NC State University

Department of Electrical and Computer Engineering

ECE 463/563: Fall 2022 (Rotenberg)

Project #2: Branch Prediction

by

Charlie Osborn

7. Tasks, Grading Breakdown
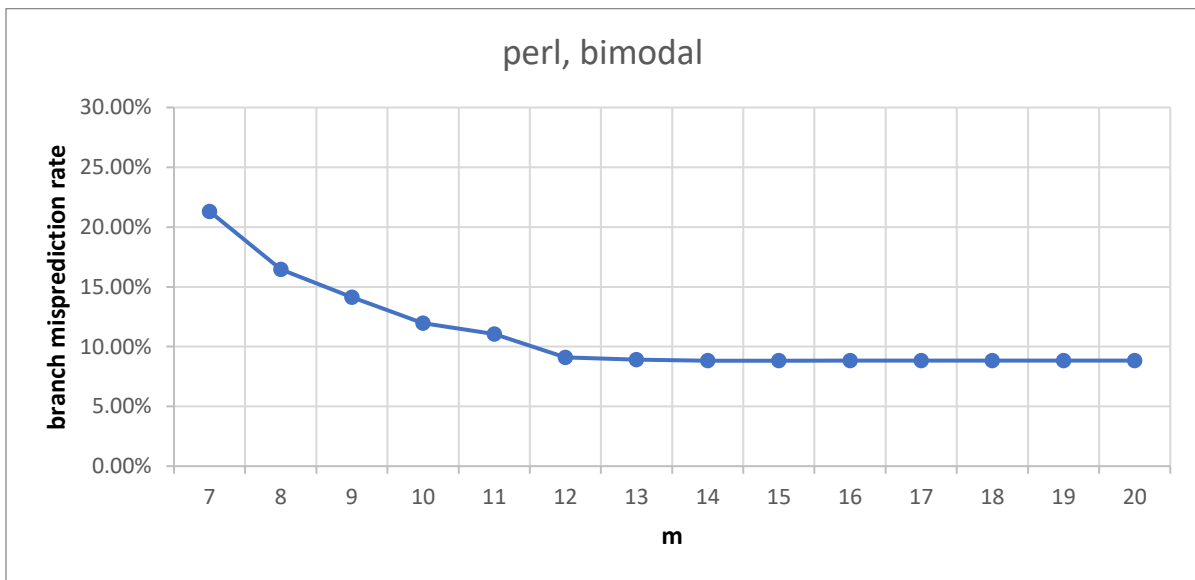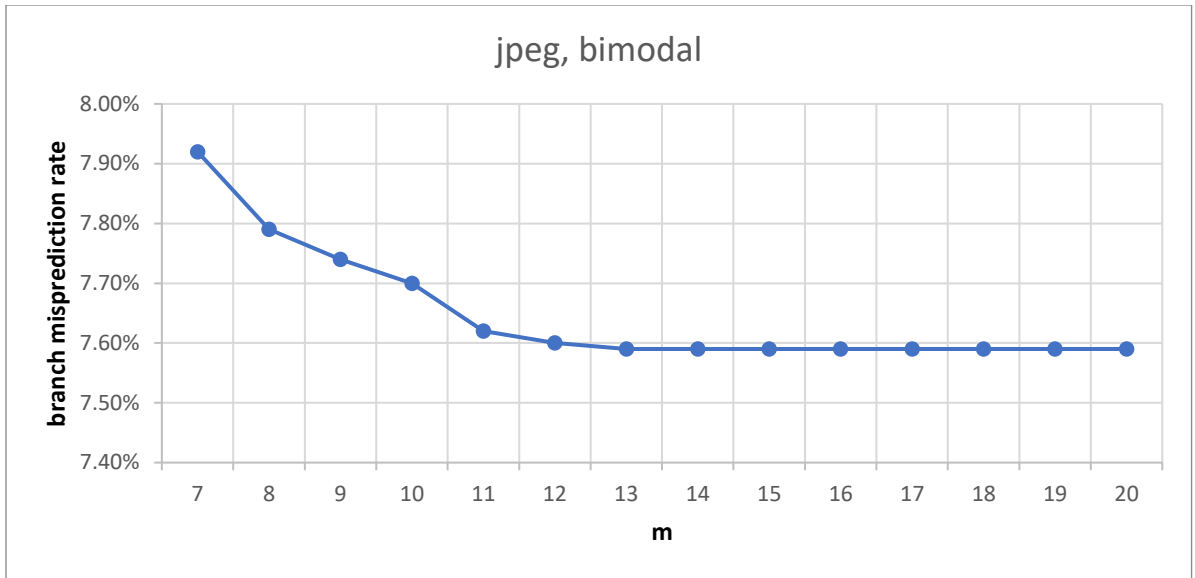
PART 1: BIMODAL PREDICTOR

(a)   [**ECE463: 25 points**] or [**ECE563: 20 points**] Gradescope will evaluate your simulator on the four validation runs "val_bimodal_1.txt", "val_bimodal_2.txt", "val_bimodal_3.txt", and "val_bimodal_4.txt", posted on the website for the BIMODAL PREDICTOR. Gradescope will also evaluate your simulator on one bimodal predictor mystery run. Each validation run and mystery run is worth ⅕ of the points for this part (5 or 4 points each). Gradescope must say that you match <u>all four validation runs</u> to get credit for the experiments with the bimodal predictor, however.

(b)   [**ECE463: 25 points**] or [**ECE563: 20 points**] Simulate BIMODAL PREDICTOR for different sizes ($7 \leq m \leq 20$). Use the traces ***gcc***, ***jpeg***, and ***perl***.
[20 or 15 points]   Graphs: Produce one graph for each benchmark.   Graph title: "*<benchmark>*, bimodal". Y-axis: branch misprediction rate. X-axis: *m*. Per graph, there should be only one curve consisting of 14 datapoints (connect the datapoints with a line).

jpeg, bimodal

perl, bimodal
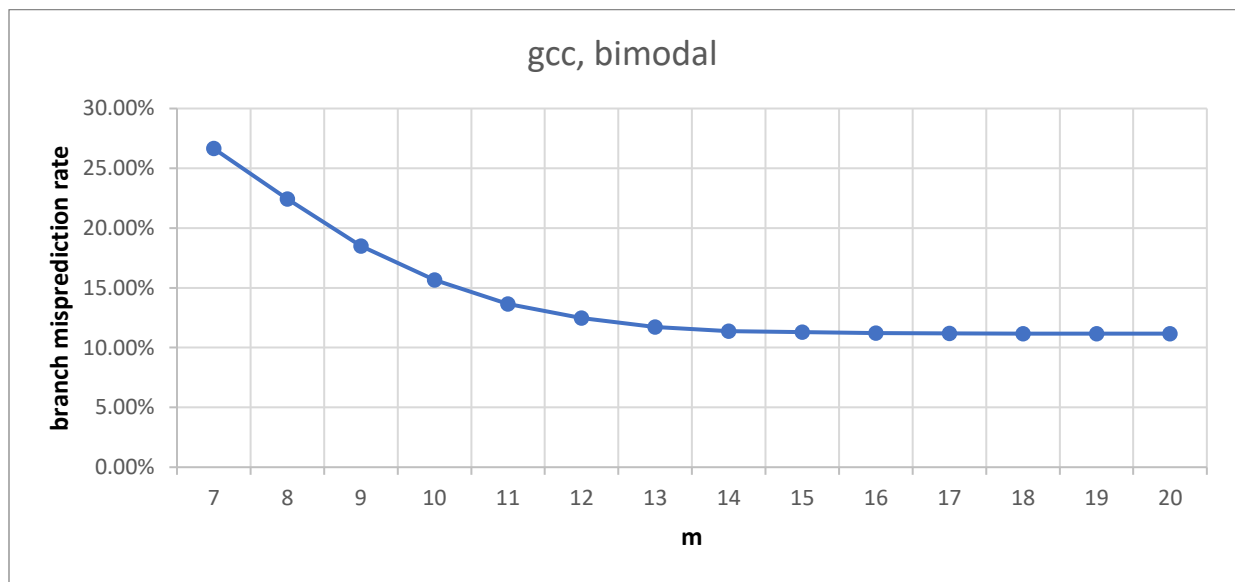
[5 points] Analysis:

1. [1 point] As the bimodal predictor's table size increases, the branch misprediction rate decreases.

2. [2 points] For each benchmark, indicate the minimum value of m at which the misprediction rate bottoms-out (reaches its minimum) and indicate its minimum misprediction rate. Fill in the table below.

Benchmark	Minimum "m" at which misprediction rate reaches its minimum	Minimum misprediction rate
gcc	$m = 18$	11.17%
jpeg	$m = 13$	7.59%
perl	$m = 14$	8.82%

3. [1 point] At some point, increasing the bimodal predictor's table size is of no value. At this point, each static branch (*i.e.*, each static branch PC) is allocated a dedicated index in the table. Given that interference among different static branches is eliminated at this point, the only way to improve accuracy further is a better prediction algorithm.

4. [1 point] I infer that **gcc** has more static branches (*i.e.*, unique static branch PCs) than **jpeg**, because **gcc** requires more table entries than **jpeg** before its misprediction rate bottoms-out.

PART 2: GSHARE PREDICTOR

(a)   [**ECE463: 25 points**] or [**ECE563: 20 points**] Gradescope will evaluate your simulator on the four validation runs "val_gshare_1.txt", "val_gshare_2.txt", "val_gshare_3.txt", and "val_gshare_4.txt", posted on the website for the GSHARE PREDICTOR. Gradescope will also evaluate your simulator on one gshare predictor mystery run. Each validation run and mystery run is worth ⅕ of the points for this part (5 or 4 points each). Gradescope must say that you match all four validation runs to get credit for the experiments with the gshare predictor, however.

(b)   [**ECE463: 25 points**] or [**ECE563: 20 points**] Simulate GSHARE PREDICTOR for different sizes ($7 \leq m \leq 20$), and for each size, *i.e.*, for each value of m, sweep the global history length n from 0 to m. Use only the trace **gcc**.
[20 or 15 points] Graphs: Produce one graph for **gcc**. Graph title: "gcc, gshare". Y-axis: branch misprediction rate. X-axis: n (spanning $n=0$ to $n=20$). For this graph, there should be a total of 203 datapoints plotted as 14 curves. Datapoints having the same value of m (same predictor size) are connected with a line, *i.e.*, one curve for each value of m. Note that not all curves have the same number of datapoints; see the listing below for the number of datapoints for each of the 14 curves, $m=7$ through $m=20$. The rationale for this graph is to study the effect of global history length for each predictor size.
m=7 curve has 8 datapoints: $0 \leq n \leq 7$
m=8 curve has 9 datapoints: $0 \leq n \leq 8$
m=9 curve has 10 datapoints: $0 \leq n \leq 9$
m=10 curve has 11 datapoints: $0 \leq n \leq 10$
m=11 curve has 12 datapoints: $0 \leq n \leq 11$
m=12 curve has 13 datapoints: $0 \leq n \leq 12$
m=13 curve has 14 datapoints: $0 \leq n \leq 13$
m=14 curve has 15 datapoints: $0 \leq n \leq 14$
m=15 curve has 16 datapoints: $0 \leq n \leq 15$
m=16 curve has 17 datapoints: $0 \leq n \leq 16$
m=17 curve has 18 datapoints: $0 \leq n \leq 17$
m=18 curve has 19 datapoints: $0 \leq n \leq 18$
m=19 curve has 20 datapoints: $0 \leq n \leq 19$
m=20 curve has 21 datapoints: $0 \leq n \leq 20$

<< INSERT GRAPH "gcc, gshare" HERE >>

[5 points] Analysis:

Insight: With the bimodal predictor (n=0: no global history), a given static branch is predicted using only a single 2-bit counter. With the addition of global history, that single counter is *specialized* or *multiplied* into many more counters. All these counters are used by the same static branch for more specialized predictions among its dynamic instances, promising higher accuracy *as long as there are adequate counters available in the table* (not just for this static branch, but for all static branches). Thus, with the addition of global history, there is a need for more 2-bit counters … the key idea being that, at some point, bimodal cannot even take advantage of more counters (see your analysis section for bimodal, above) whereas gshare *can*. Summarizing: *Gshare needs an abundance of counters and, unlike bimodal, it can exploit abundant counters for higher accuracy.*

1. [0.5 points] At small table sizes, global history can <help/hurt> accuracy. This is because there are <too few/abundant> counters.

2. [0.5 points] At large table sizes, global history can <help/hurt> accuracy. This is because there are <too few/abundant> counters.

3. [2.5 points] For each table size (m), indicate the global history length (n) that yields the lowest misprediction rate, indicate what that lowest misprediction rate is, and indicate the misprediction rate for bimodal. Fill in the table below.

m	Global history length (n) that yields the lowest misprediction rate	Lowest misprediction rate	Bimodal misprediction rate (*i.e.*, for $n=0$)
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

4. [0.5 points] The *smallest* bimodal predictor that achieves the best bimodal accuracy (lowest misprediction rate among all bimodal configurations) is as follows:
Best bimodal: $m =$ _____, misp. rate = _____.

5. [0.5 points] The *smallest* gshare predictor that achieves the best gshare accuracy (lowest misprediction rate among all gshare configurations) is as follows:
Best gshare: $m =$ _____, $n =$ _____, misp. rate = _____.

6. [0.5 points]   In conclusion, with adequate predictor storage budget, gshare <rocks/sucks>.

PART 3: HYBRID PREDICTOR (ECE563 students only)

[ECE563: 20 points] Gradescope will evaluate your simulator on the two validation runs "val_hybrid_1.txt" and "val_hybrid_2.txt" posted on the website for the HYBRID PREDICTOR. Gradescope will also evaluate your simulator on two hybrid predictor mystery runs. Each validation run and mystery run is worth ¼ of the points for this part (5 points each).