

June 3, 2018

@Championship Value Prediction

ISCA 2018

Organized and Sponsored by:

Qualcomm

# CVP1 Infrastructure & Traces

Arthur Perais

Qualcomm Technologies, Inc.



# Championship Value Prediction Infrastructure

- Value predictors can be characterized with :
  - Coverage :  $\frac{\textit{correct}}{\textit{predictable}}$
  - Accuracy :  $\frac{\textit{correct}}{\textit{correct}+\textit{incorrect}}$
- Contrarily to Branch Prediction, not predicting is *not equivalent* to mispredicting = not predicting is advised if value is unpredictable.
  - All predictors may not predict the same instructions.
  - Metrics not conclusive w.r.t. performance
- Need an infrastructure *that provides cycles*, not just accuracy/coverage/MPKI

# A Low Complexity Simulator for CVP



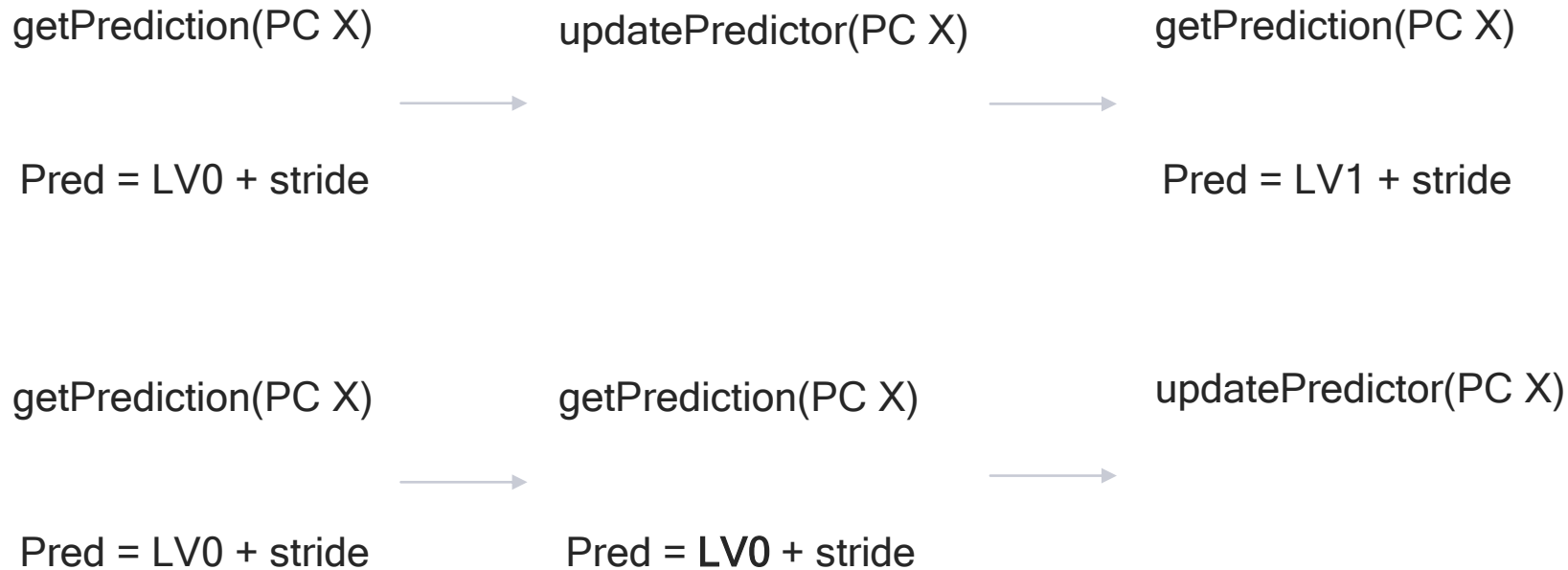
- Courtesy of Eric Rotenberg and Vinesh Srinivasan from NCSU.
- Finite-size instruction window (256).
- 16-wide frontend, perfect branch prediction.
- Dataflow-only constrained scheduling.
- Variable latency instructions (e.g., ALU vs. MUL)
- 3 levels of caches + main memory
  - 32KB 4-way, 2 cycles.
  - 1MB 8-way, 12-cycle
  - 8MB 16-way, 60-cycle
  - Main memory, 150-cycle
- Provides execution cycles to rank predictors using meaningful metric.

# Simulator API for Contestants

- **bool getPrediction() :**
  - returns "true" if microarch. simulator should speculate based on the prediction for this instruction, "false" if it should not speculate based on the prediction for this instruction.
  - Also provides prediction.
- **void updatePredictor() :**
  - Provides actual result and allows contestants to update their predictor non-speculatively.

# Simulator API for Contestants

- Assume stride predictor :

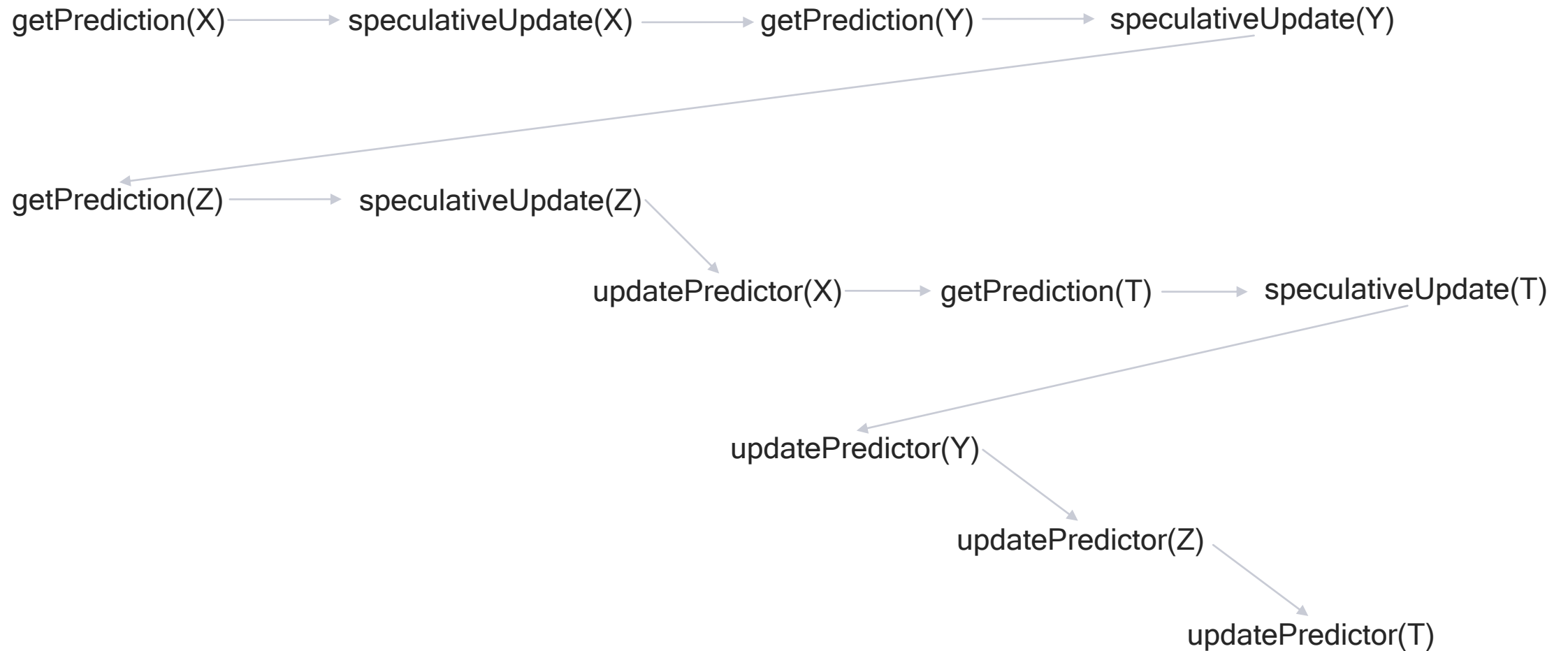


- Speculative update + fixup on misprediction in real pipeline.
- No pipeline flush in infrastructure, how do we enable this behavior?

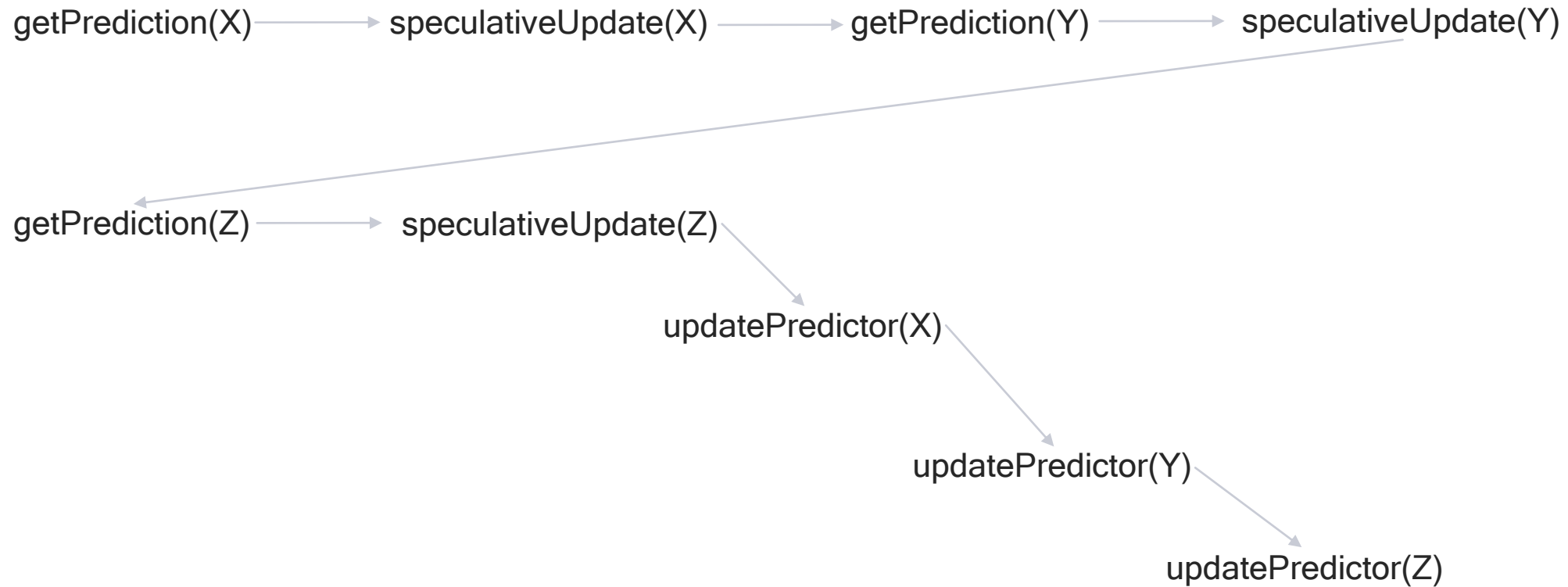
# Simulator API for Contestants

- `bool getPrediction()`
- `void updatePredictor()`
- `void speculativeUpdate()` :
  - If `getPrediction()` instructed the simulator to speculate, then `speculativeUpdate()` will reveal whether or not the predicted value is correct, immediately after `getPrediction()`.
  - If `getPrediction()` instructed the simulator to **NOT** speculate, then this call will not have the correct value. The contestants must wait until the non-speculative `updatePredictor()` function to see whether or not the prediction was correct.

# Simulator Flow - No misprediction



# Simulator Flow - Value Misprediction (Z)



- No call to `getPrediction()` until all pending calls to `updatePredictor()` have been drained from window (= full pipe flush on value misprediction).



# Workloads




- Public & Secret Traces :
  - SPEC'06/'17
  - Database-class
  - Crypto
  - **Workloads used to drive the design of state-of-the-art chips.**
  
- Public (135 traces) :
  - 30M instructions.
  
- Secret (2013 traces) :
  - 100M instructions.
  - Superset of Public traces.

# Program

|             |  |
|-------------|--|
| 8:30-8:40   | Introduction and Welcome   |
| 8:40-8:50   | CVP1 Infrastructure & Traces<br><i>Arthur Perais (Qualcomm)</i>  |
| 8:50-9:15   | Exploring Value Prediction with the EVES predictor<br><i>André Seznec (INRIA/IRISA, France)</i>  |
| 9:15-9:40   | H3VP: History-based High-reliable Hybrid Value Predictor<br><i>Kenichi Koizumi, Kei Hiraki and Mary Inaba (The University of Tokyo, Japan)</i>   |
| 9:40-10:05  | Context-Base Computational Value Predictor with Value Compression<br><i>Yasuo Ishii (ARM)</i>  |
| 10:05-10:30 | Break  |
| 10:30-10:55 | DFCM++: Augmenting DFCM with Early Update and Data Dependency-driven Value Estimation<br><i>Nayan Deshmukh, Snehil Verma, Prakhar Agrawal, Biswabandan Panda, Mainak Chaudhuri (Indian Institute of Technology Kanpur)</i>   |
| 10:55-11:50 | <b>Panel Discussion: "Speculation: Past, Present and Future"</b><br><i>Gurindar S. Sohi (Wisconsin-Madison), Avi Mendelson (Technion)</i><br><i>Chris Wilkerson (Nvidia), Andreas Moshovos (Toronto), Daniel Jiménez (Texas A&amp;M)</i><br>Moderator: <i>Rami Sheikh (Qualcomm)</i> |
| 11:50-12:00 | Announcement of results and awards<br><i>Arthur Perais and Rami Sheikh (Qualcomm)</i>  |



# Thank you

Follow us on:    

For more information, visit us at:

[www.qualcomm.com](http://www.qualcomm.com) & [www.qualcomm.com/blog](http://www.qualcomm.com/blog)

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

©2018 Qualcomm Technologies, Inc. and/or its affiliated companies. All Rights Reserved.

Qualcomm is a trademark of Qualcomm Incorporated, registered in the United States and other countries. Other products and brand names may be trademarks or registered trademarks of their respective owners.

References in this presentation to “Qualcomm” may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes Qualcomm’s licensing business, QTL, and the vast majority of its patent portfolio. Qualcomm Technologies, Inc., a wholly-owned subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of Qualcomm’s engineering, research and development functions, and substantially all of its product and services businesses, including its semiconductor business, QCT.