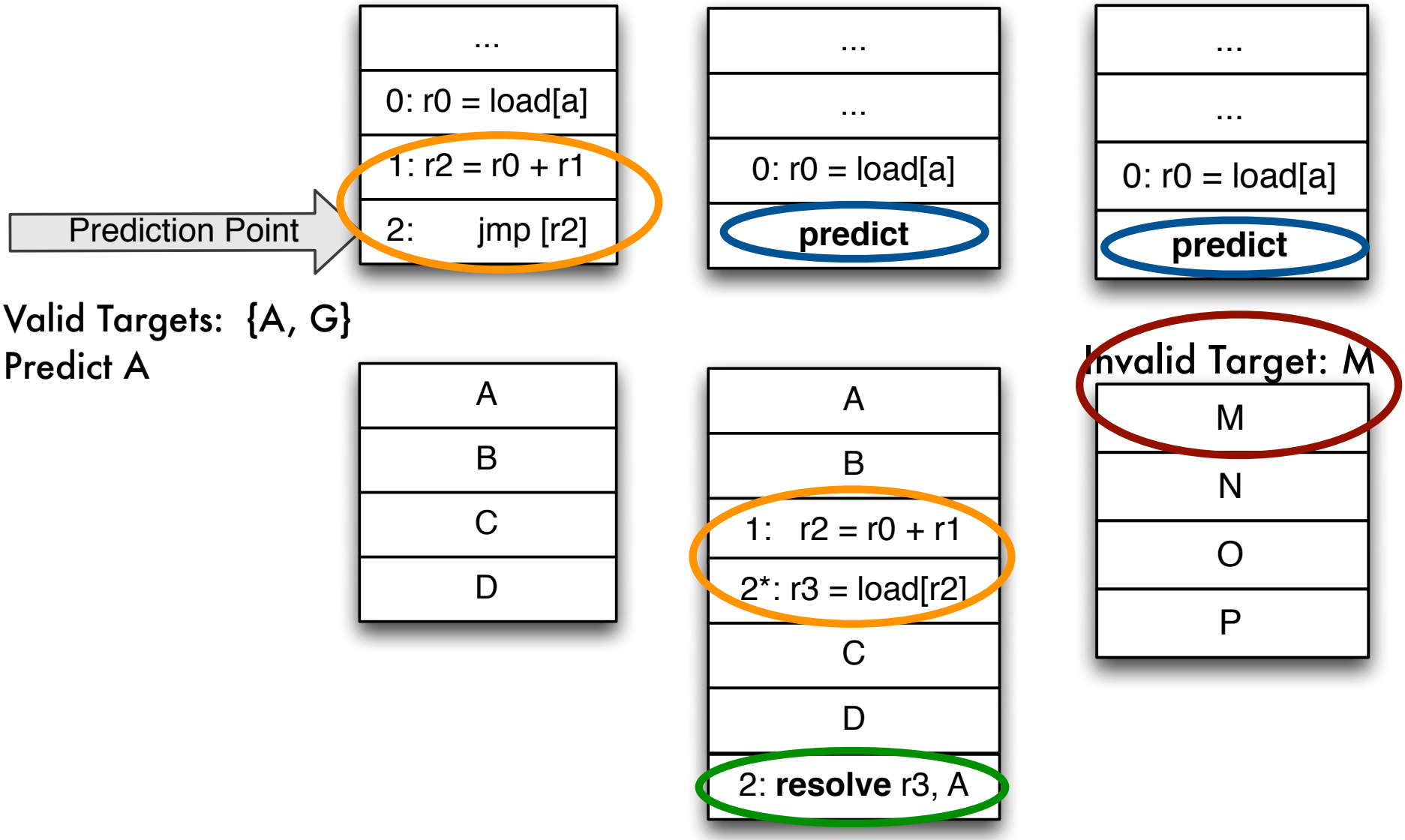
# Challenge: Invalid Predicted Target Address



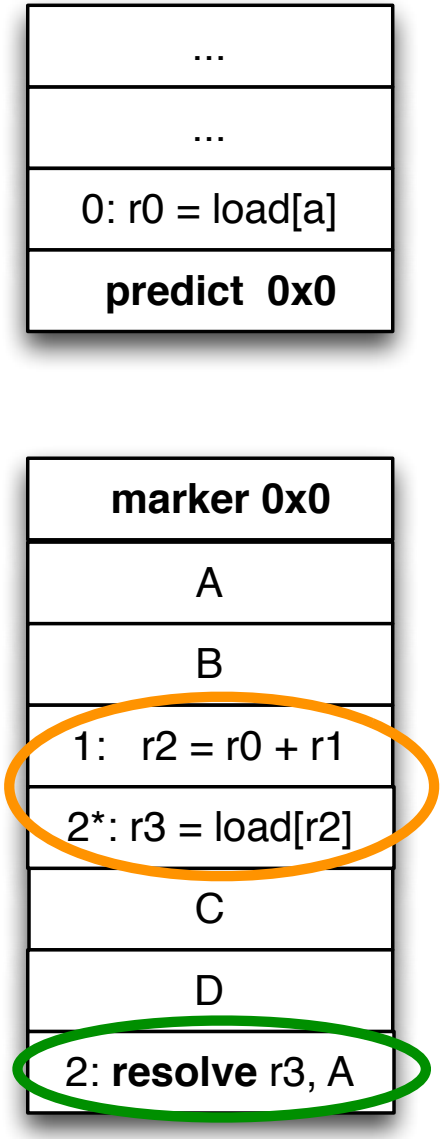
# Challenge: Invalid Predicted Target Address



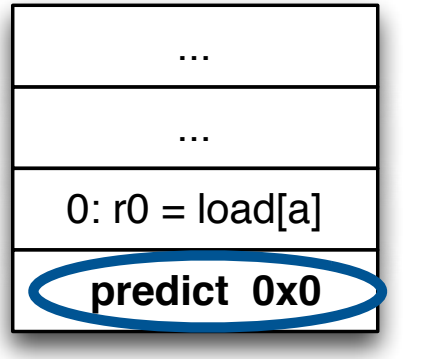
# Challenge: Invalid Predicted Target Address

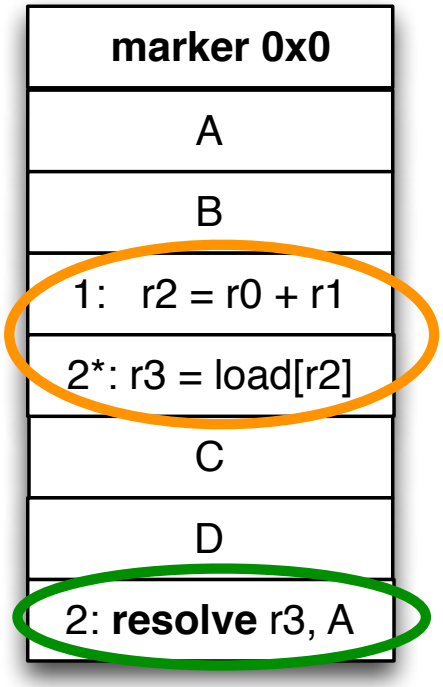


# Solution: Landing Pad

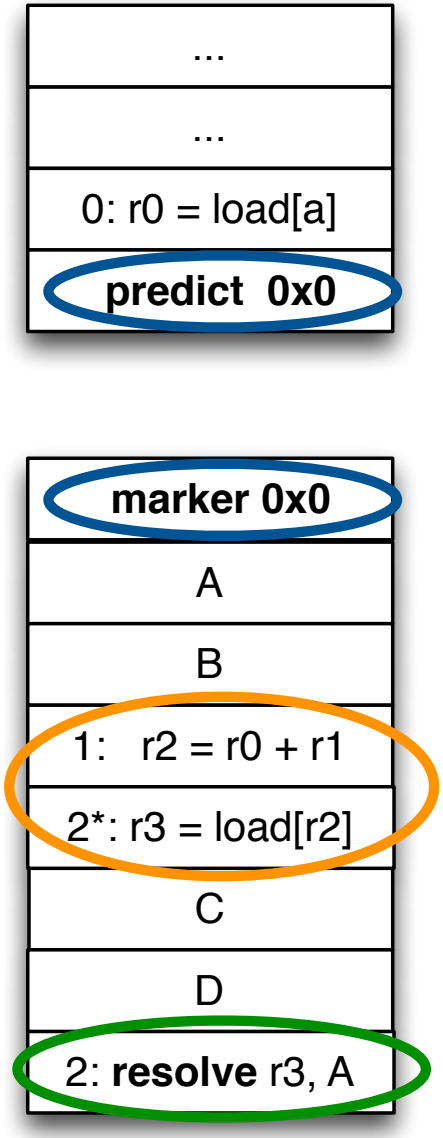


# Solution: Landing Pad

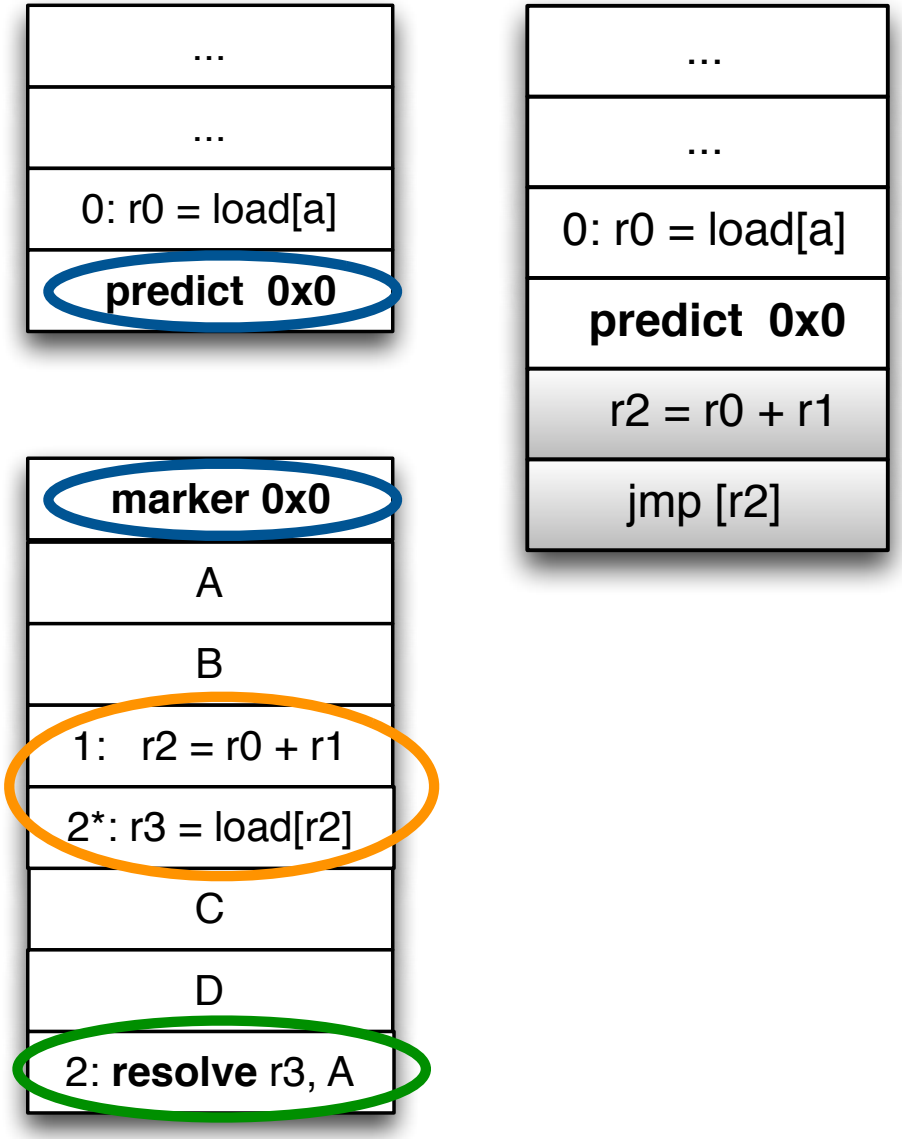


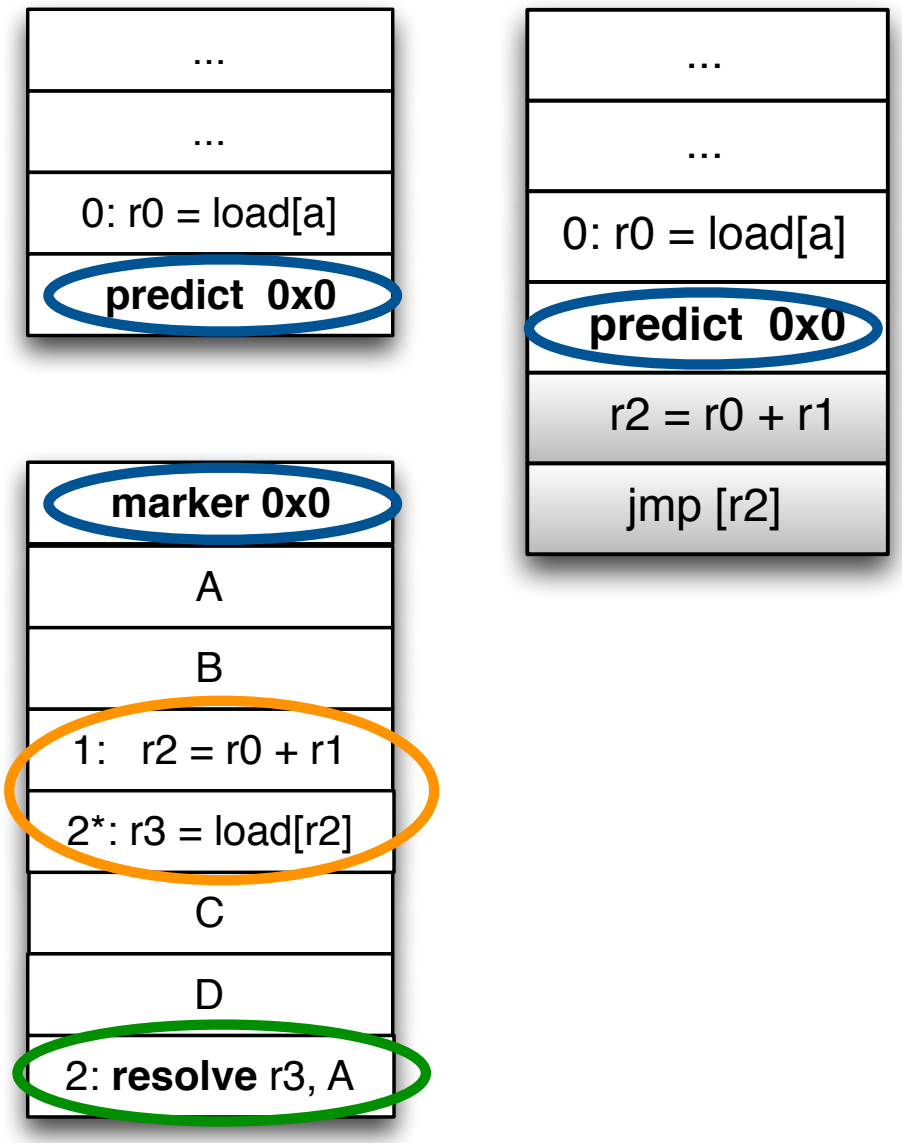
# Solution: Landing Pad



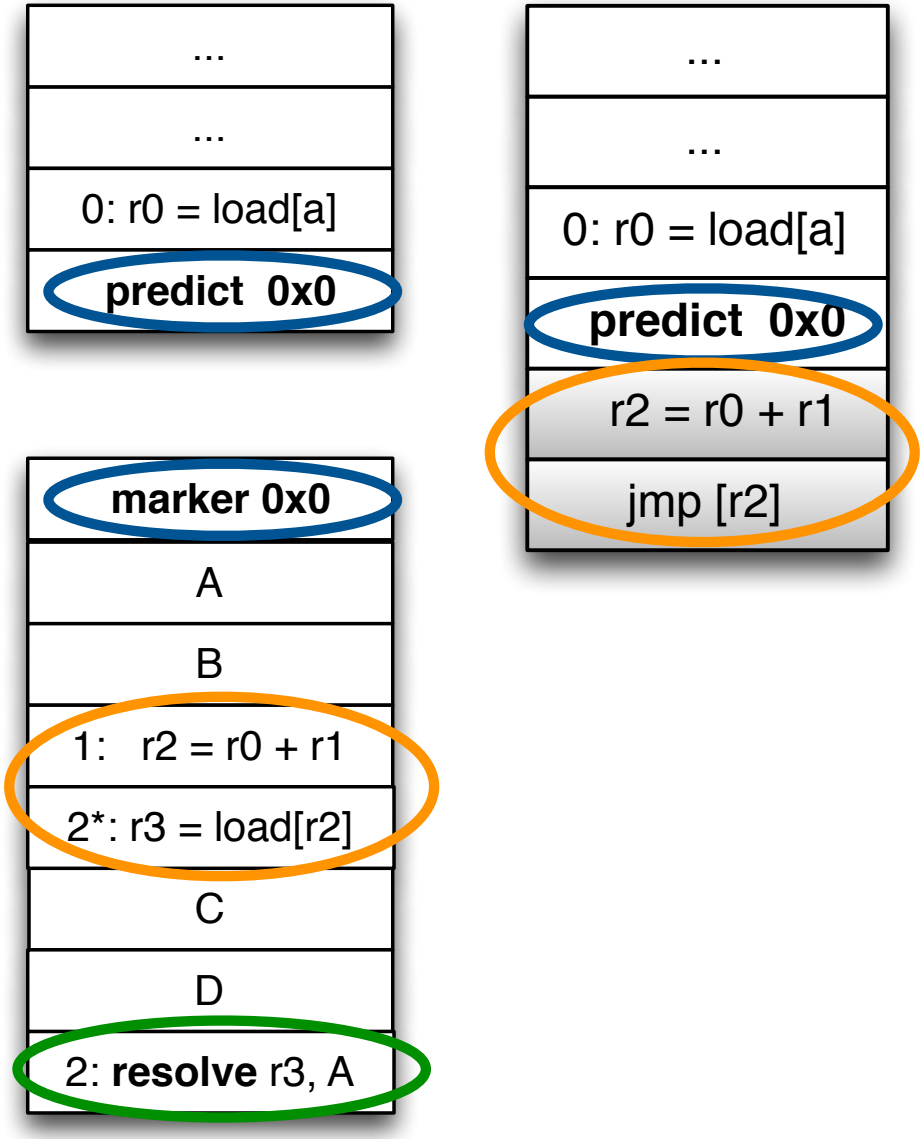
# Solution: Landing Pad



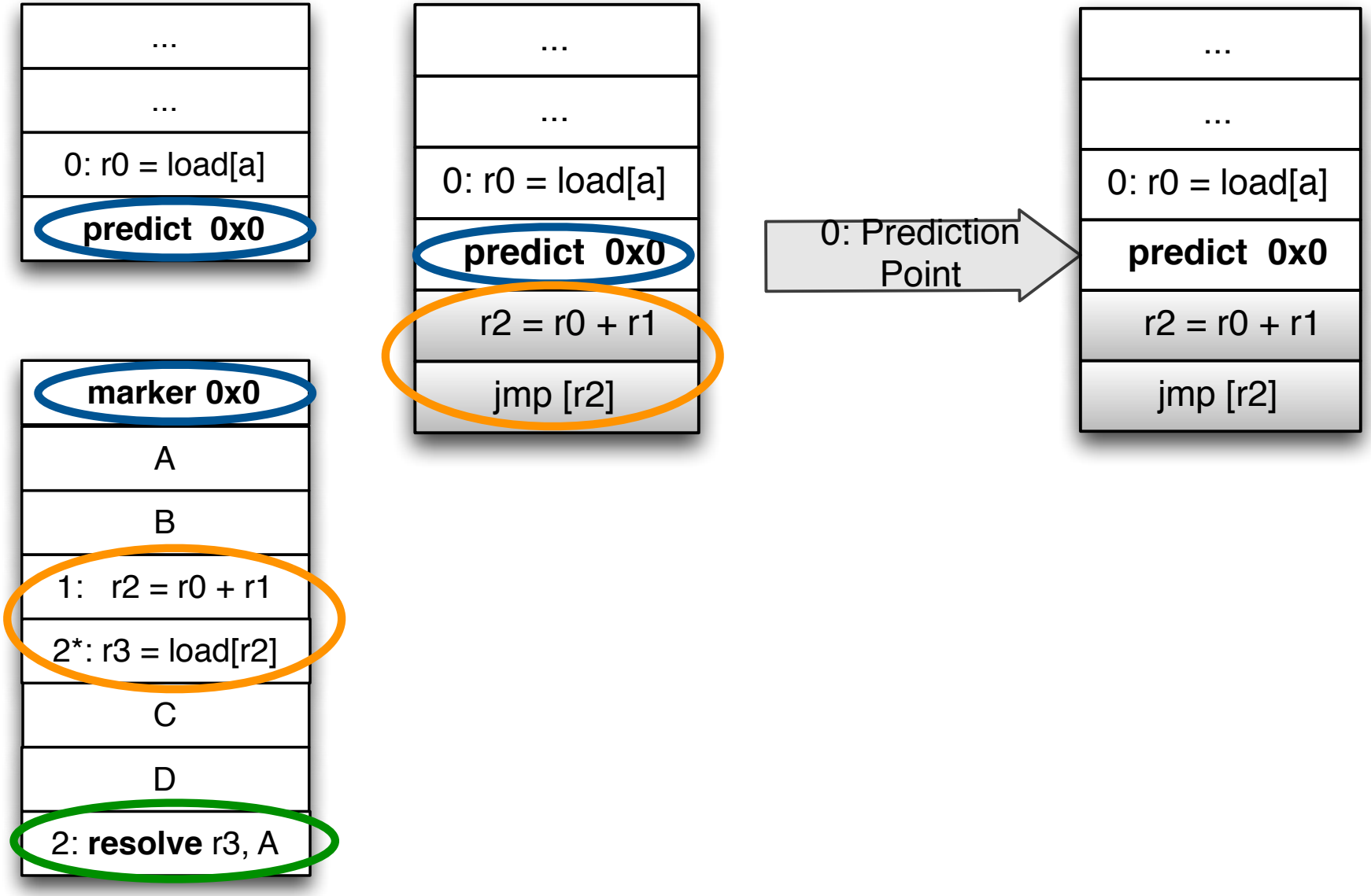
# Solution: Landing Pad



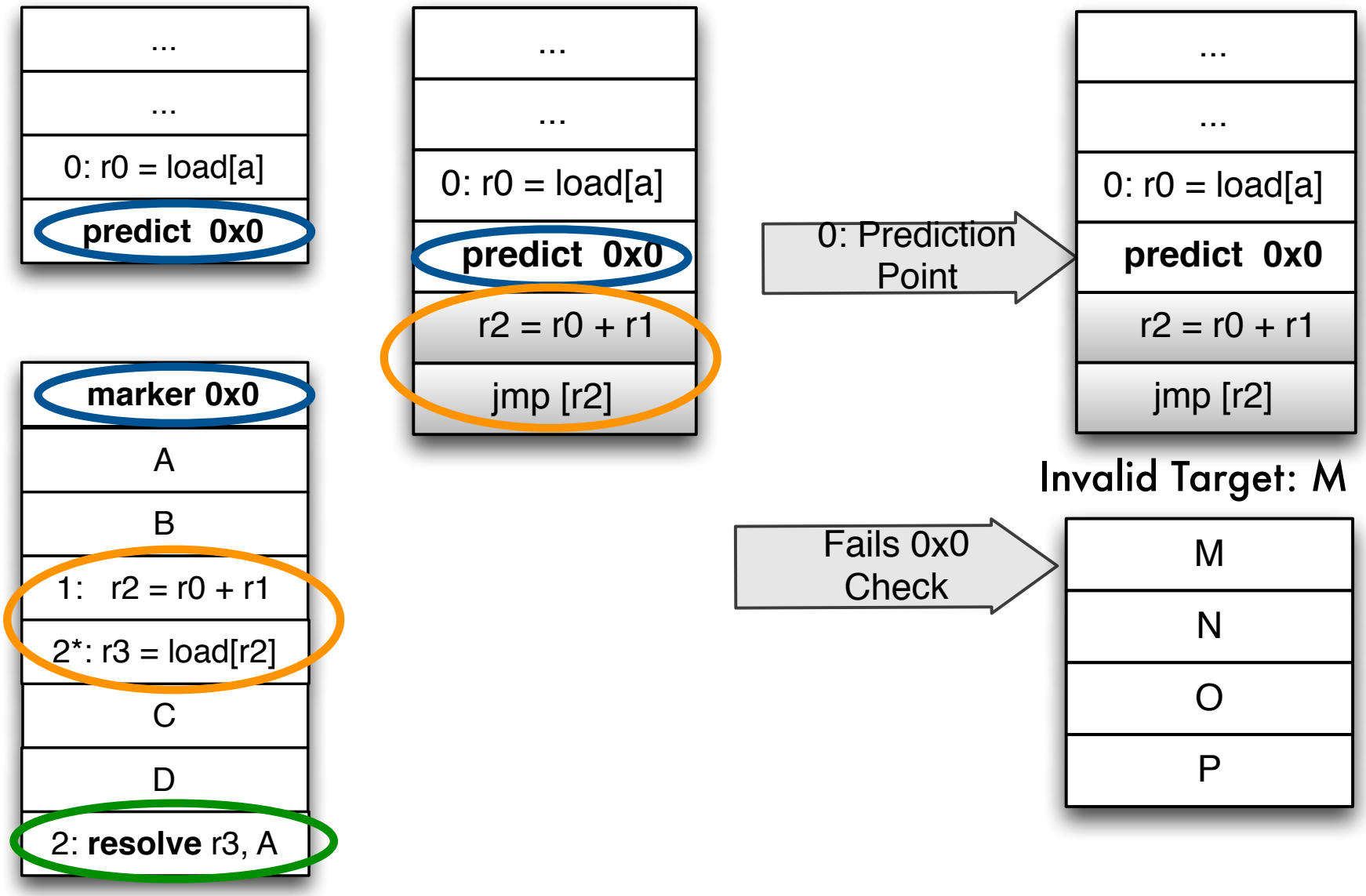
# Solution: Landing Pad



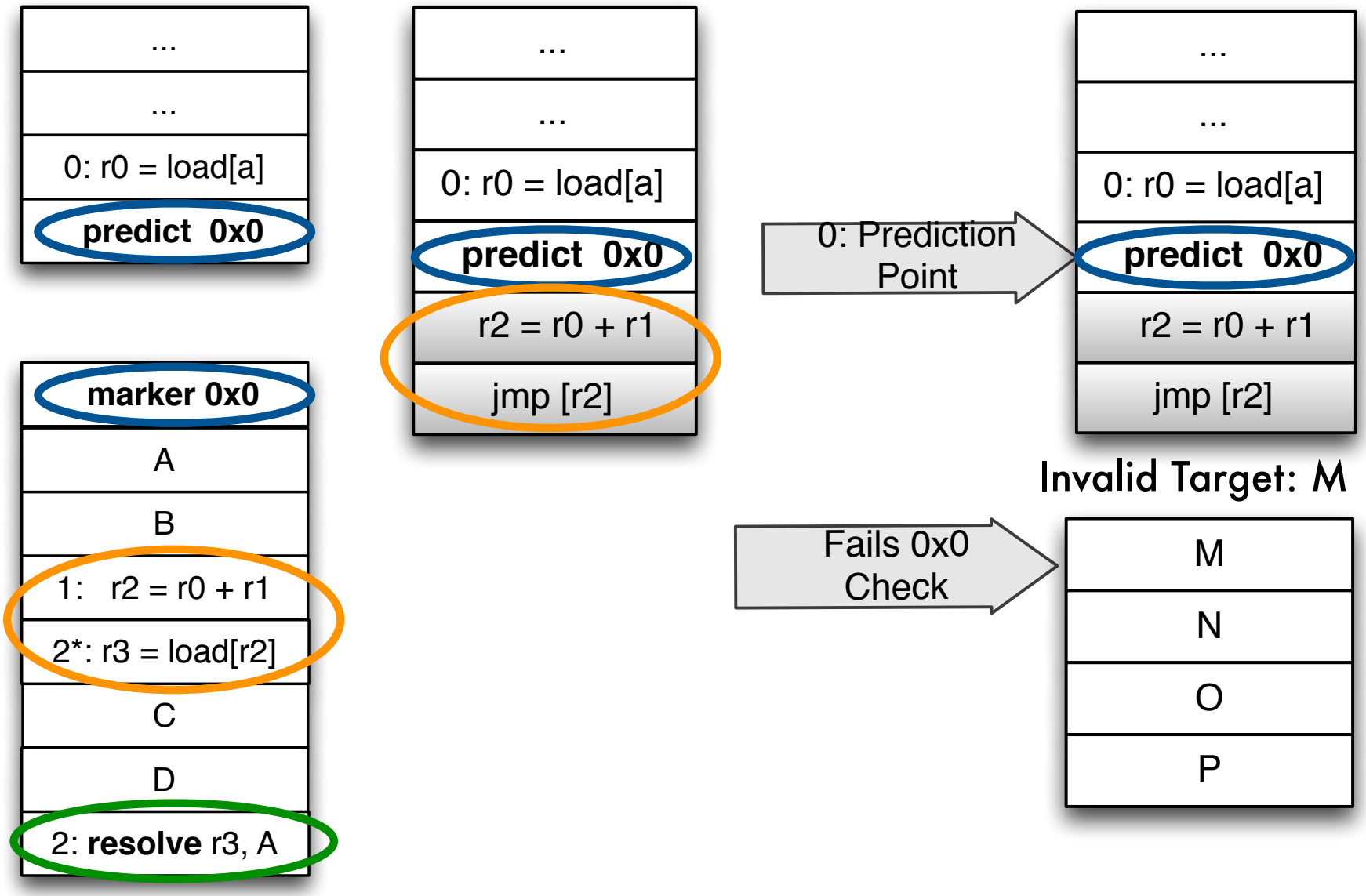
# Solution: Landing Pad



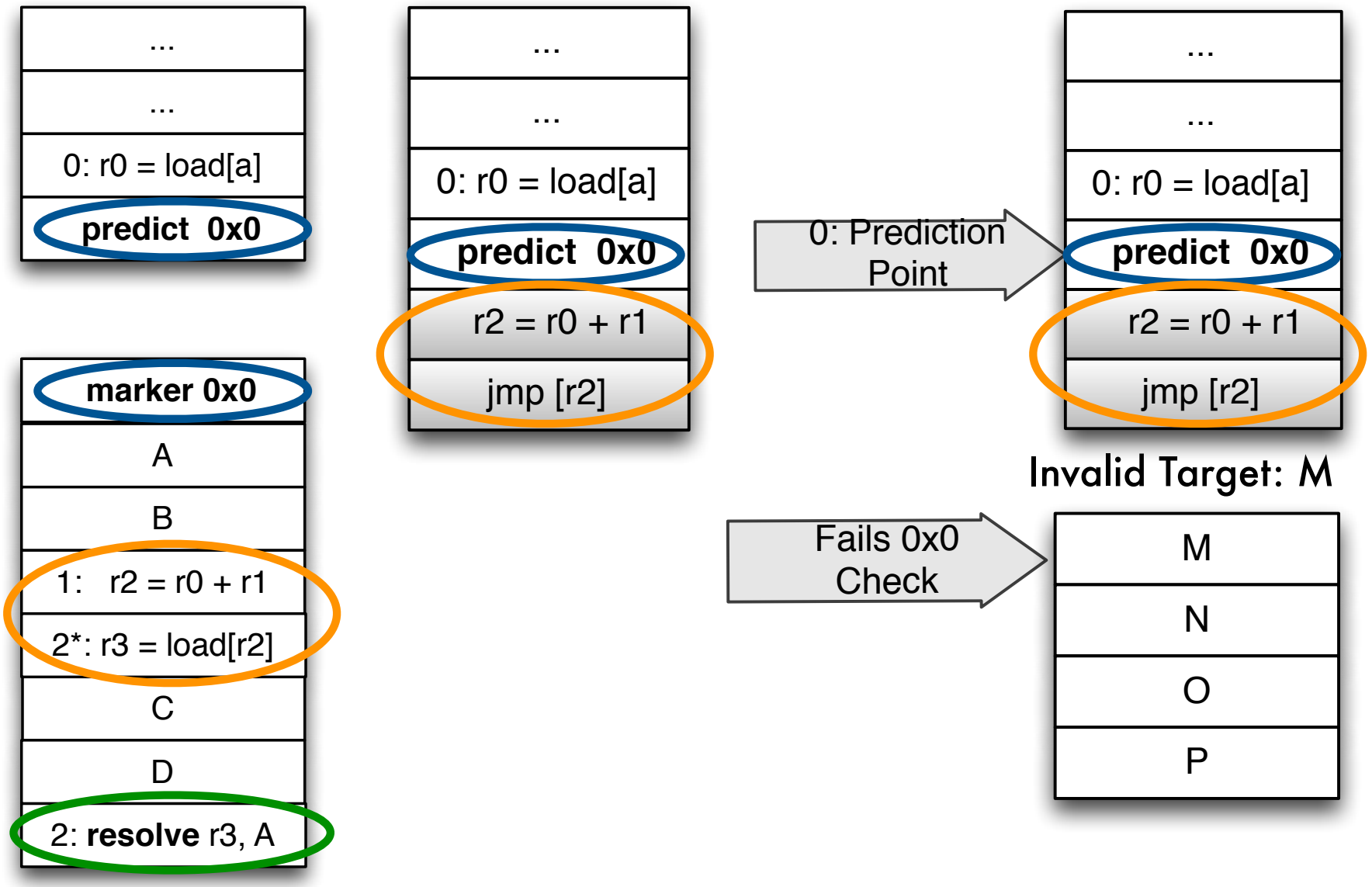
# Solution: Landing Pad



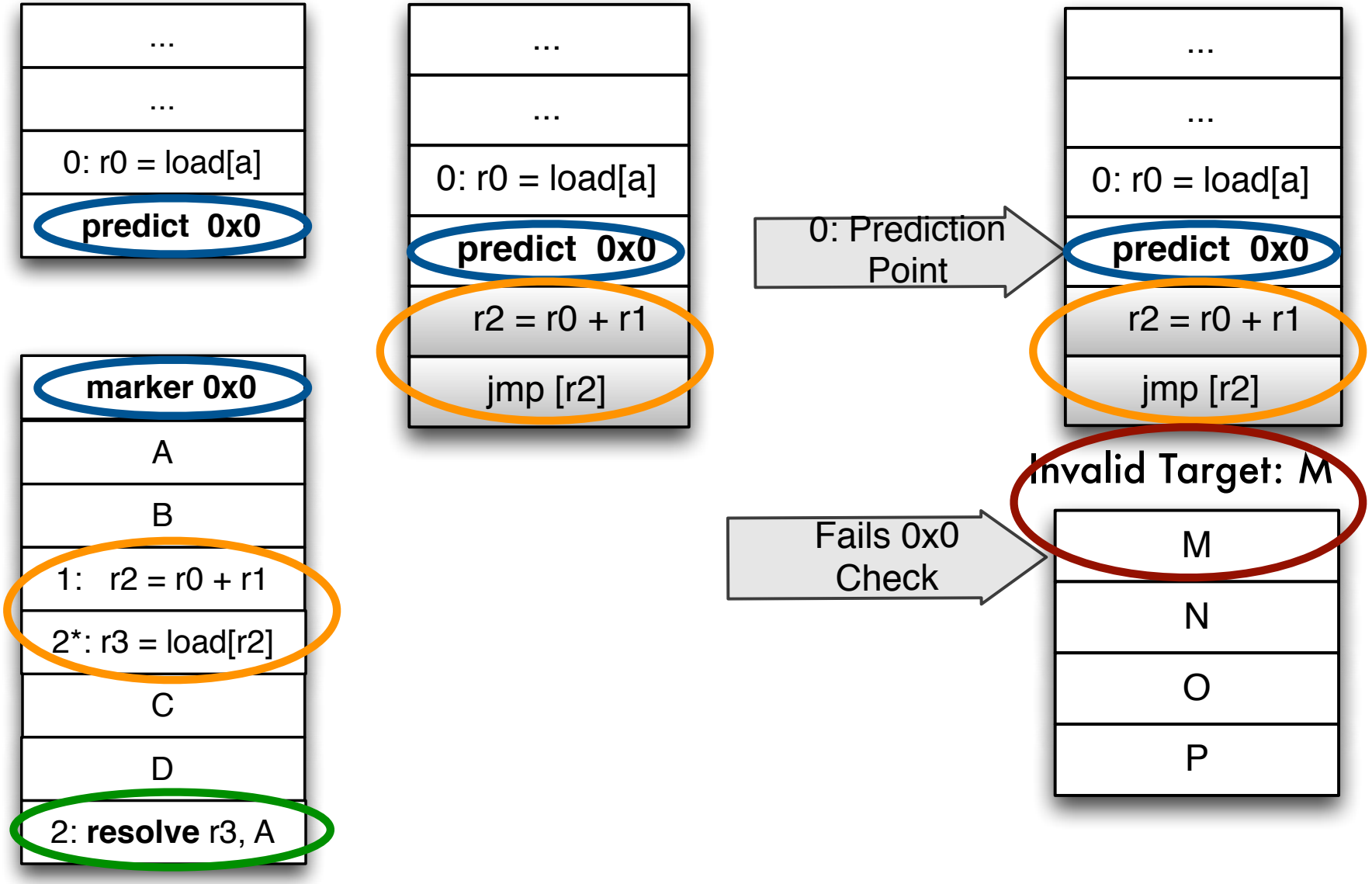
# Solution: Landing Pad



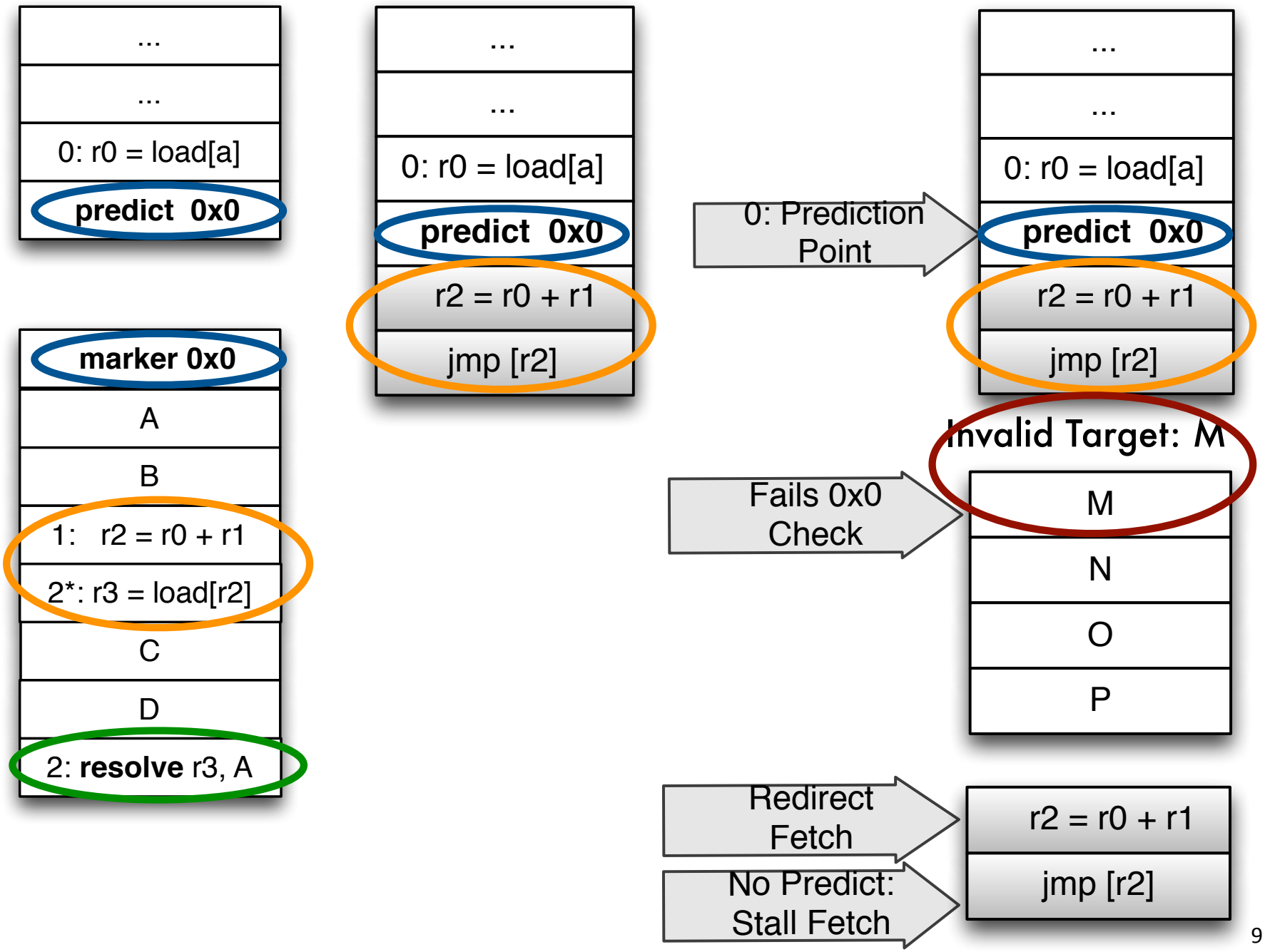
# Solution: Landing Pad



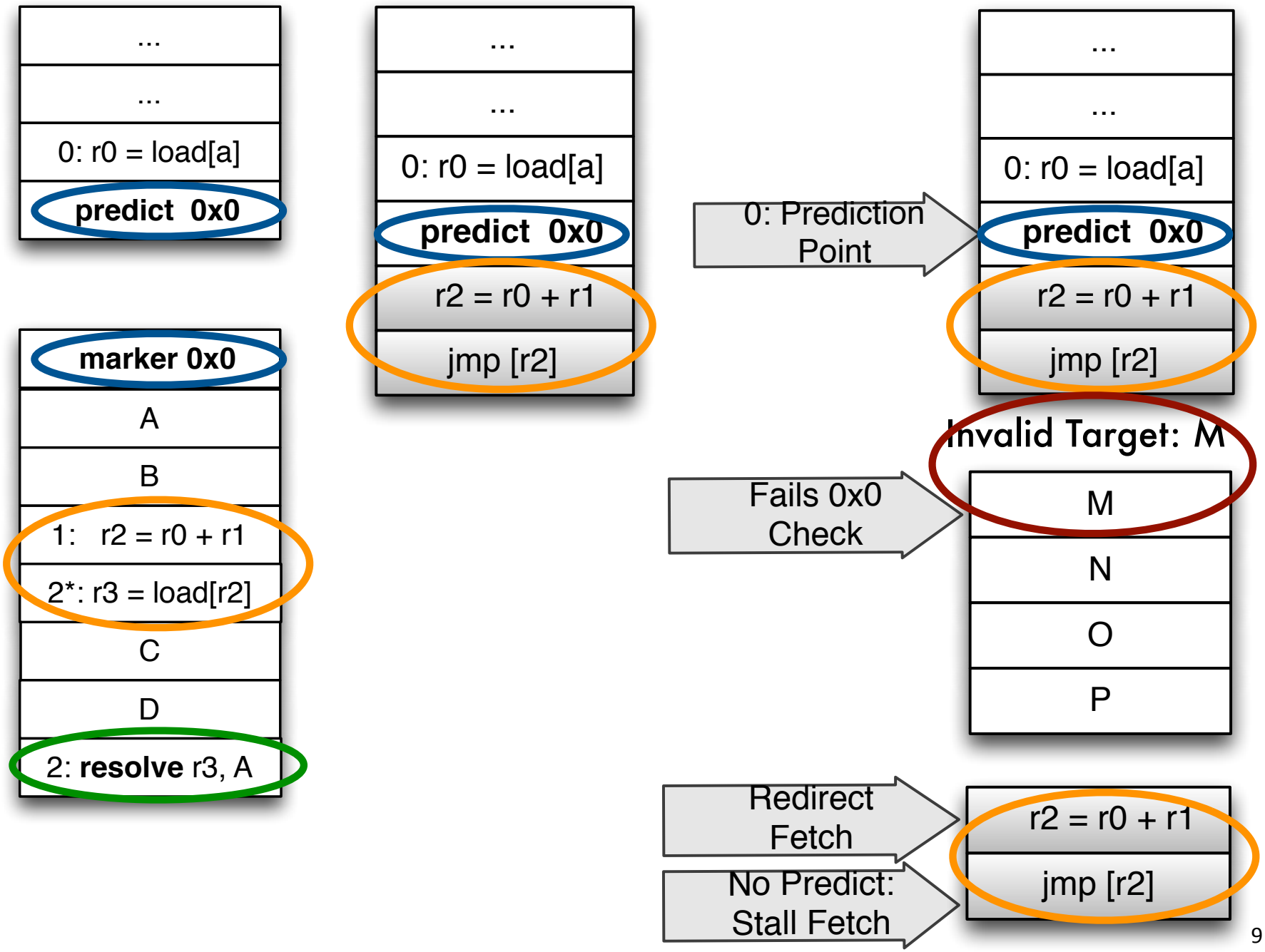
# Solution: Landing Pad



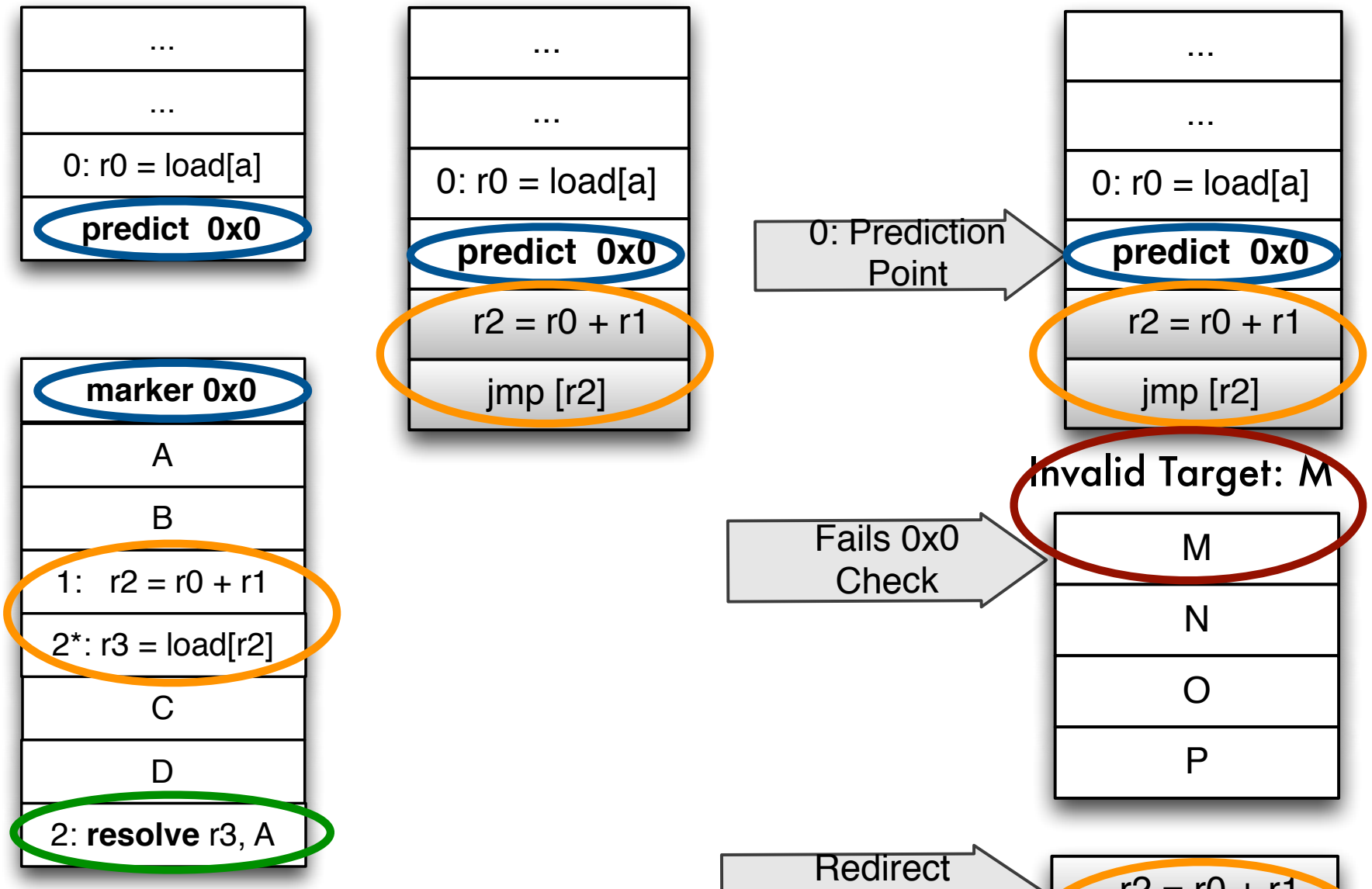
# Solution: Landing Pad



# Solution: Landing Pad



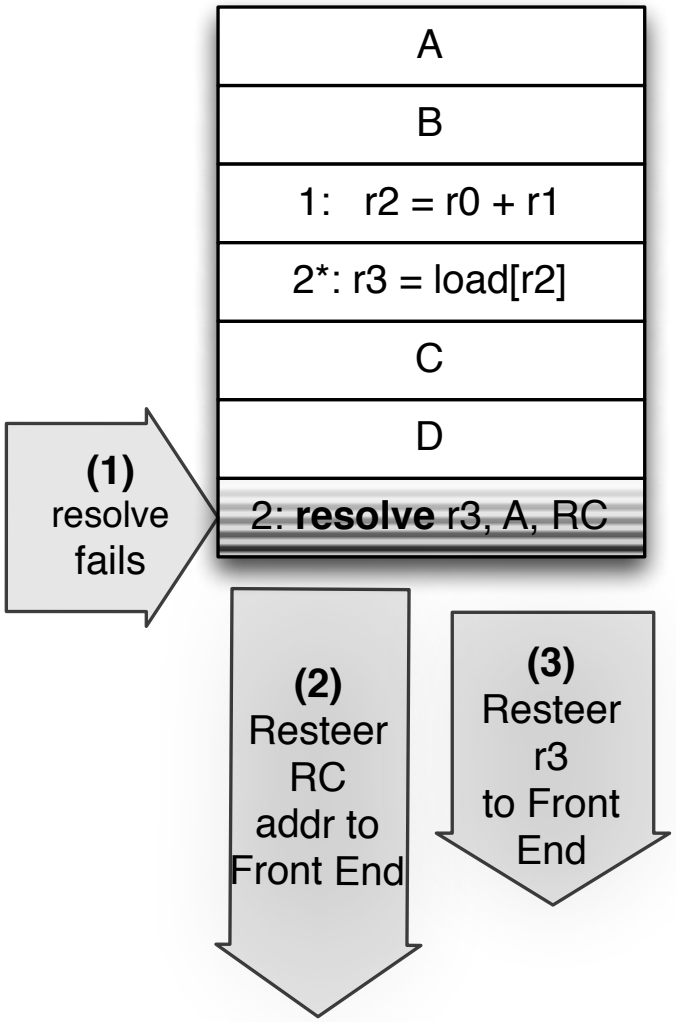
# Solution: Landing Pad



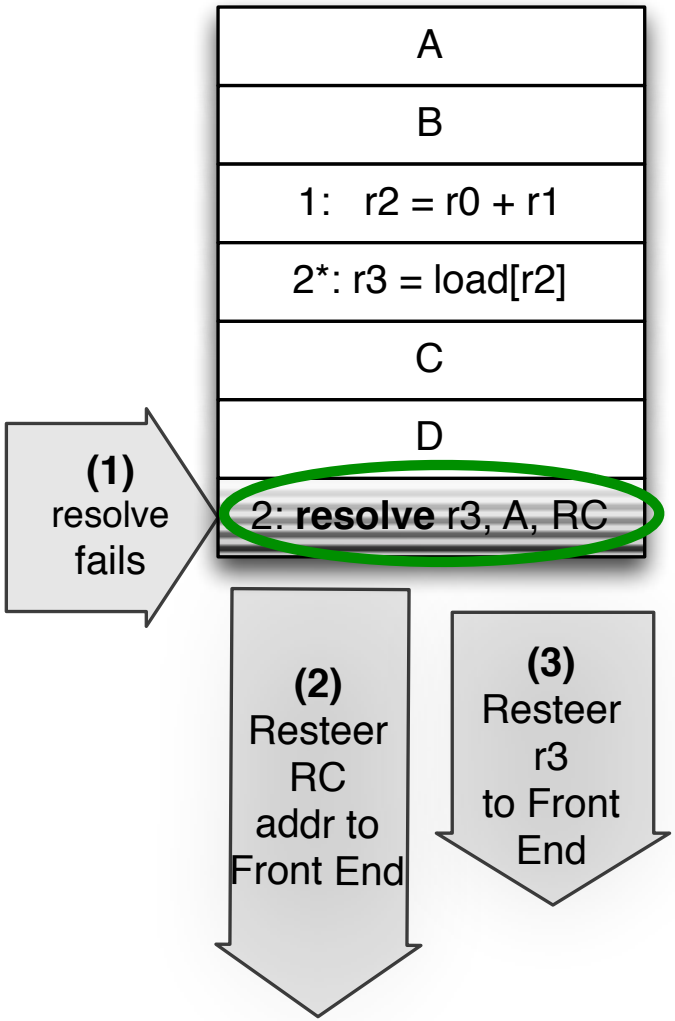

**Necessitates some changes to indirect call /return handling**

# Recovery From Misprediction

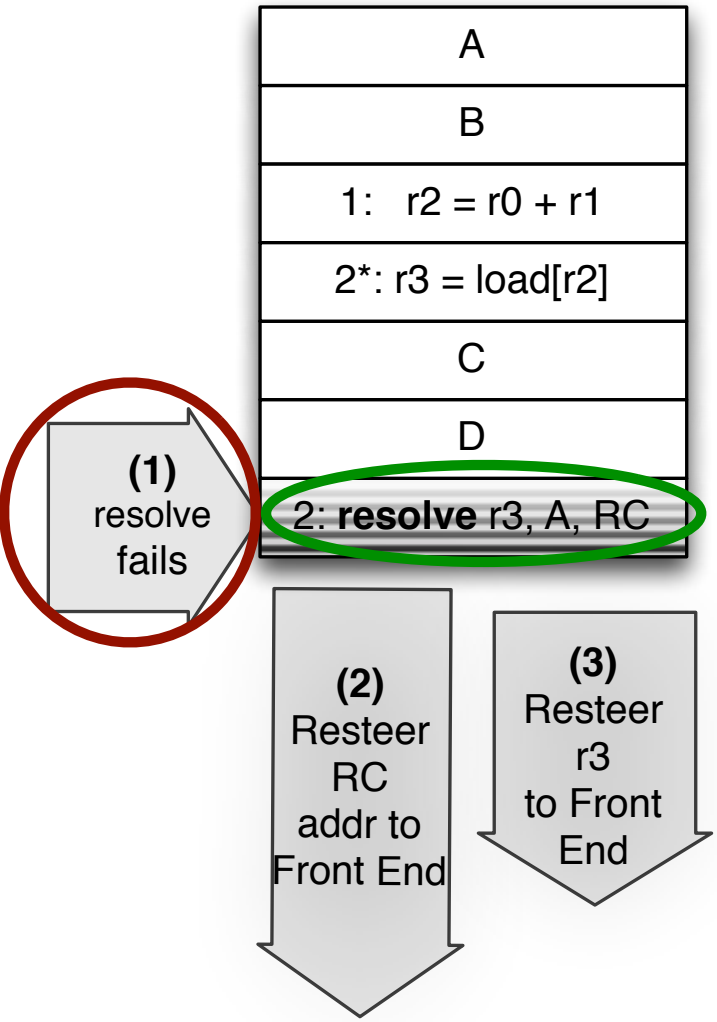
# Recovery From Misprediction



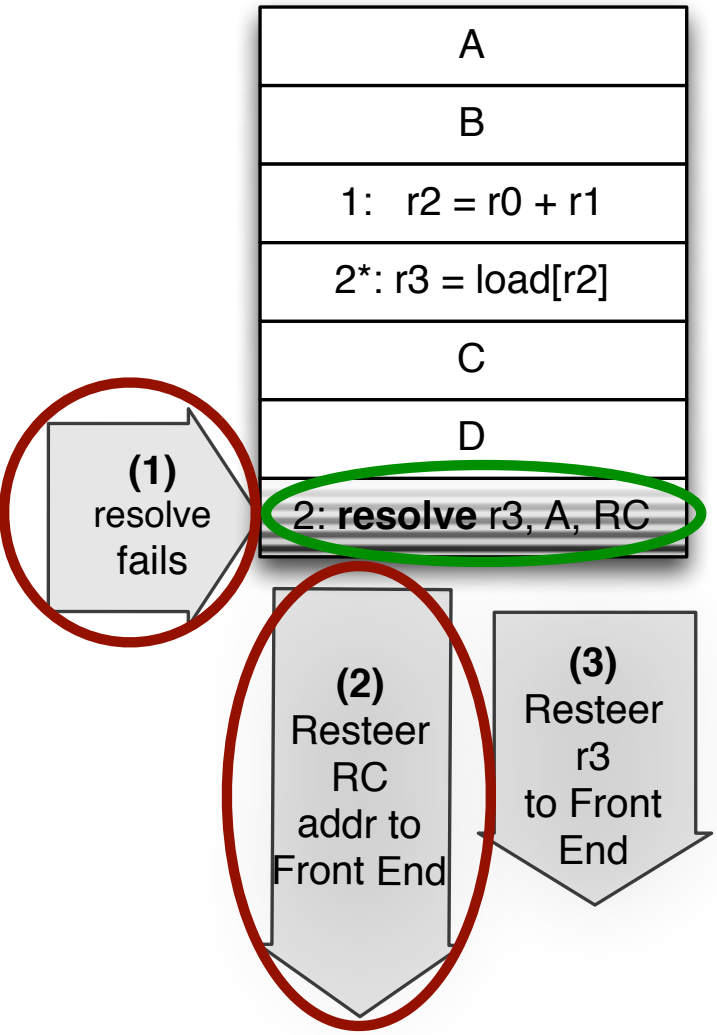
# Recovery From Misprediction



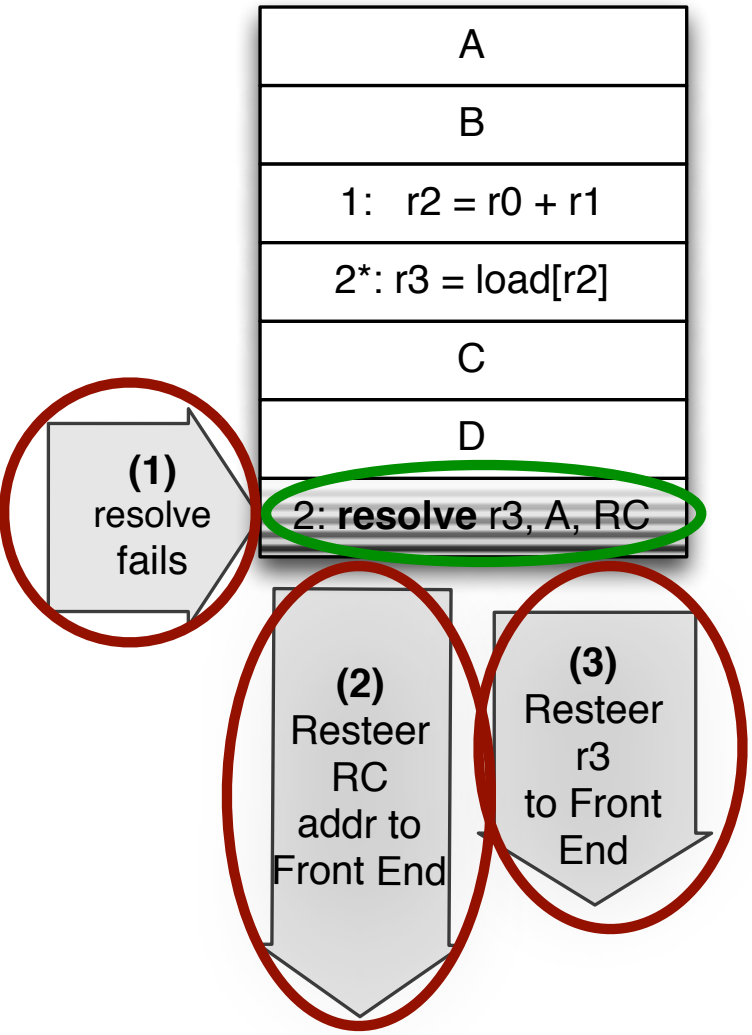
# Recovery From Misprediction



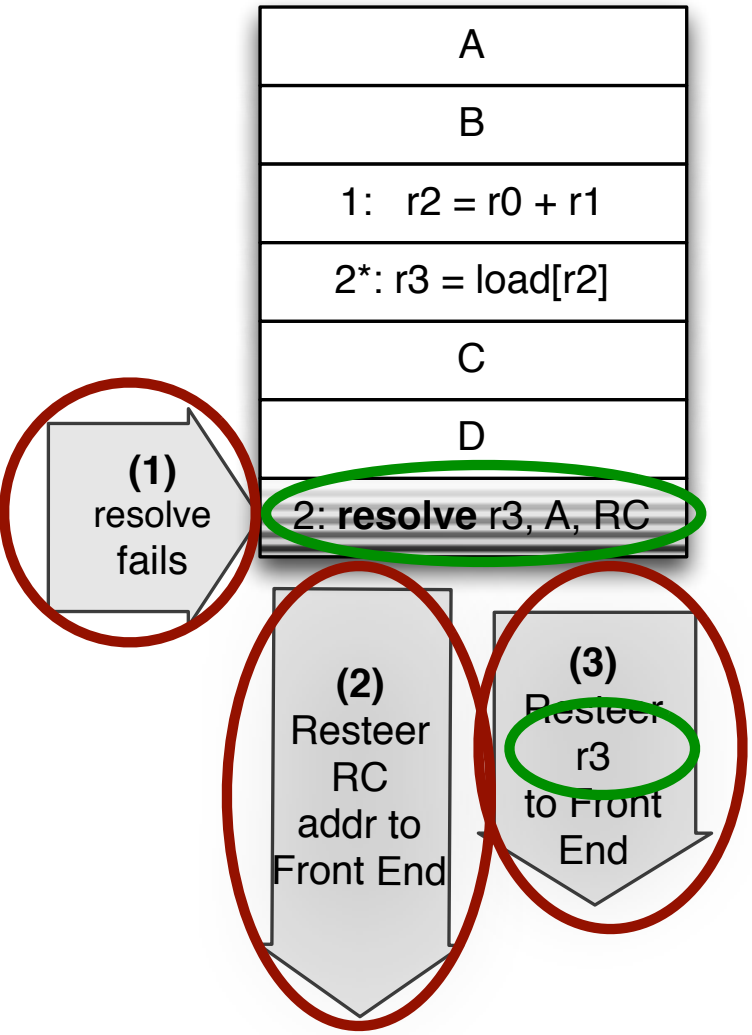
# Recovery From Misprediction



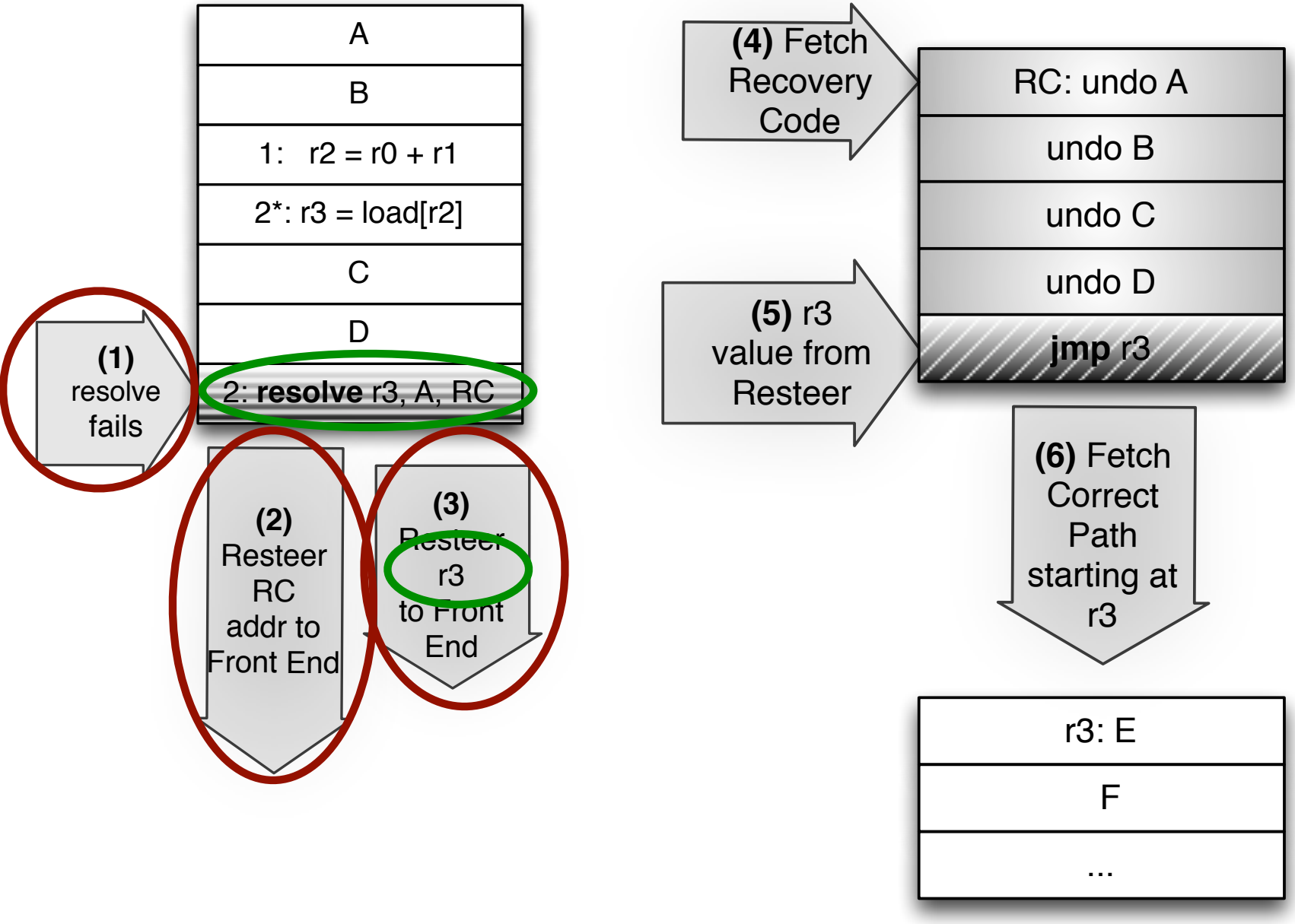
# Recovery From Misprediction



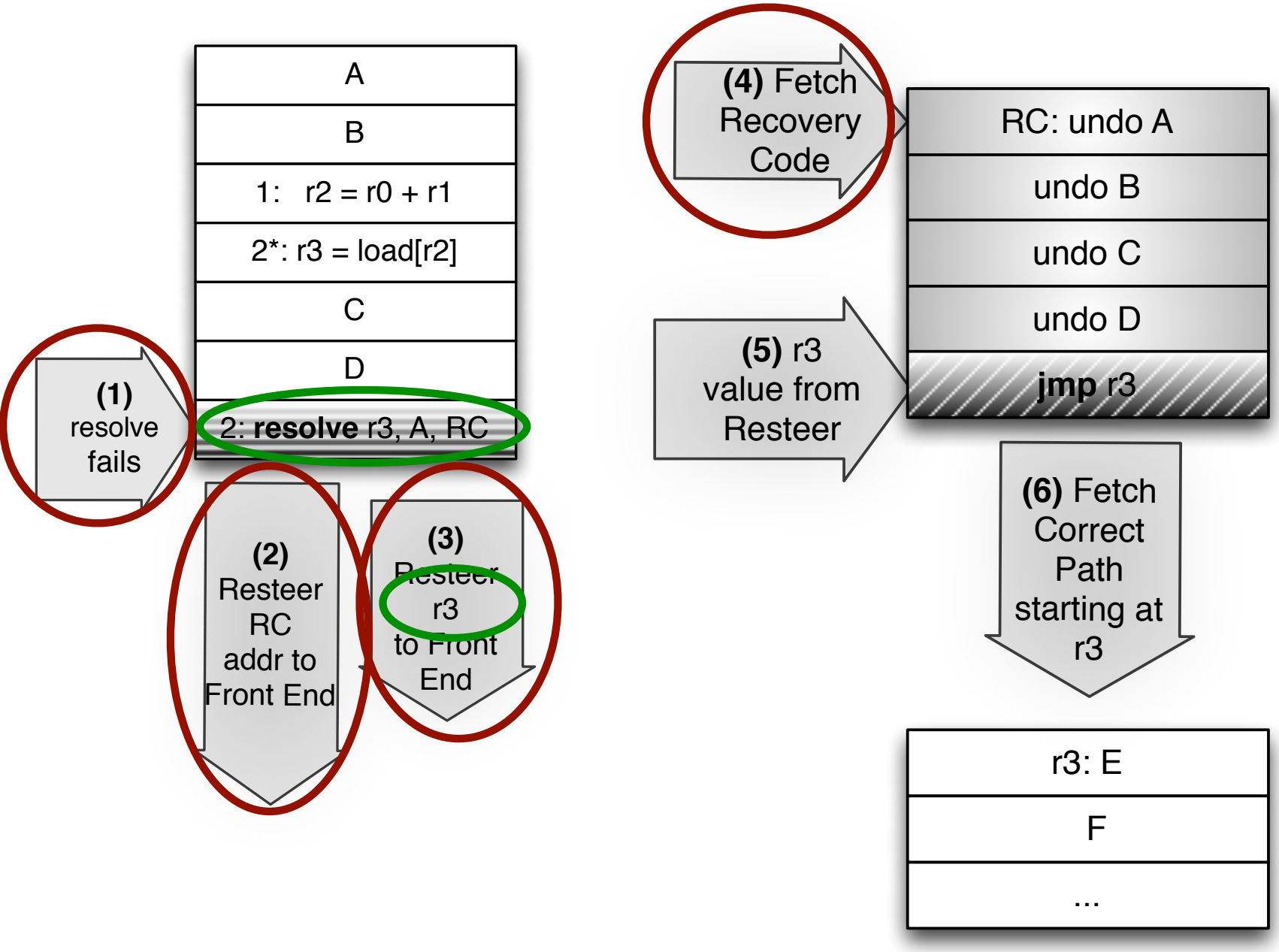
# Recovery From Misprediction



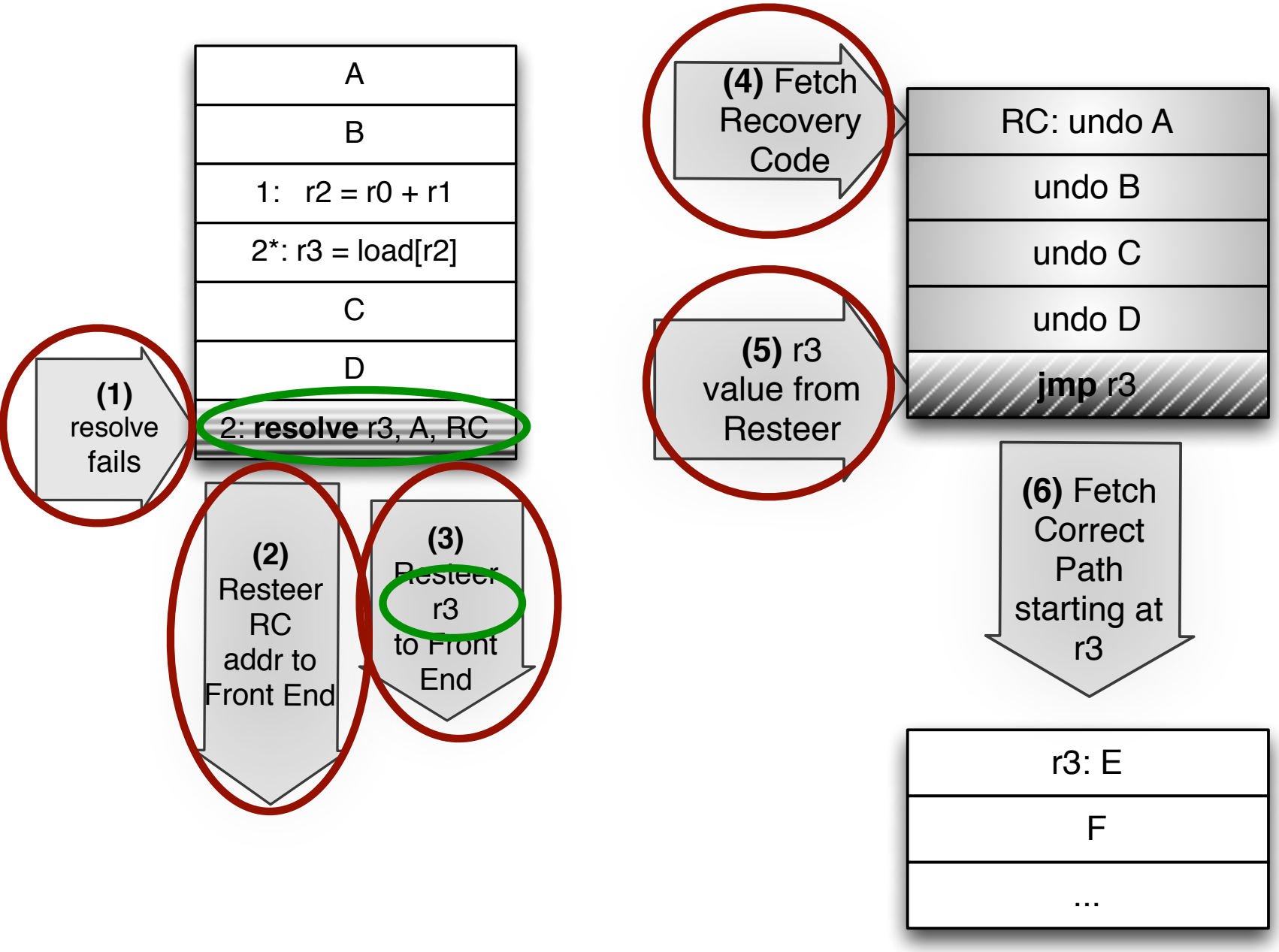
# Recovery From Misprediction



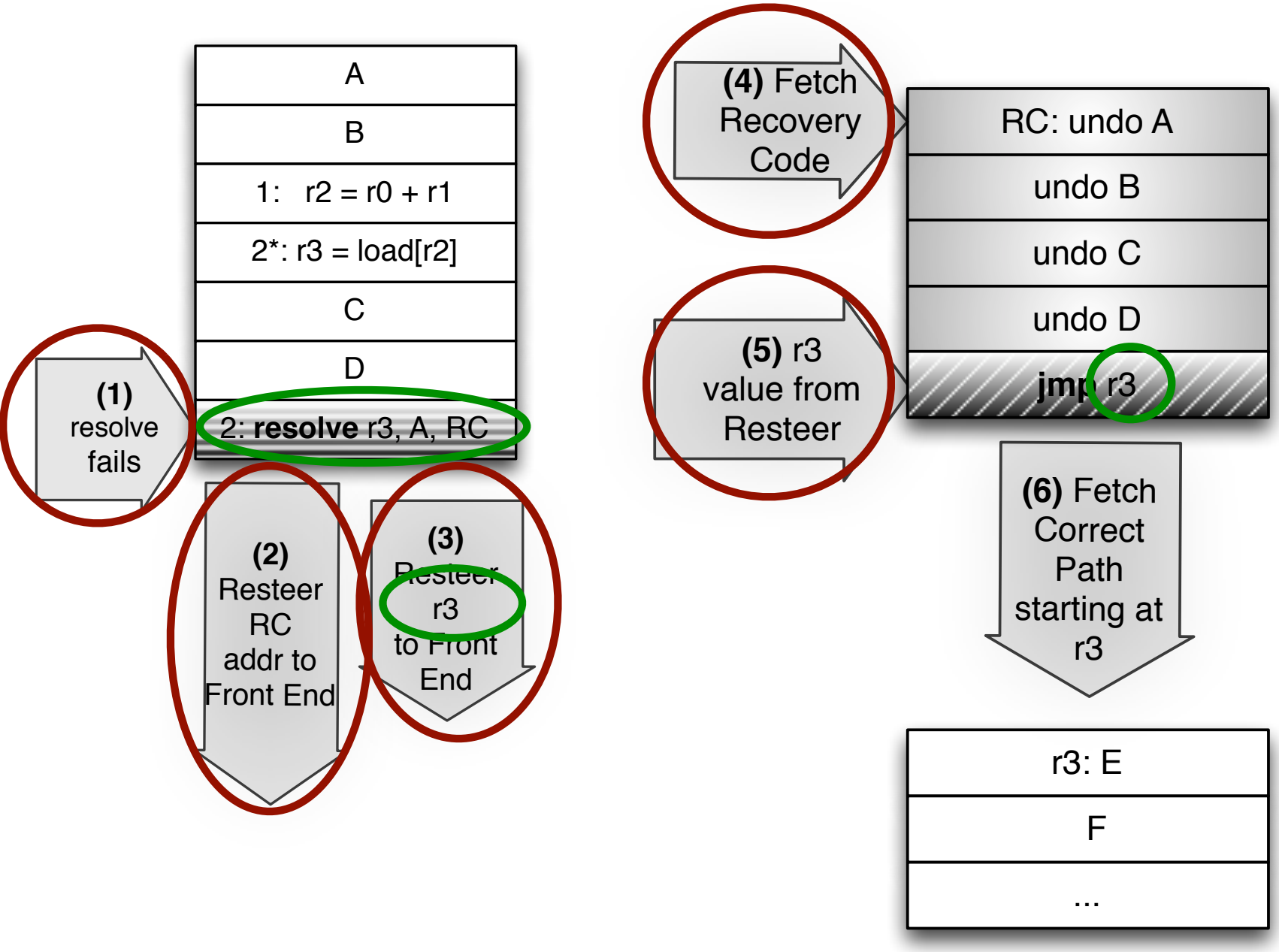
# Recovery From Misprediction



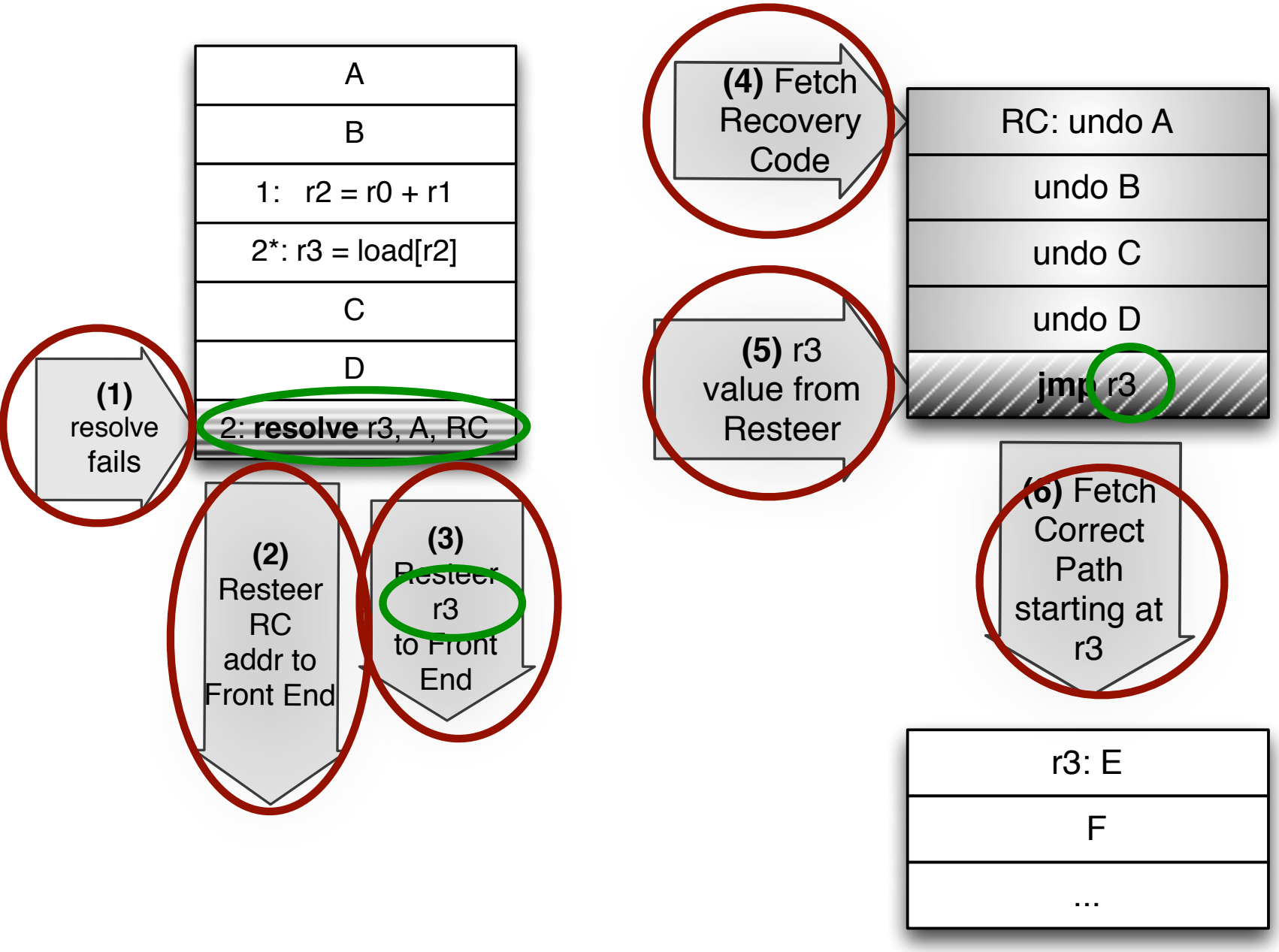
# Recovery From Misprediction



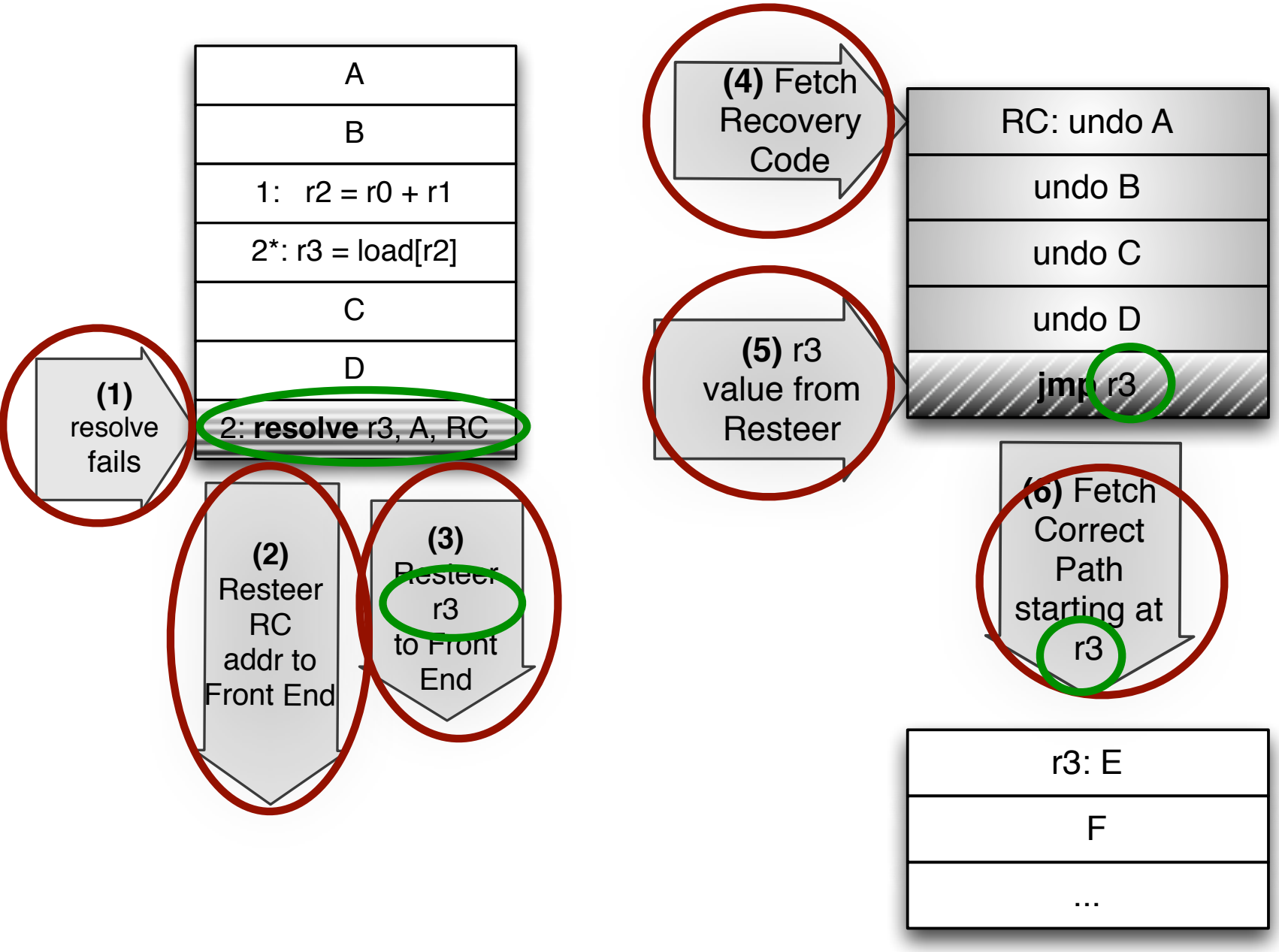
# Recovery From Misprediction



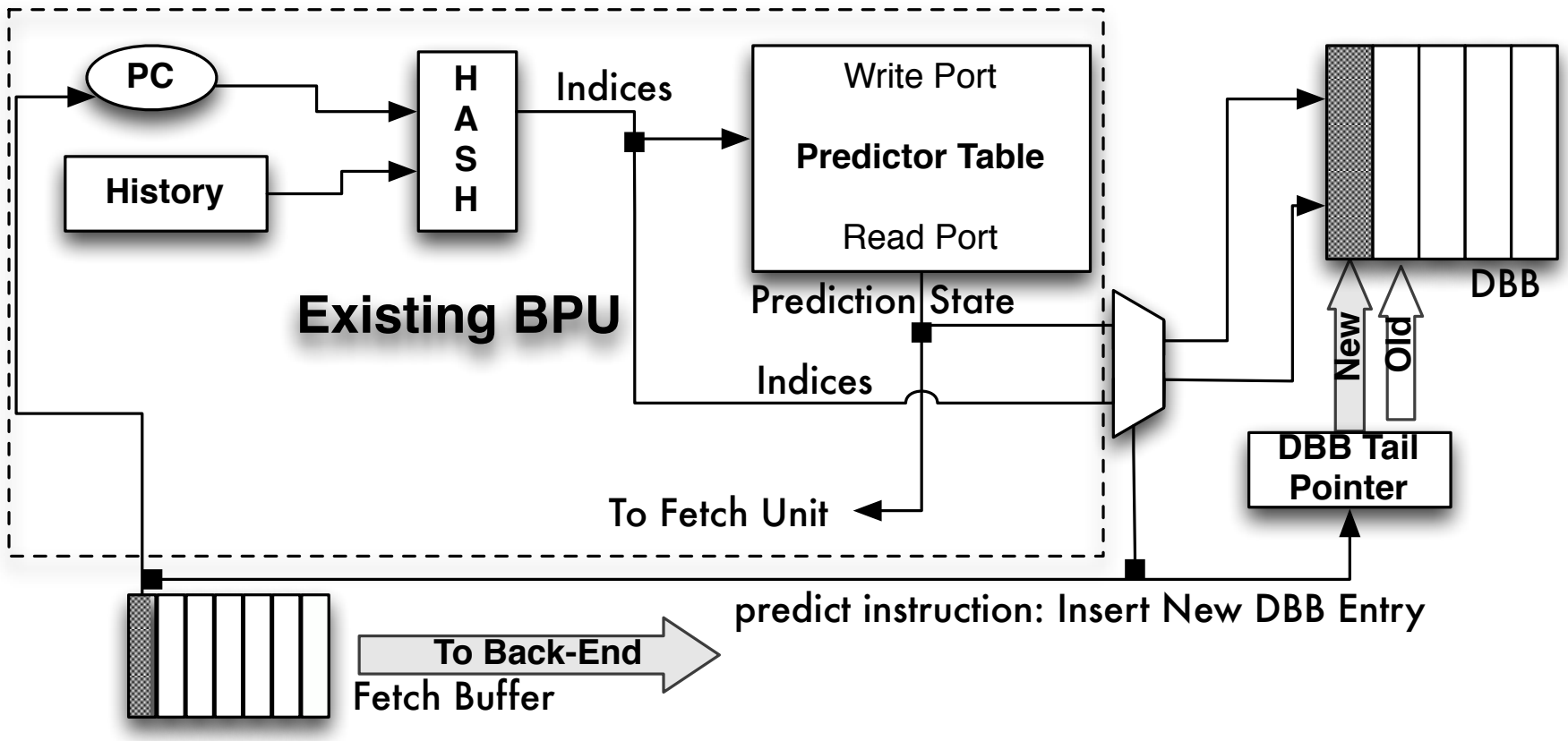
# Recovery From Misprediction



# Recovery From Misprediction



# Hardware Requirements

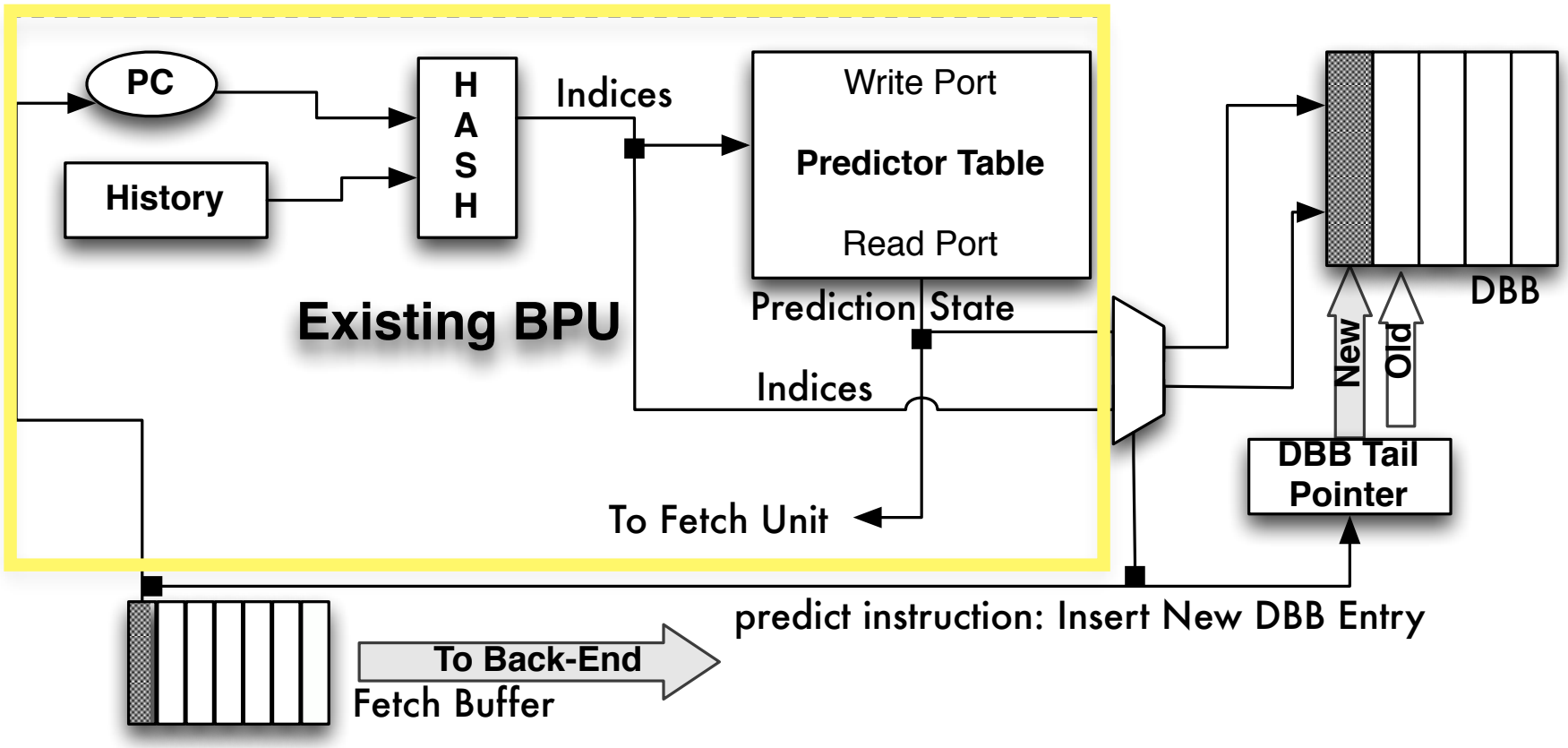



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

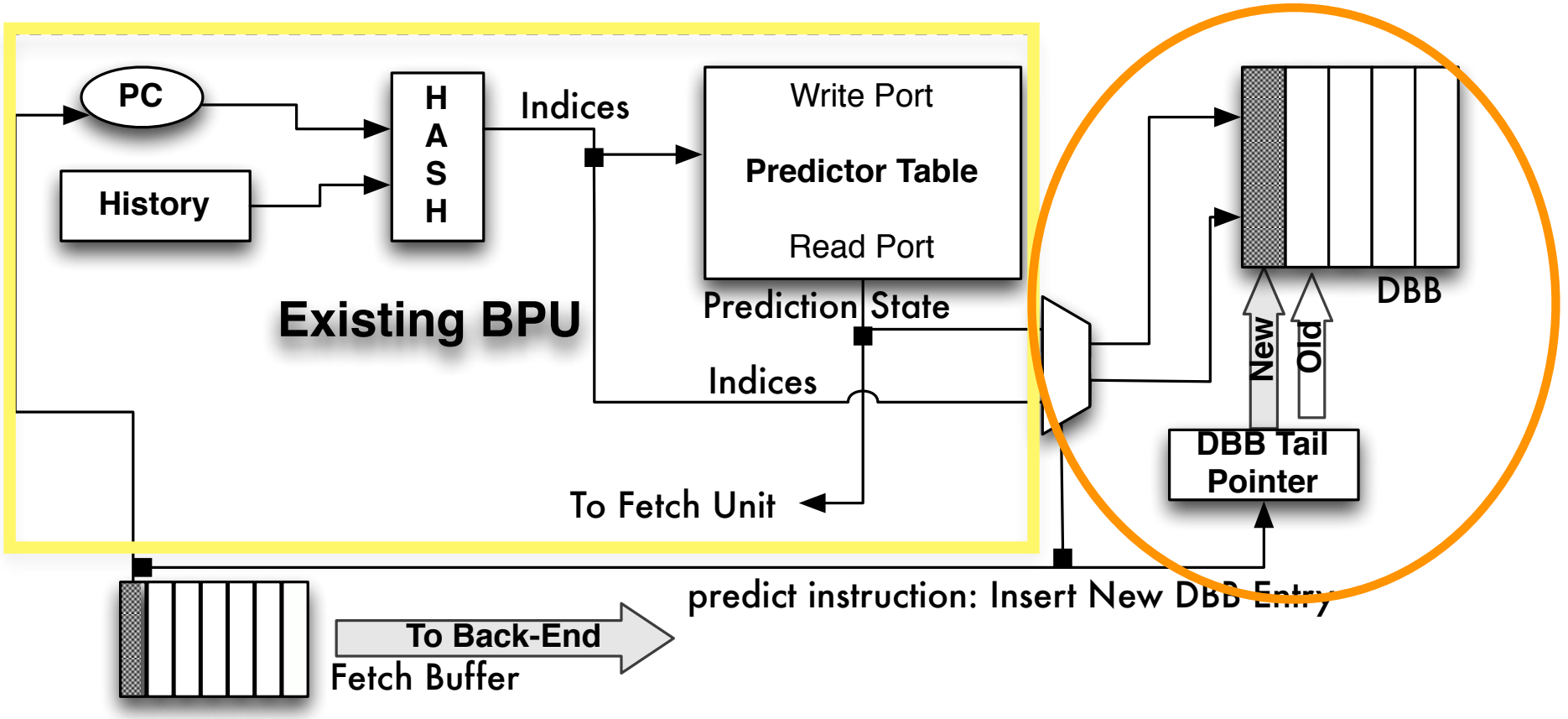



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

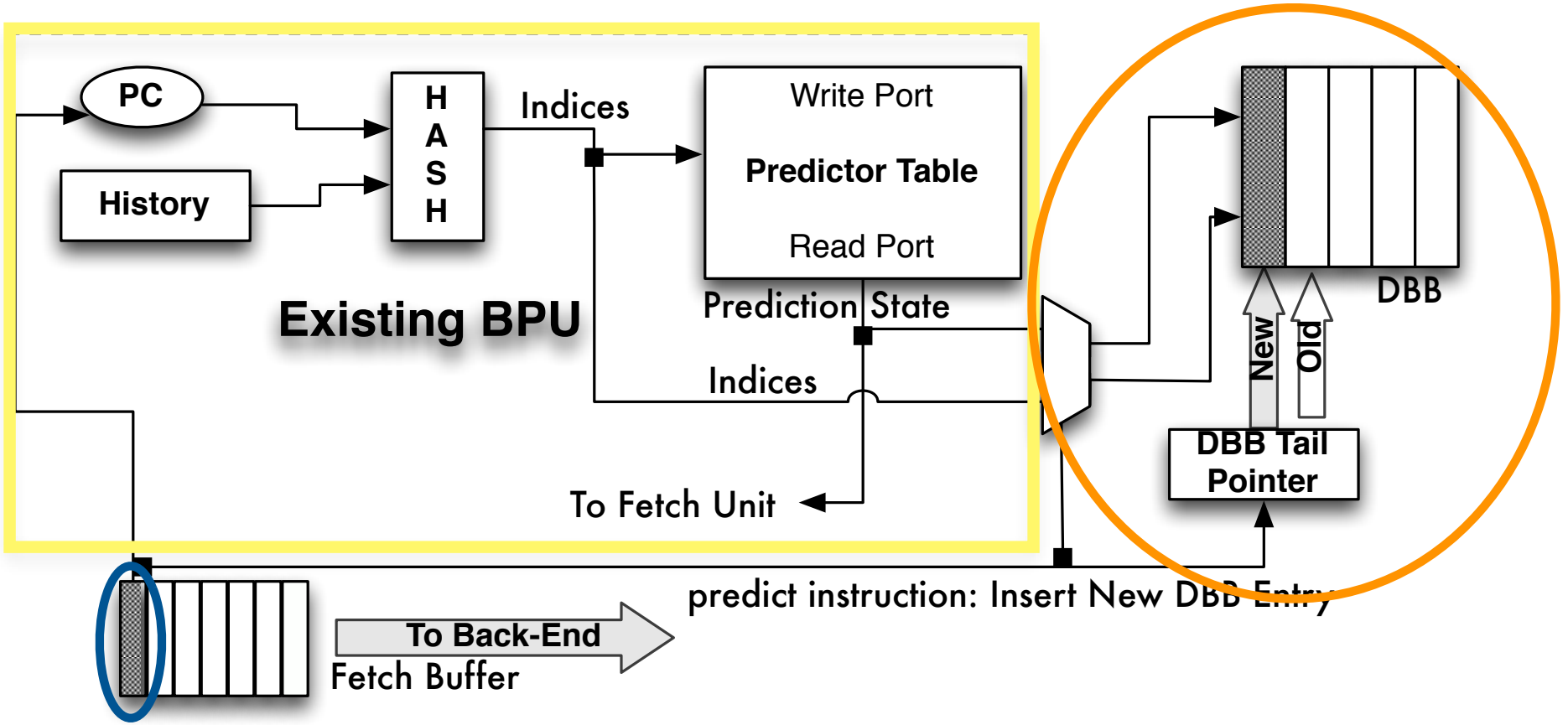



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

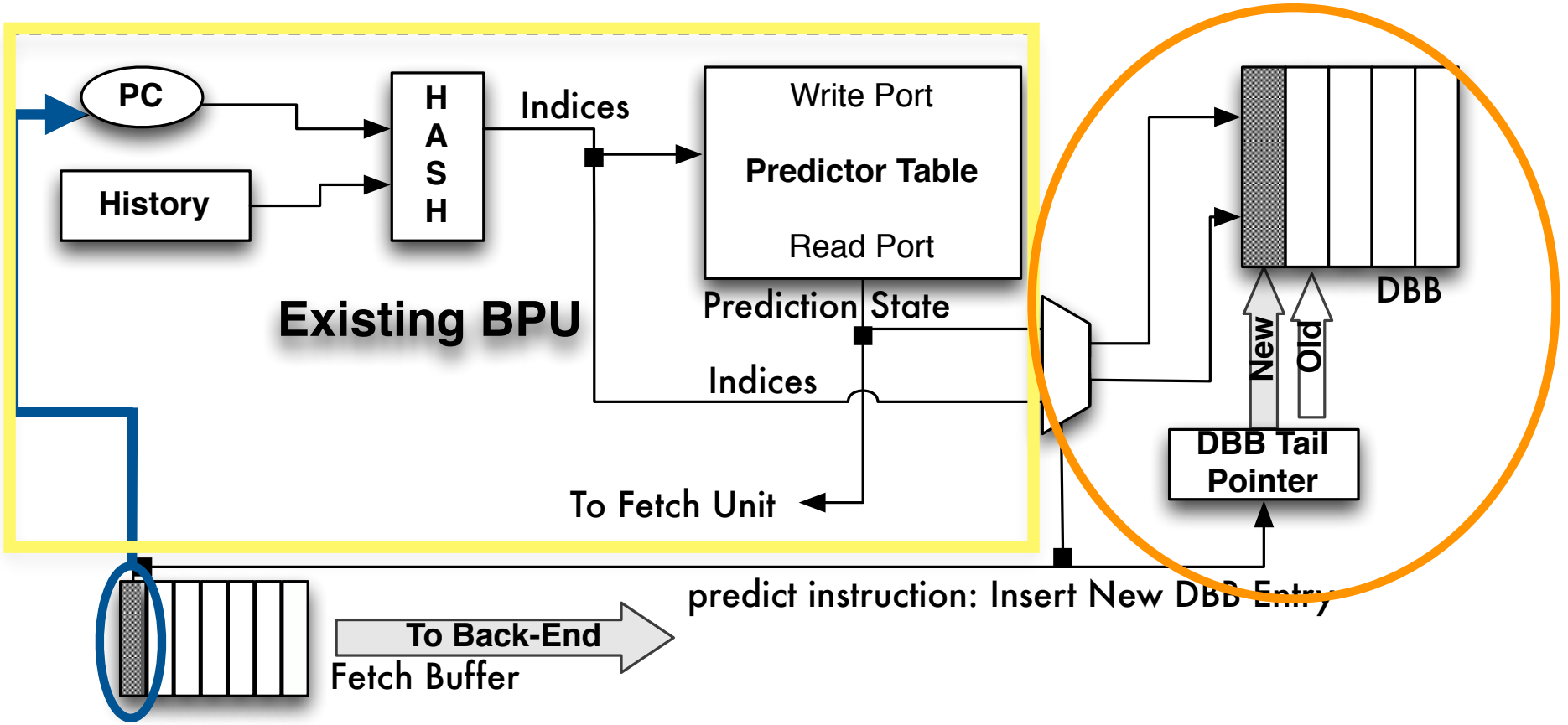



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

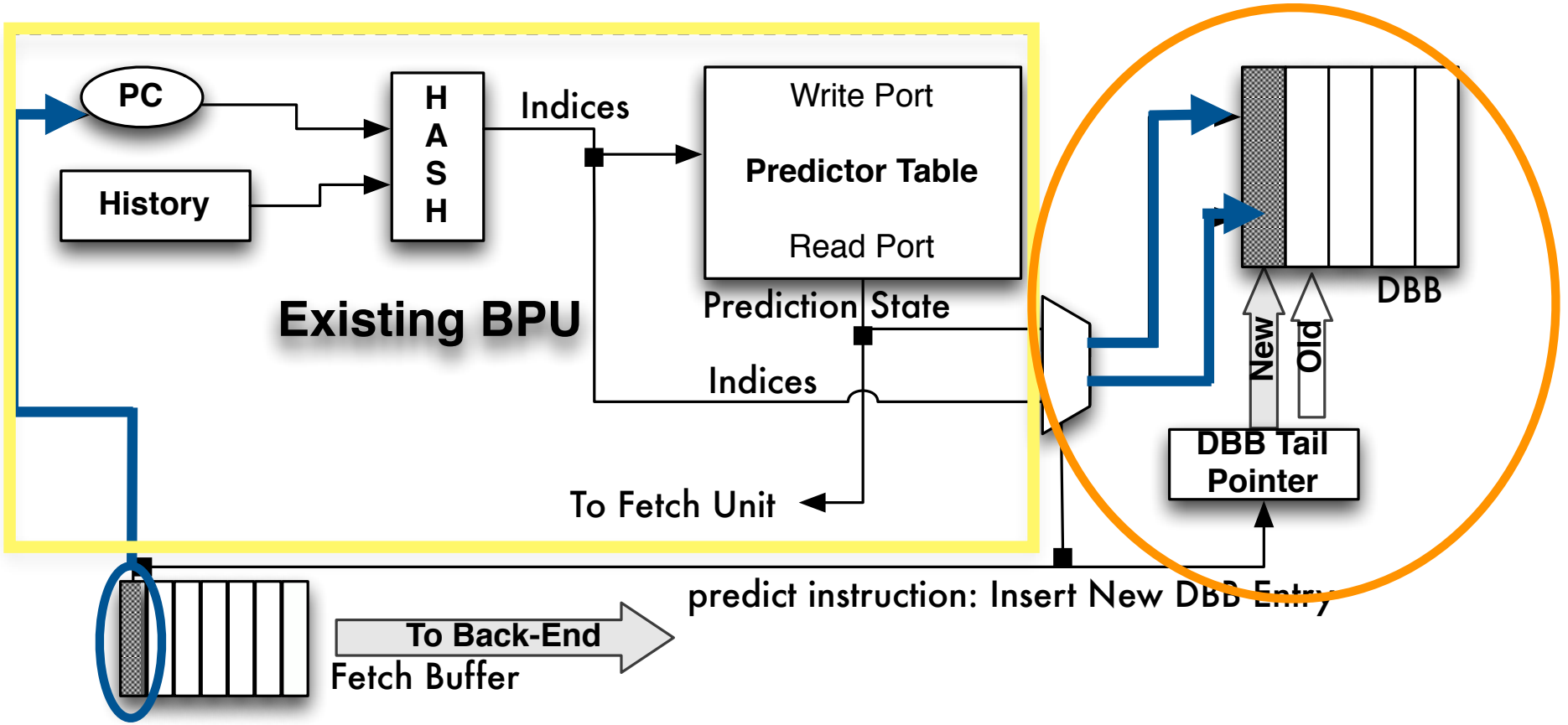



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

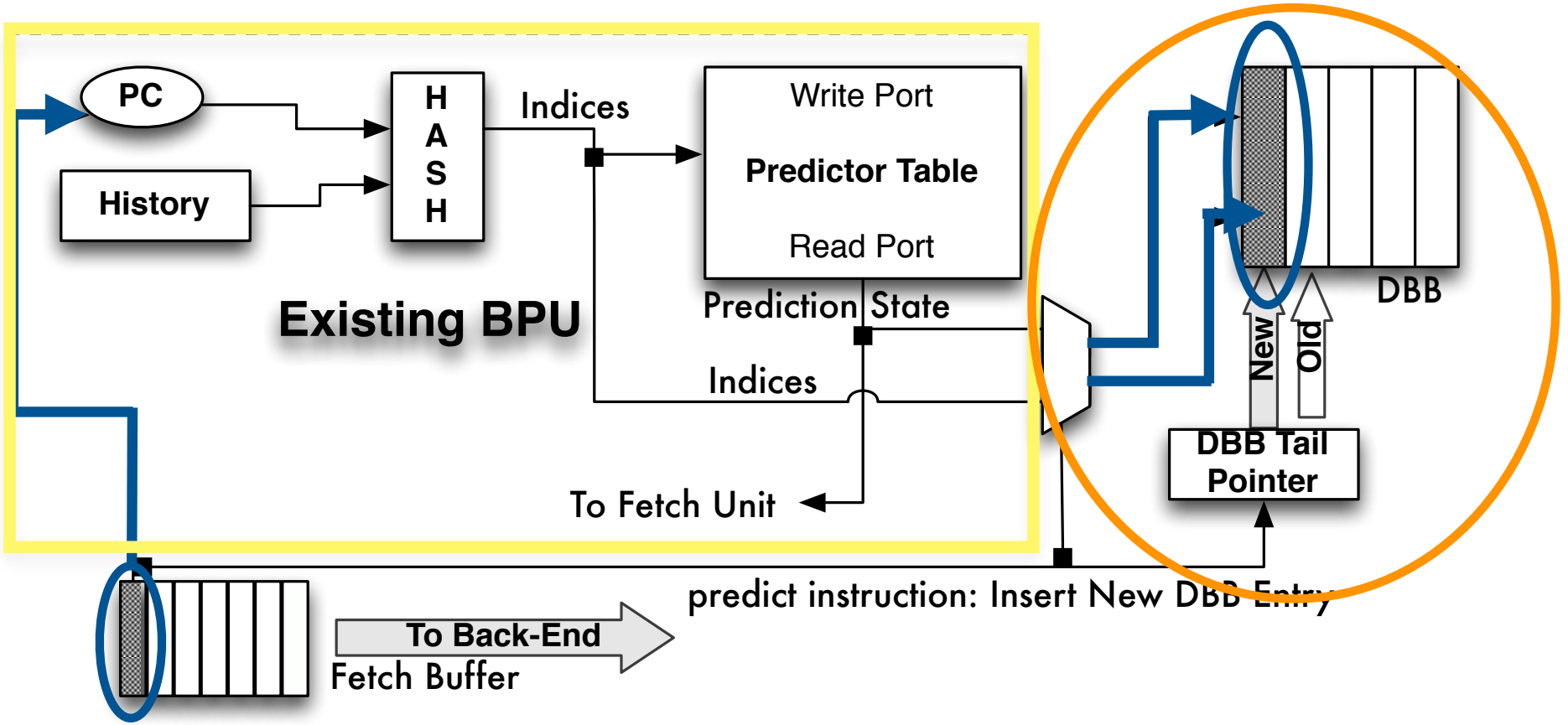



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

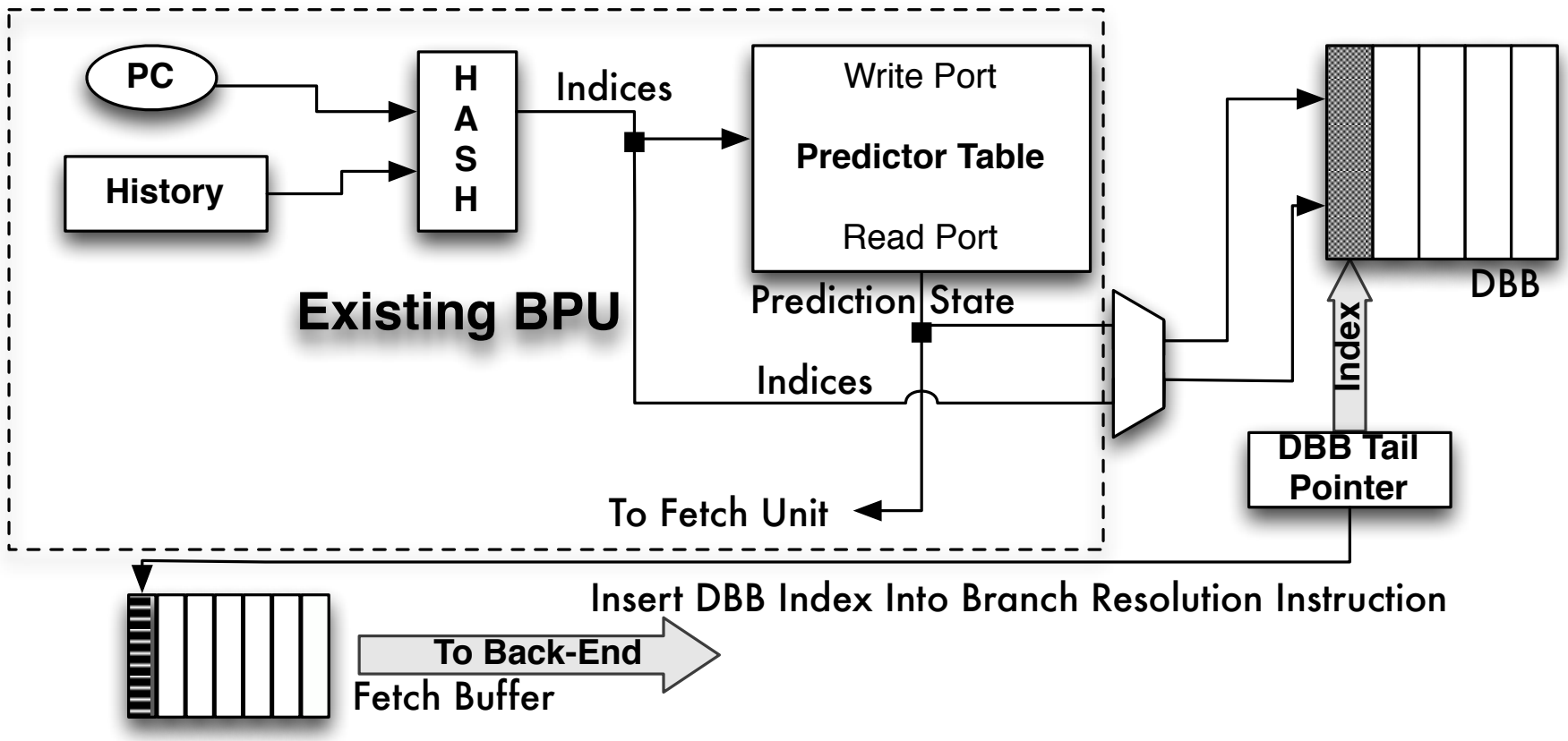



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

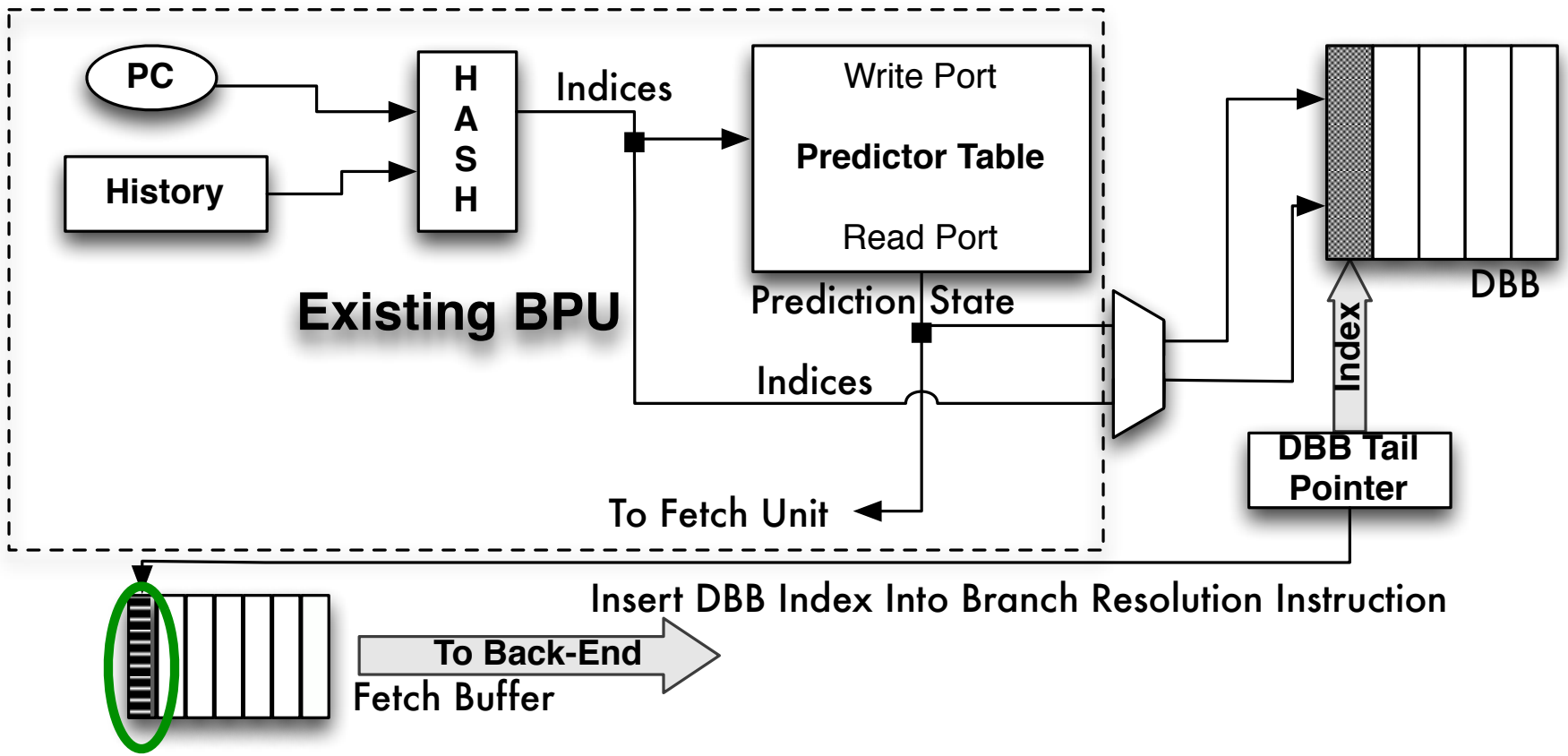
- Single pointer, single R/W port

# Hardware Requirements



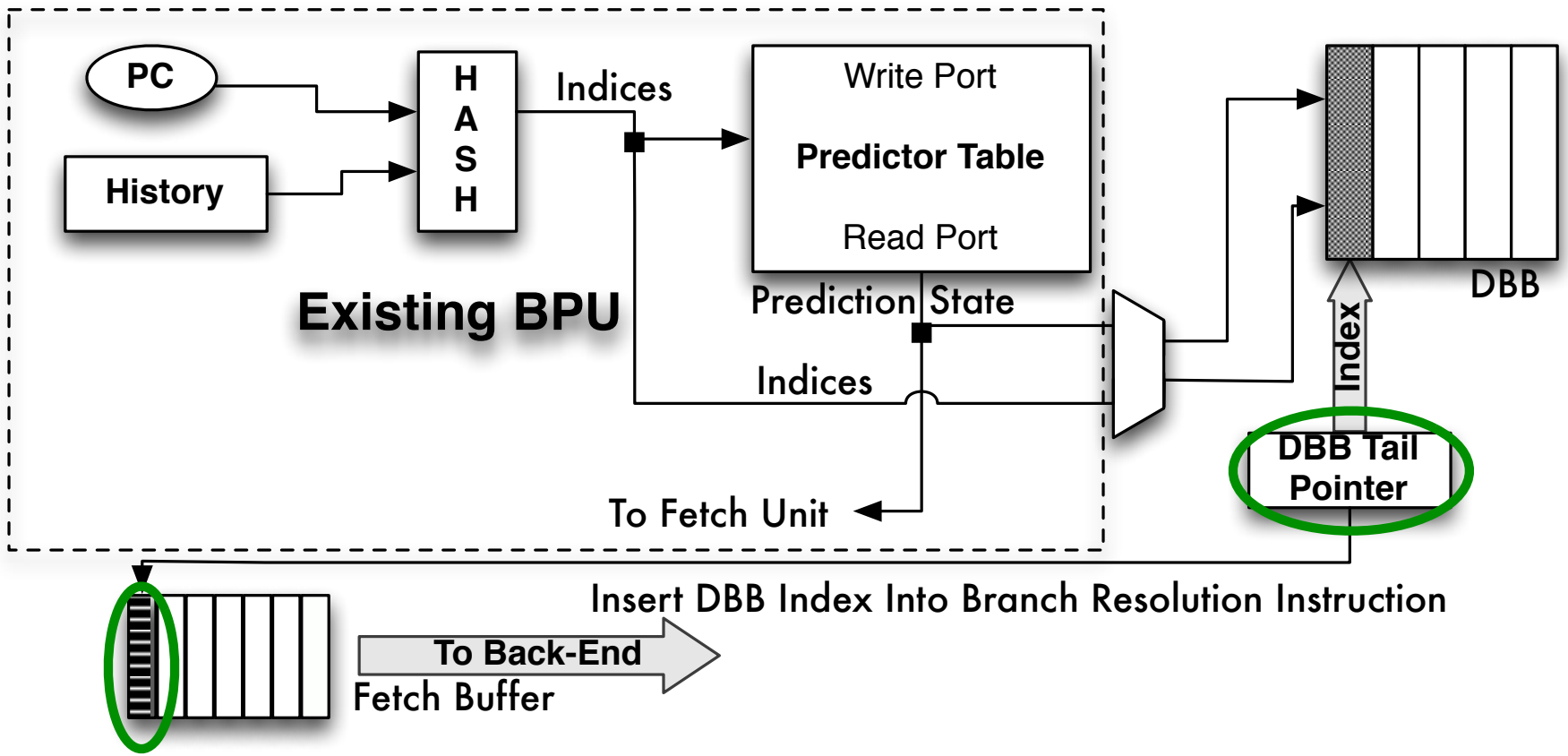

- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions
  - Dovetails with existing structures for outstanding branches
- Single pointer, single R/W port

# Hardware Requirements



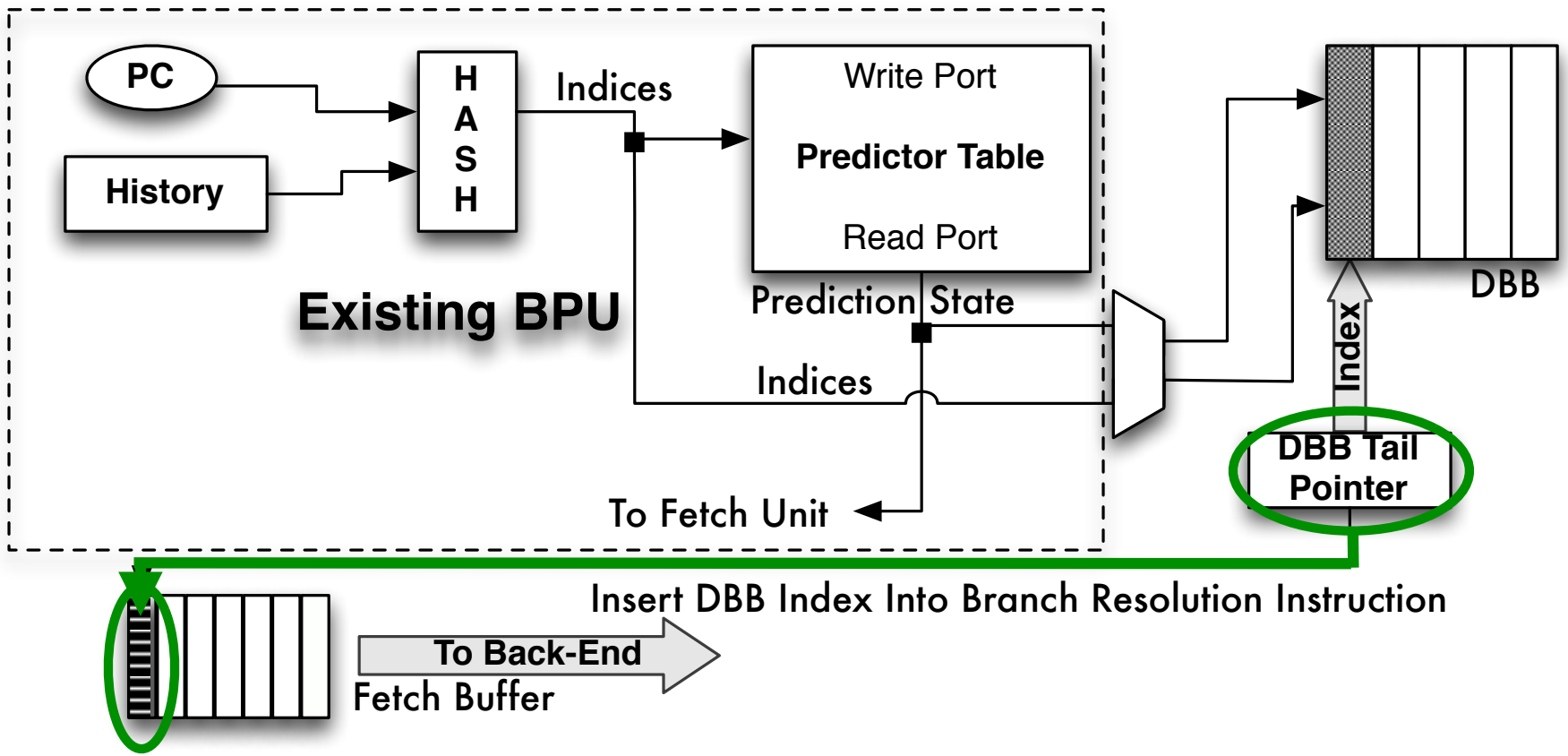

- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions
  - Dovetails with existing structures for outstanding branches
- Single pointer, single R/W port

# Hardware Requirements



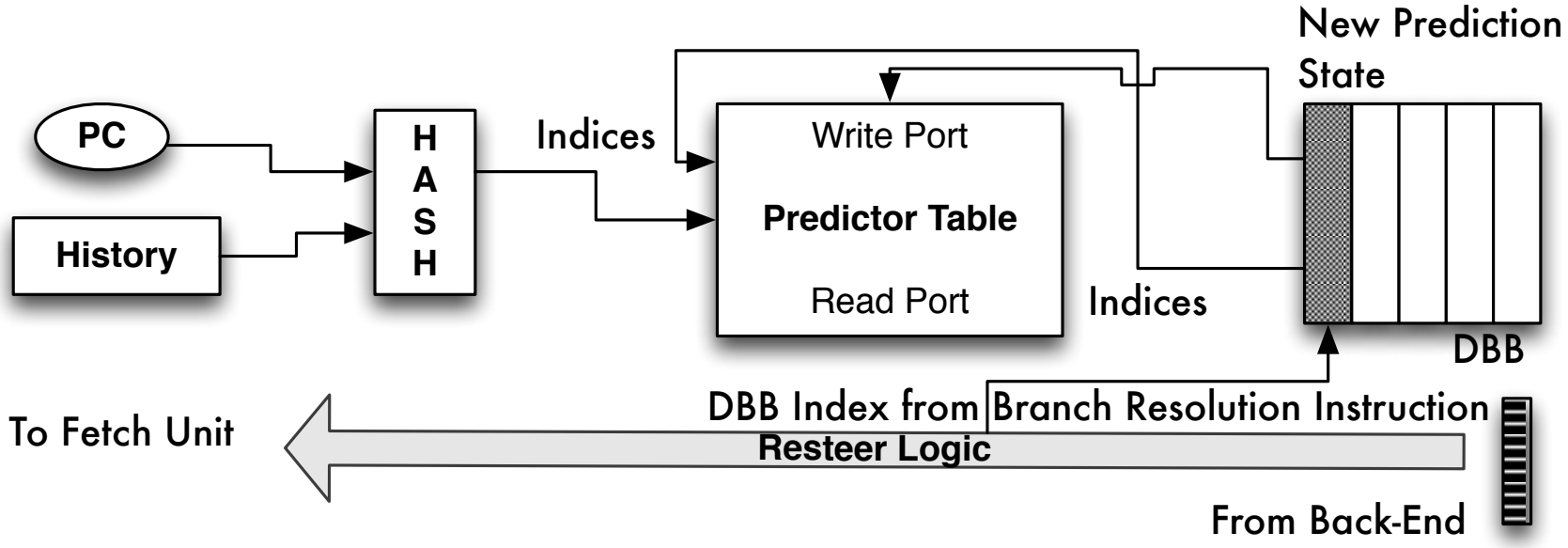

- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions
  - Dovetails with existing structures for outstanding branches
- Single pointer, single R/W port

# Hardware Requirements




- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions
  - Dovetails with existing structures for outstanding branches
- Single pointer, single R/W port

# Hardware Requirements

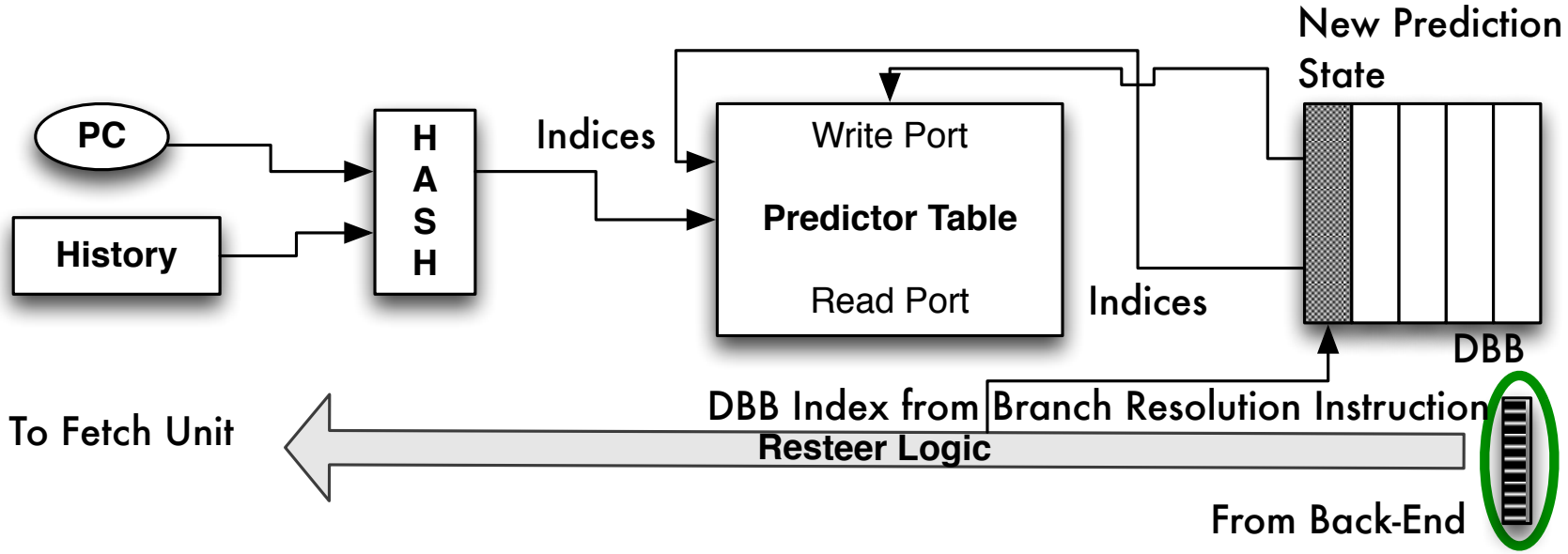



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

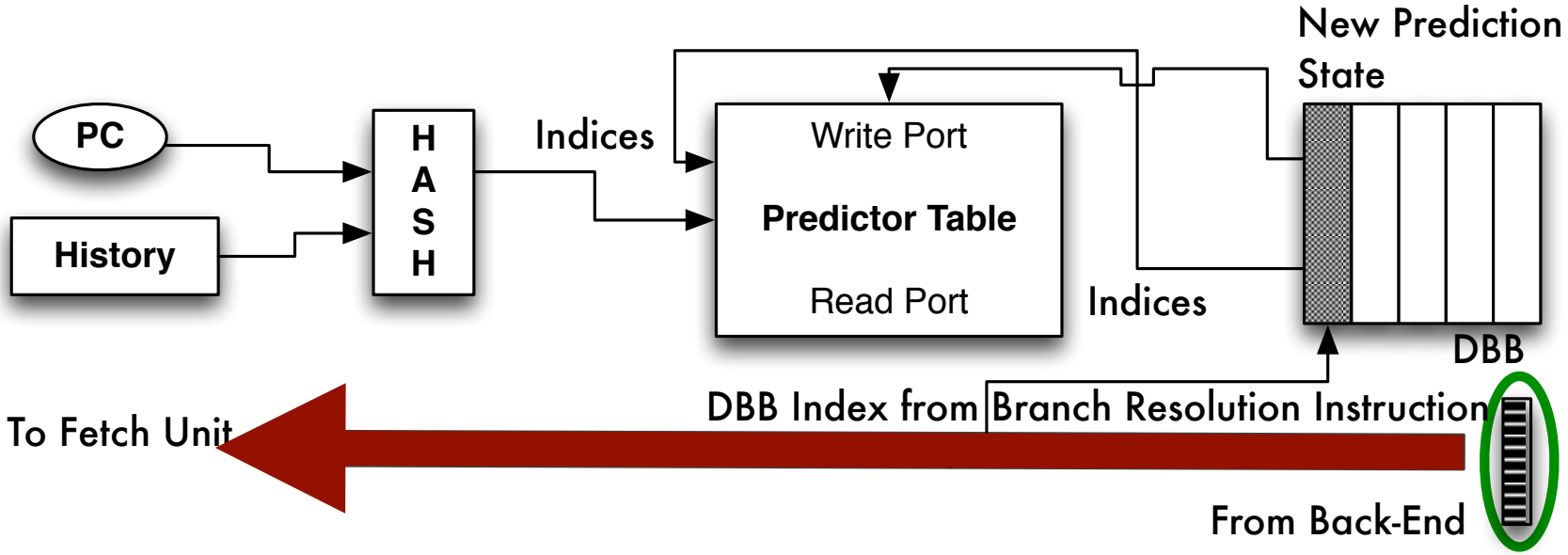



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

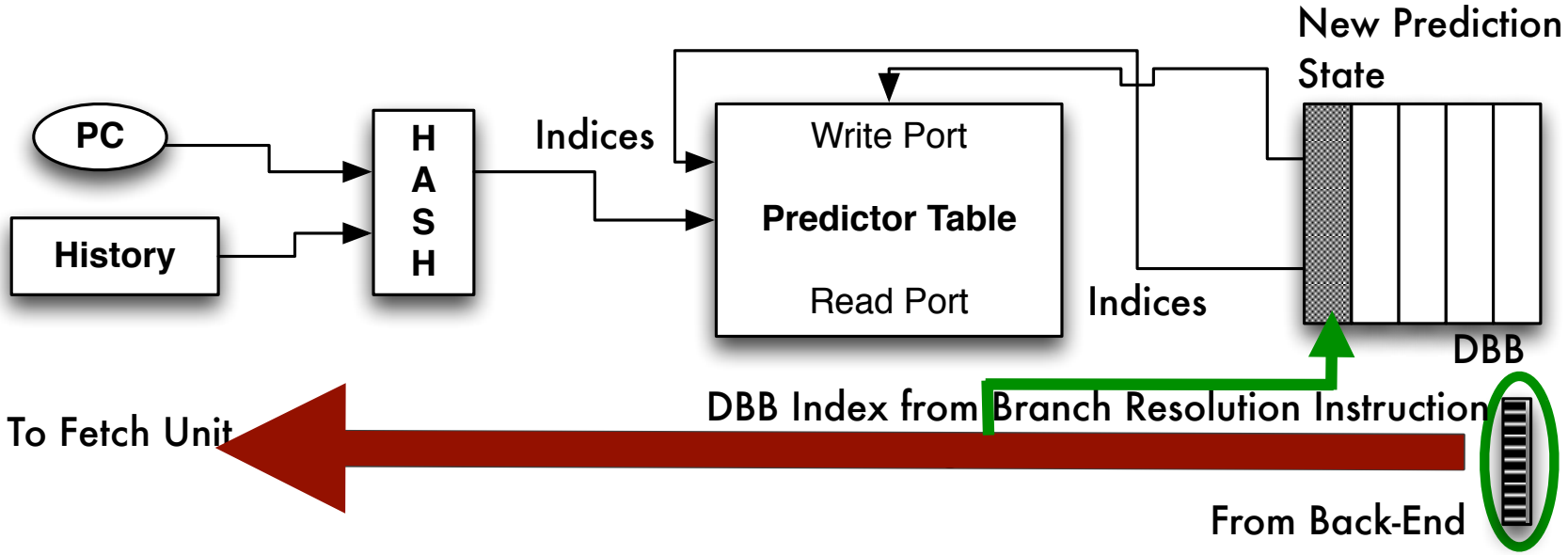



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

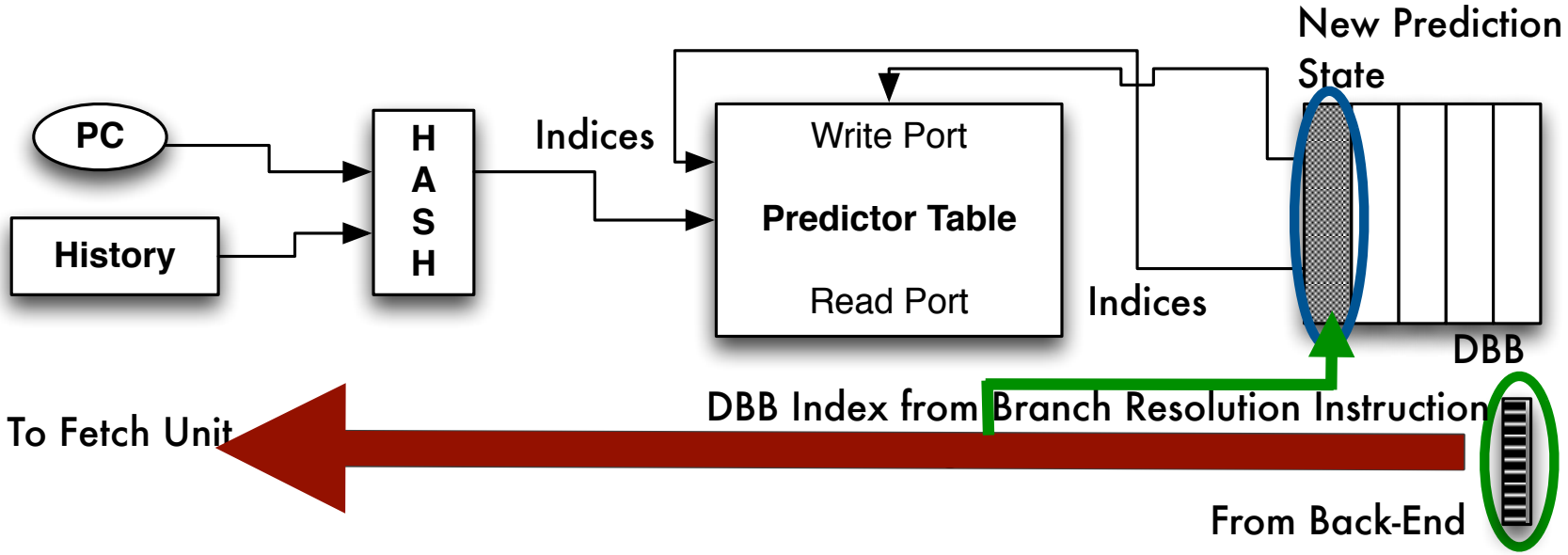



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements

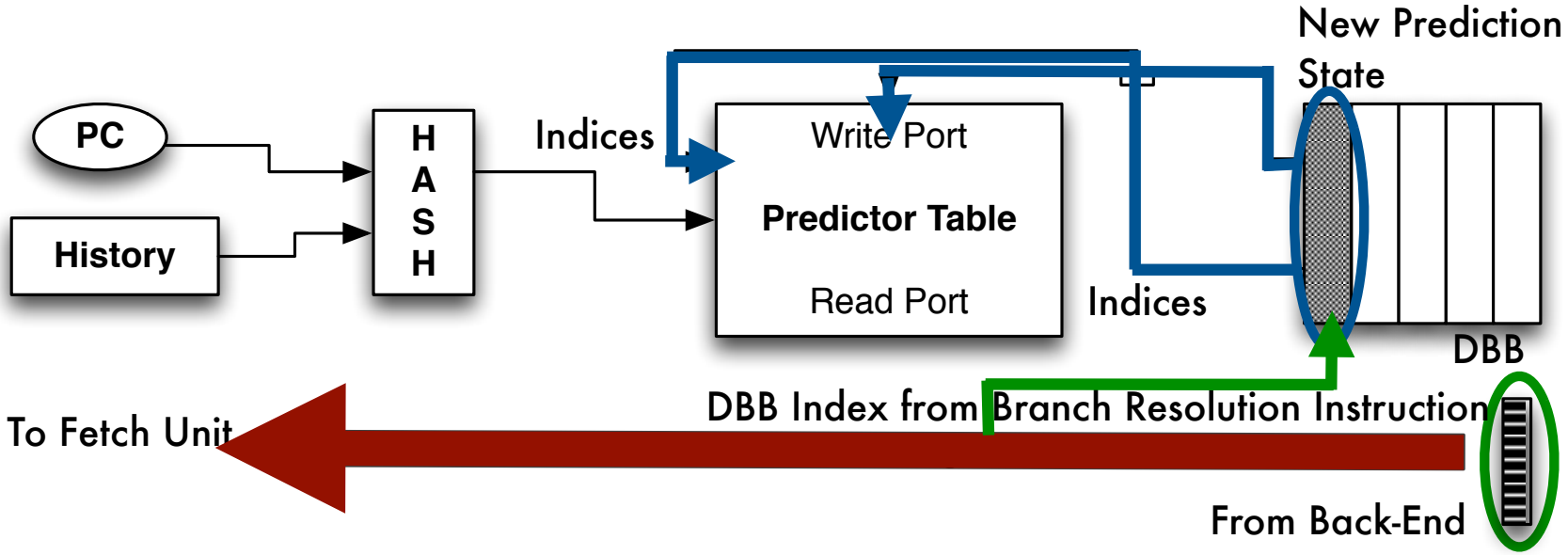



- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

- Single pointer, single R/W port

# Hardware Requirements




- Maintain correspondence between prediction/resolution
- Small FIFO: not many branches outstanding and no reordering of prediction and resolution instructions

–Dovetails with existing structures for outstanding branches

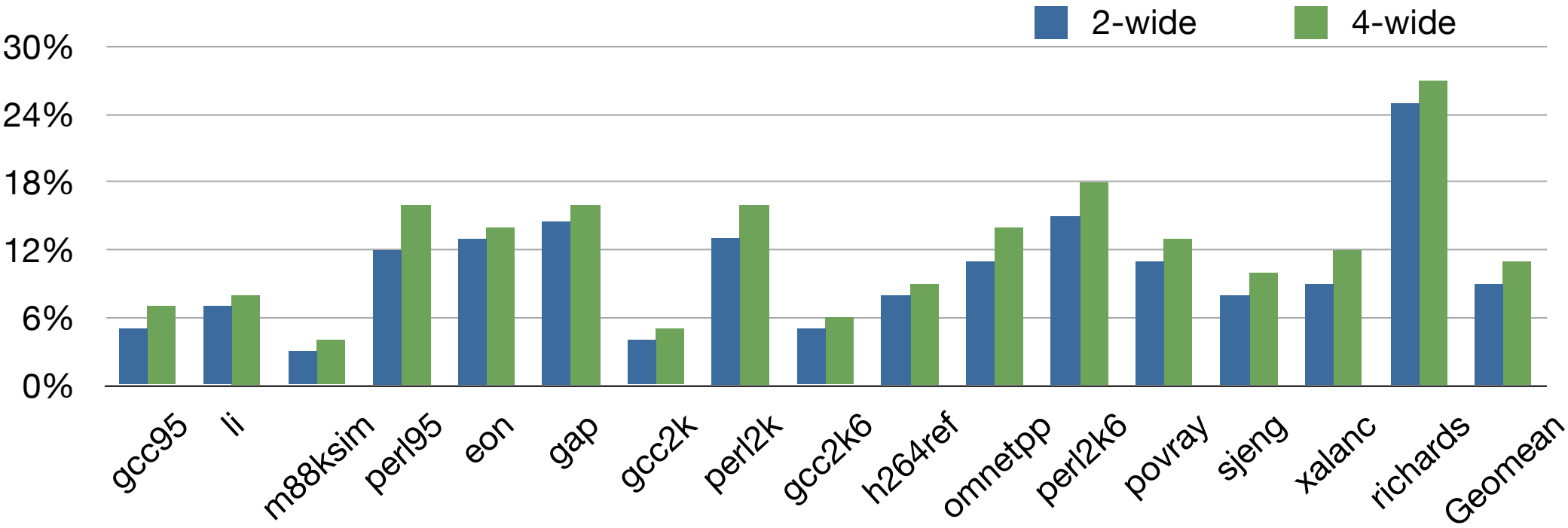
- Single pointer, single R/W port

# Experimental Methodology

| Key Structures | Configuration Parameters |
|---|---|
| **Bpred** | PTLSim default: GShare, 24 KB 3-table direction predictor, 4K-entry BTB, 64-entry RAS |
| **Front-End** | 5 stages, **Experimentally Varied 2/4 wide Fetch/Decode/Dispatch**, 32-entry FetchBuffer |
| **Execution Ports** | **Experimentally Varied 2/4** |
| **Functional Units** | Up to 2 x LD/ST, 2 x INT/SIMD-Permute, 4 x 64-bit SIMD/FP, 1-cycle bypass |
| **L1 Caches** | 8-way 32 KB L1-D$, 4-way 32 KB L1-I$, 64B lines, 4-cycle latency |
| **L2 Cache** | 16-way 256KB Unified, 12-cycle latency |
| **L3 Cache** | 32-way 4MB LLC, 25-cycle latency |
| **Miss Handling** | 64-entry Miss Buffer, 64-entry Load Fill Request Queue |
| **Main Memory** | 140-cycle latency |

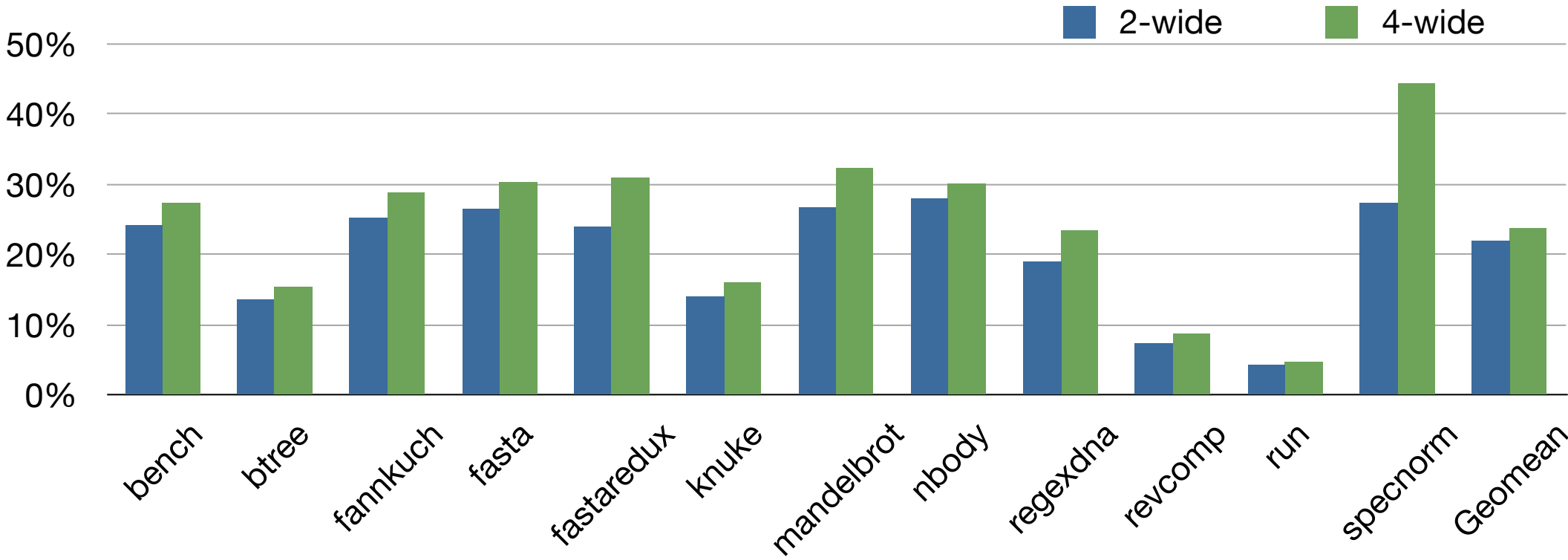
- Profile Guided Optimization LLVM 3.5
- Benchmarks with 0.3% dynamic instruction stream indir branch
- SPEC **TRAIN** input, PHP and Python **first** input
- Cycle Accurate x86 simulator PTLSim provides predictability
- Transform non-loop branches with pred > bias and (pred - bias) > 3%
- Run SPEC, PHP, Python on PTLSim using **PGO** binaries with and without prediction guide on **REF**
- Speculation support : (alias/shadow registers) for baseline and experimental
- Improved Indirect Branch Predictor: **VPC** (**TAGE** study in paper)

# Performance: SPEC95, SPEC2K, SPEC2K6



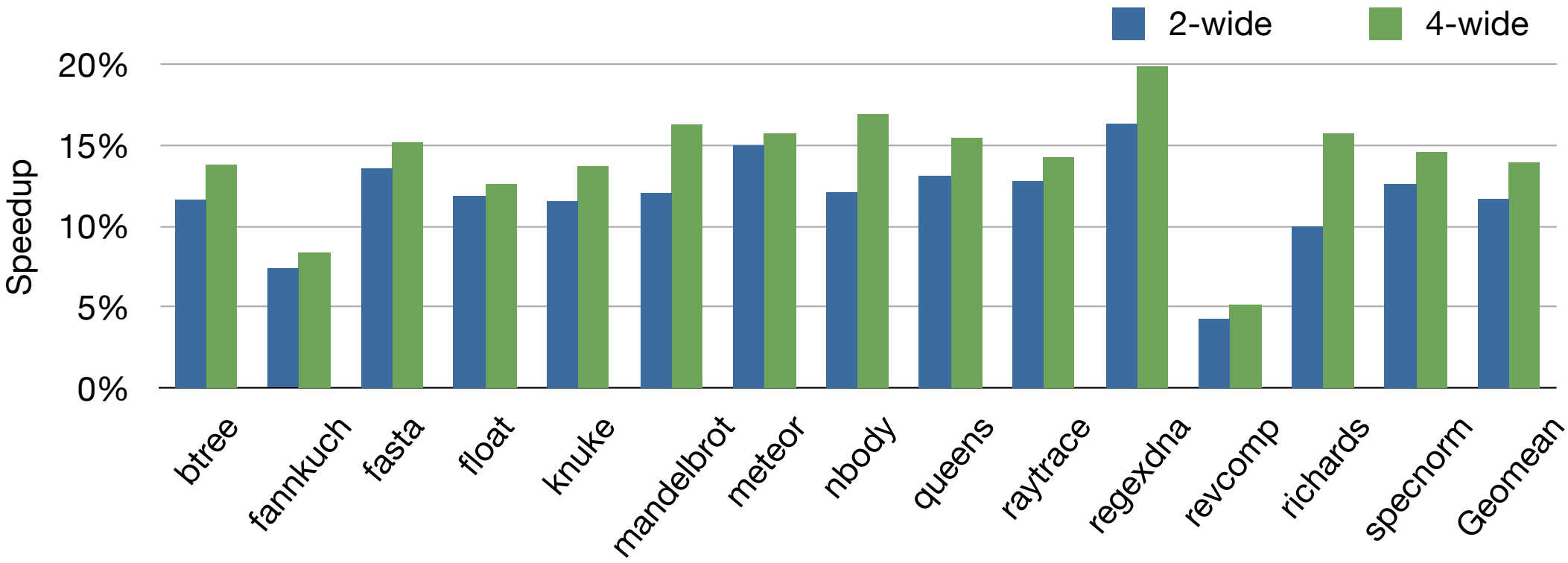

- Performance proportional to % of dynamic instructions which are indirect branches (**PDS**) and amenable to transformation (weighted averaged bias: **WAB**)
- richards: 4% **PDS**, 41% **WAB**
- m88ksim: 0.4% **PDS**, 57% **WAB**
- Geomean: 1.1% **PDS**, 62% **WAB**

# Performance: PHP



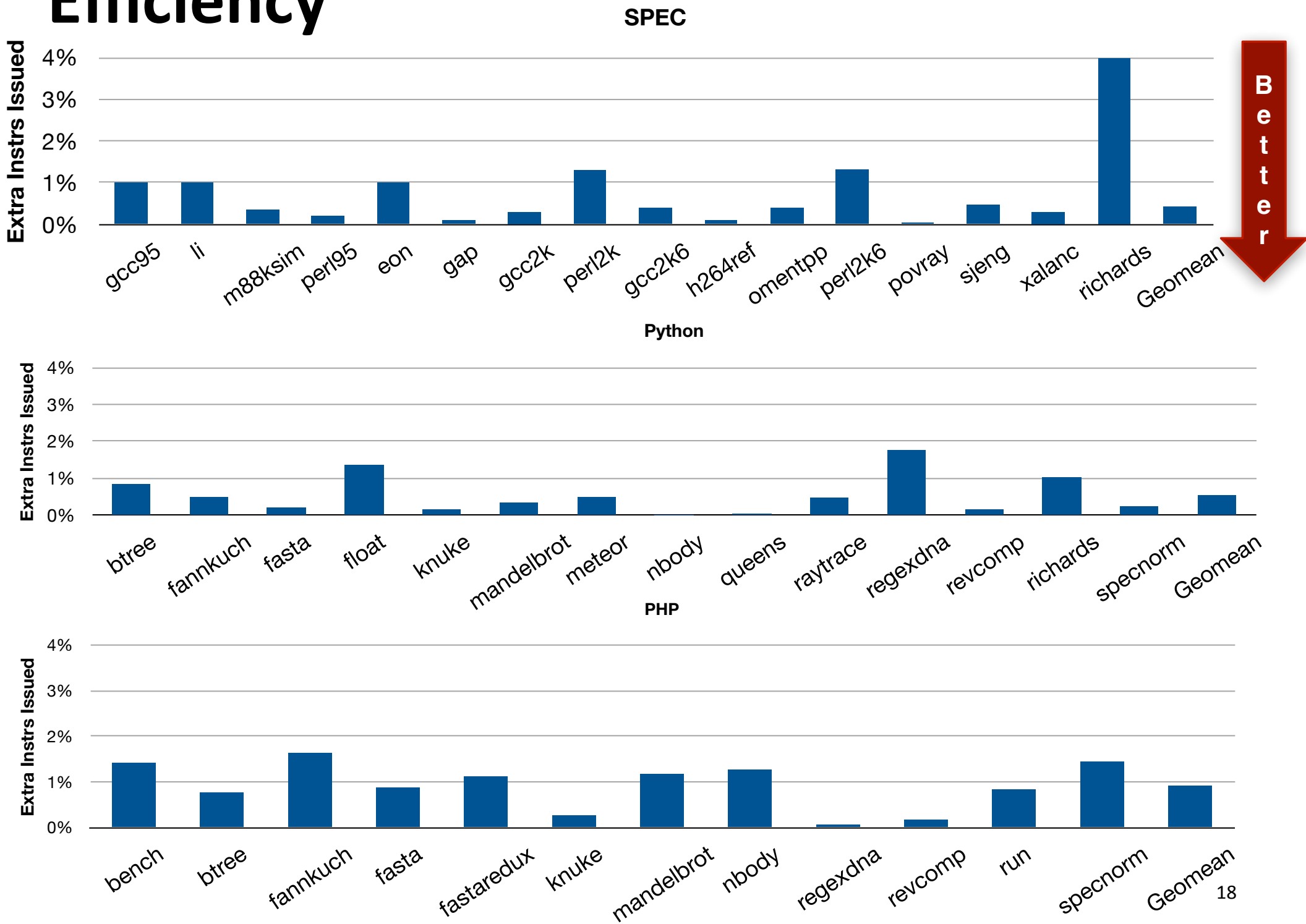

- Specnorm: 1.4% **PDS**, 18% **WAB**
- run: 0.6% **PDS,** 49% **WAB**
- Geomean: 1.4% **PDS,** 37% **WAB**

# Performance: Python




- regexdna: 2.7% **PDS**, 72% **WAB**
- revcomp: 1.0% **PDS**, 90% **WAB**
- Geomean: 1.8% **PDS**, 73% **WAB**

# Efficiency


SPEC
Extra Instrs Issued
4%
3%
2%
1%
0%
gcc95
li
m88ksim
perl95
eon
gap
gcc2k
perl2k
gcc2k6
h264ref
omentpp
perl2k6
povray
sjeng
xalanc
richards
Geomean
Better
Python
Extra Instrs Issued
4%
3%
2%
1%
0%
btree
fannkuch
fasta
float
knuke
mandelbrot
meteor
nbody
queens
raytrace
regexdna
revcomp
richards
specnorm
Geomean
PHP
Extra Instrs Issued
4%
3%
2%
1%
0%
bench
btree
fannkuch
fasta
fastaredux
knuke
mandelbrot
nbody
regexdna
revcomp
run
specnorm
Geomean

# Conclusions

- Straightforward, low-cost “enabling” transformation
  - Leverages DBT profiling and speculation facilities
- Modest Hardware Requirements
- Leverages Advances In Indirect Branch Prediction
- Good Performance across Integer and Floating Point
- Maintains the Efficiency of the In-Order

# Thank You

Carnegie Mellon
Electrical & Computer
ENGINEERING

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

# Thank You

# Questions?

Carnegie Mellon
Electrical & Computer ENGINEERING

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN