PDPM: Perceptron-Driven Pipeline Morphing for Adaptive RISC-V Processors

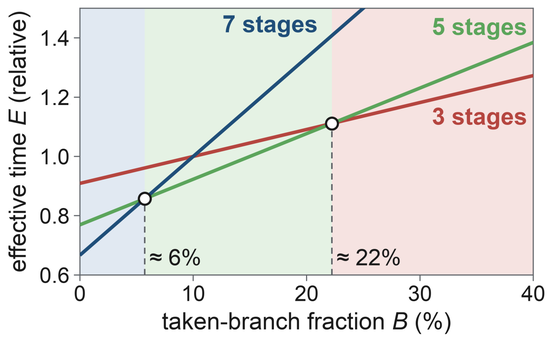
Paper ID **2026337954**

Dawit Chun[1] **Minhee Jun**[2]

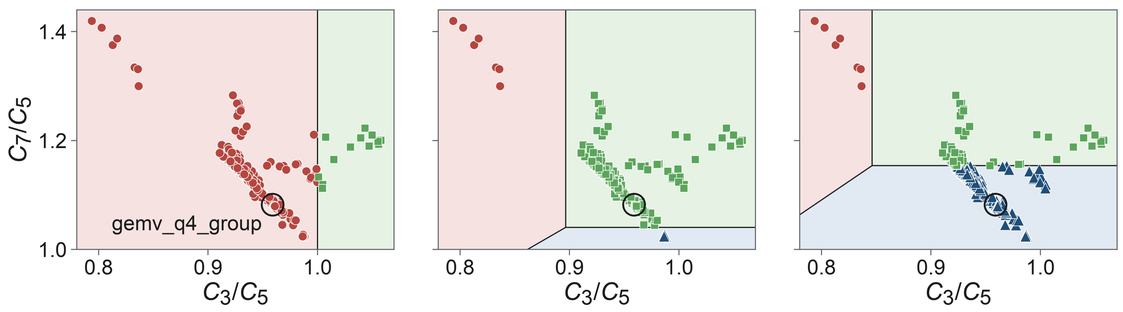
[1]Taejae University, Seoul, Republic of Korea
[2]The Catholic University of America, Washington, DC, USA

Motivation

The fastest depth changes with the workload, between phases of one program, and with the clock plan, yet processors fix it at design time.



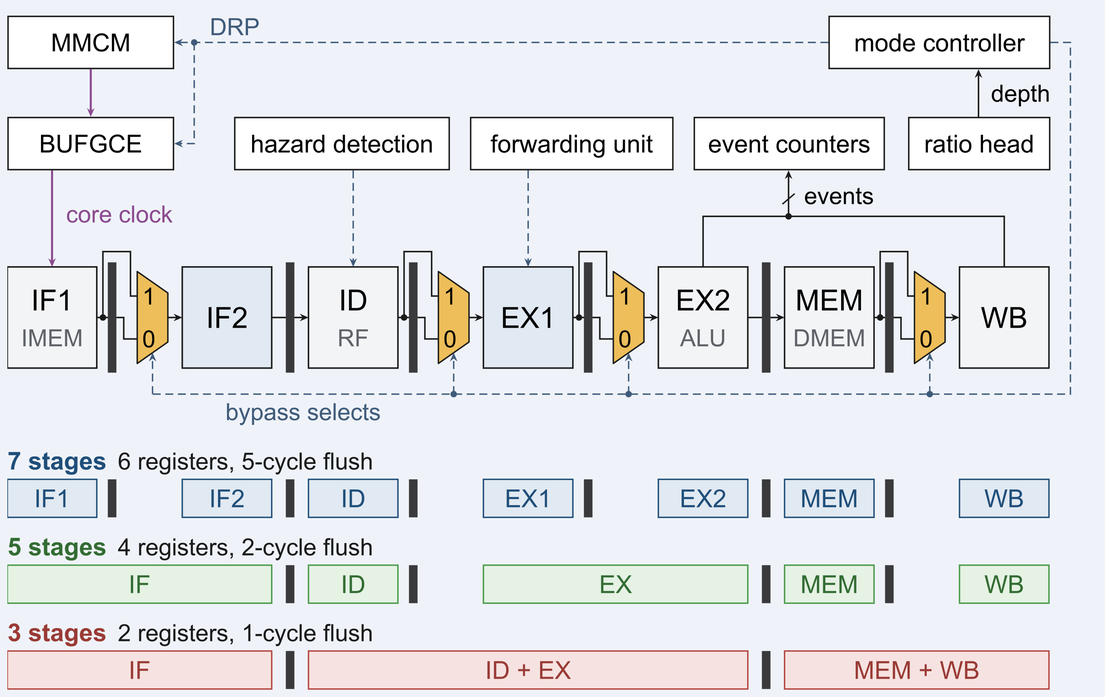
Deeper clocks faster, but flushes cost more. Effective time is E = CPI / α. Seven, five, and three stages clock at 1.5 : 1.3 : 1.1 and, with no branch predictor, lose 5, 2, and 1 cycles per taken branch: seven stages win below about 6% taken branches, three above 22%.



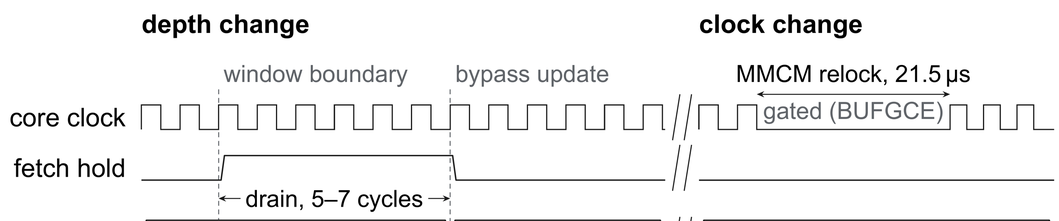
Depth is a property of the workload and the clock plan together. Across three clock plans, most of the 300 kernels flip from 3 to 5 to 7 stages, and actively switching to the best depth per phase saves 0.45% on average over the best fixed depth.

Architecture

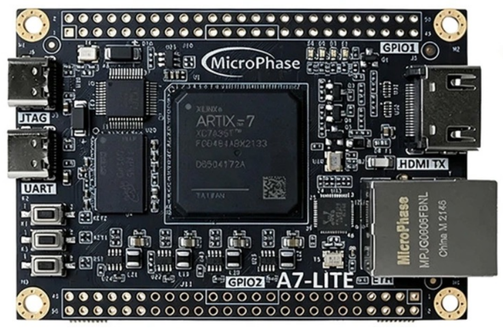
One seven-stage RV32IM pipeline; four bypass multiplexers make registers transparent to form 5- and 3-stage modes over one register file and memory.



A learned head picks the depth once per instruction window. Event counters feed the ratio head; its argmin tells the mode controller to drain and set the four bypass selects. Register file and memory are shared by all depths.



A depth change is cheap, a clock change is not. Depth: a 5–7-cycle drain, no state transferred. Clock: the core waits 21.5 µs while the MMCM relocks.

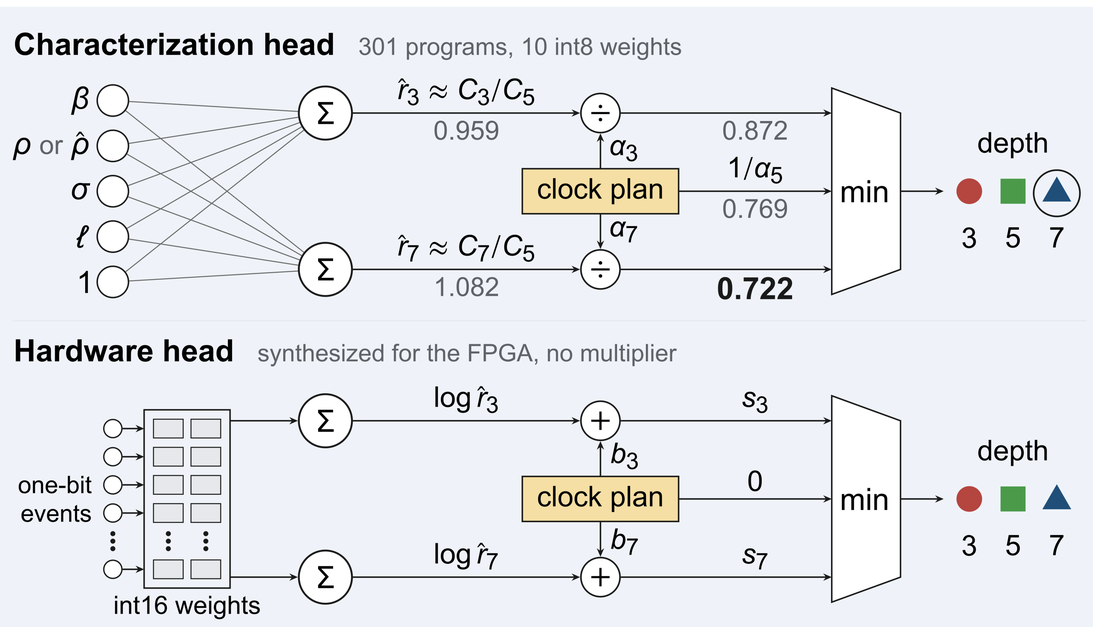


Clock, three-depth head	87.1 MHz
Clock, one-ratio head	90.32 MHz
Predictor	**478 LUT, 0.75%**
With the mode FSM	1.30%
Byte-identical runs	**641 / 641**

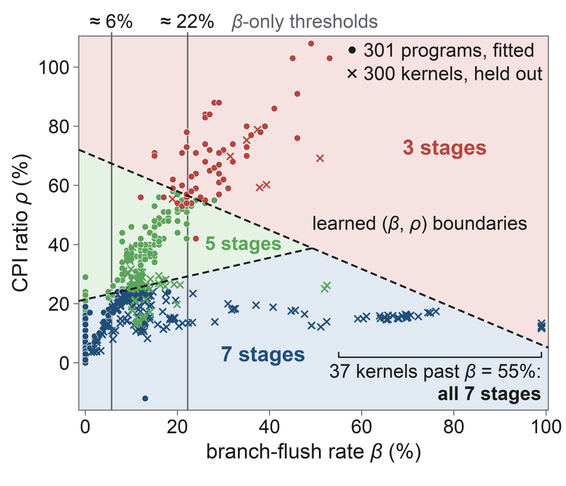
Closed loop on an Artix-7 XC7A100T, no host in the loop. Board image: MicroPhase A7-Lite.

Learned depth selection

Instead of classifying twelve (depth, clock) configurations, the head regresses two cycle ratios, so a new clock plan changes constants, not weights.



Two forms of one rule. Branch rate β, CPI-depth sensitivity ρ (or a single-mode proxy that a running core can measure), stall share σ, and length ℓ give two ratios, here for one measured kernel; on the board, each firing event adds its int16 weight.



Branch rate alone cannot pick the depth. The analytical thresholds (solid) assume ρ = σ = 0, yet most real programs lie between ρ = 10% and 80%.

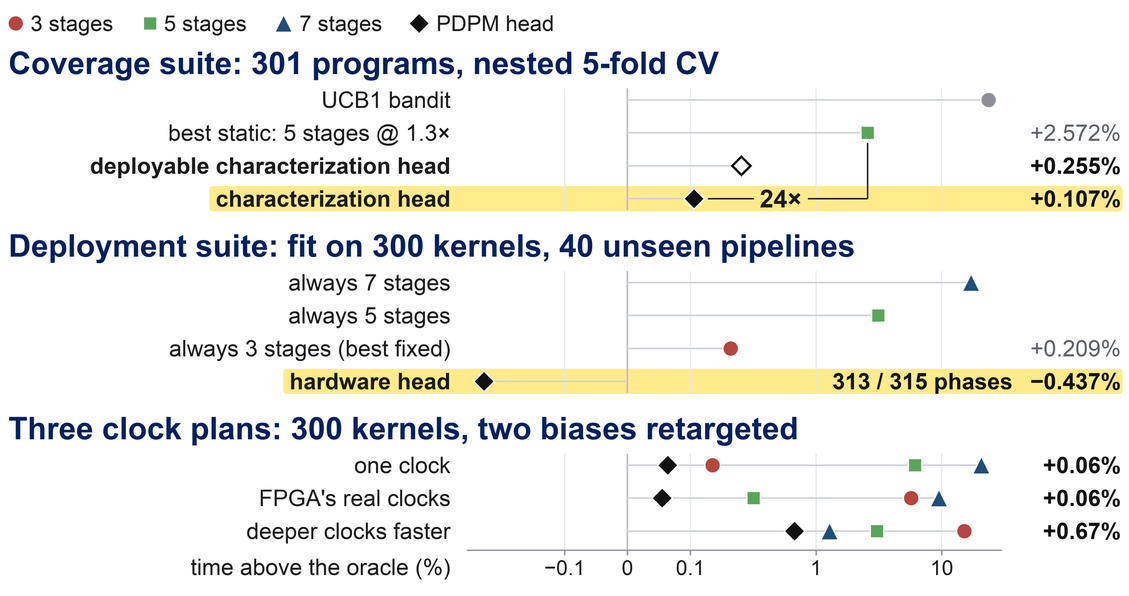
The learned boundaries (dashed) are lines in (β, ρ): they place 92.7% of the 301 fitted programs and 87.3% of 300 held-out kernels correctly. All 37 kernels past β = 55% are fastest at 7 stages; a β threshold calls each of them 3 stages.

Predictor, same four features	Weights	Accuracy
12-class linear	60	91.62%
3-class depth	15	90.73%
2 ratios, PDPM	**10**	**94.28%**
2 ratios, deployable proxy	10	89.40%

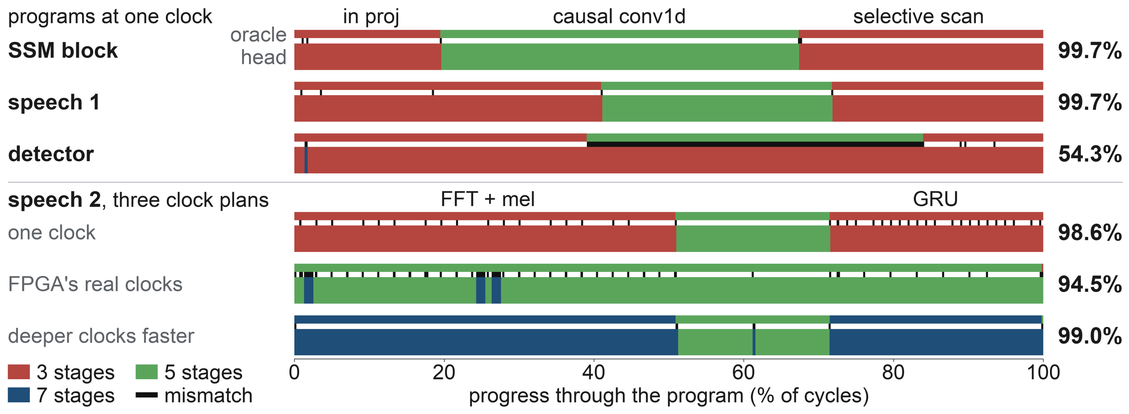
Ratios beat labels. Same features, data, and 45 trials. Two ratios with 10 weights are +3.11 pp more accurate than three classes ($p = 1.3 \times 10^{-4}$), and their errors cost less.

Evaluation

On 40 unseen pipelines, the head picks the oracle depth for **313 of 315** phases.



The head is lowest in every test, 24× below the best static configuration.



Without retraining, it adaptively changes depth inside complex programs it never saw, under every clock plan.

TAEJAE UNIVERSITY

THE CATHOLIC UNIVERSITY OF AMERICA

GitHub