Extended Scaled Neural Predictor for Improved Branch Prediction
Zihao Zhou\textsuperscript{1}, Mayank Kejriwal\textsuperscript{1} and Risto Miikkulainen\textsuperscript{1}
\textsuperscript{1} [zhouz, kejriwal, risto]@cs.utexas.edu / \textsuperscript{}The University of Texas at Austin

Abstract
A perception-based scaled neural predictor (SNP) was implemented to emphasize the most recent branch histories via the following three approaches: (1) expanding the size of tables that correspond to recent branch histories, (2) scaling the branch histories to increase the weights for the most recent histories but decrease those for the old histories, and (3) expanding most recent branch histories to the whole history path. Furthermore, hash mechanisms and saturating value for adjusting threshold were tuned to achieve the best prediction accuracy in each case. The resulting extended SNP was tested on well-known floating point and integer benchmarks. Using the SimpleScalar 3.0 simulator, while different features have different impact depending on whether the test is floating point or integer, overall such a well-tuned predictor achieves an improved prediction rate compared to prior approaches.

Motivation
- Neural inspired branch prediction is infeasible using purely digital implementation, because of poor power and latency characteristics.
- Digital units may, in some cases, be replaced with analog equivalents but at the cost of precision [12].
- Scaled Neural Predictor (SNP) was thus proposed in 2008, along with analog implementations that uses converters and current summation instead of digital hardware [12].
- SNP achieved accuracy comparable to L-TAGE [16] (state-of-the-art), but is far more competitive from a power perspective because of analog implementation.
- Motivated to propose extensions to SNP and empirically evaluate its performance on different parameter settings because of its practical advantages.

Contribution
- Extensions include using both branch address and direction histories, as well as investigating multiple hashing functions for hashing branch addresses into compressed path address.
- Conduct comparisons on a wide range of both floating point and integer benchmarks.
- Evaluate effects of individual SNP features on branch prediction accuracy.

Experimental Results
In each of the figures below, the integer benchmark results are in (a) and the floating point benchmarks results in (b). All benchmarks are on the x-axis and the prediction accuracy, scaled to [0,1], is on the y-axis.

Future Work
- Incorporate more aggressive neural computation e.g. backpropagation to compute non-linear functions of correlation, and investigate if it performs better than simple neural model in original design.
- Incorporate more complex hashing mechanisms and investigate if they perform better than simply hashing a bit or a combination of two bits.

Conclusions
- SNP had already been shown in prior work to achieve state-of-the-art performance, and has promising analog implementation.
- Based on empirical investigations, recommendations were made on how the SNP could be modified for more accurate branch prediction.
- Existing design choices were also evaluated on different parameter settings with better prediction rates obtained by modifying some of the originally proposed parameters.
- Extended Scaled Neural Predictor can potentially result in a significant hardware application of neural networks in the near future.

References