CORF: Coalescing Operand Register File for GPUs

Hodjat Asghari Esfeden[1], Farzad Khorasani[2], Hyeran Jeon[3], Daniel Wong[1], and Nael Abu-Ghazaleh[1]

[1] UC Riverside — University of California

[2] Georgia Tech

[3] SJSU — San José State University

GPUs Have Huge Register Files

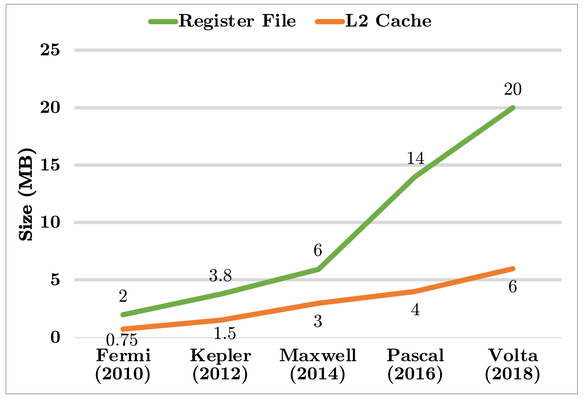


Figure 1: Register File vs. L2 Cache size.

- Register File is a critical structure in GPUs.

Narrow-width Values Are Prominent

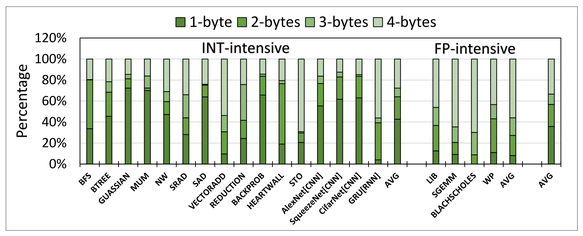


Figure 2: Width distribution of accessed registers.

- On average, 65% of all register operations contain narrow-width values.

Coalescing Opportunities & Register Packing

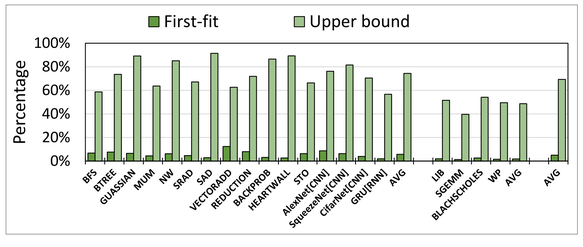
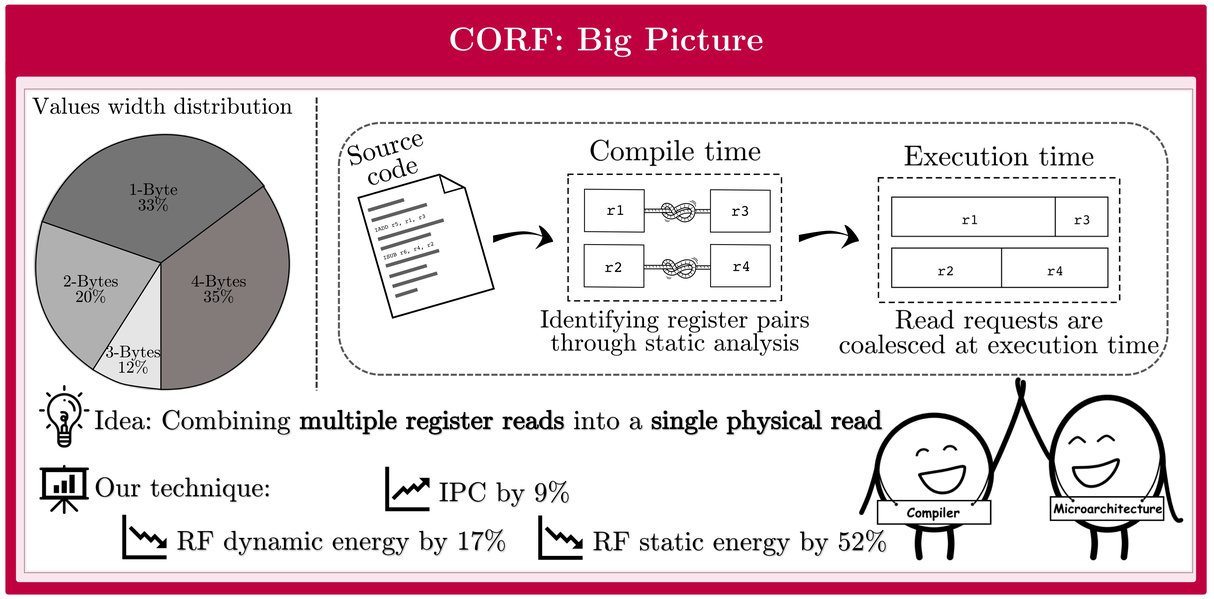


Figure 3: Instructions with coalesce-able register reads.

- First fit is clearly weak in promoting coalescing.

CORF: Big Picture



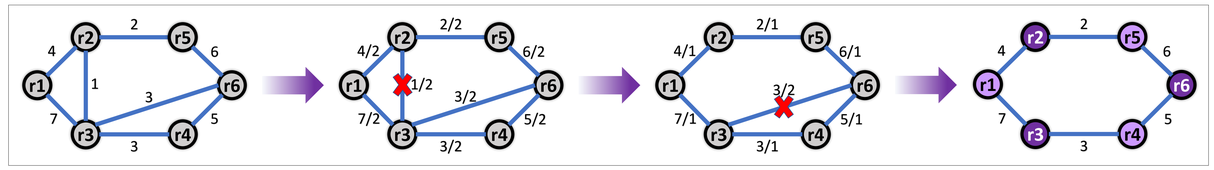


Figure 4: CORF++ register assignment heuristic example.

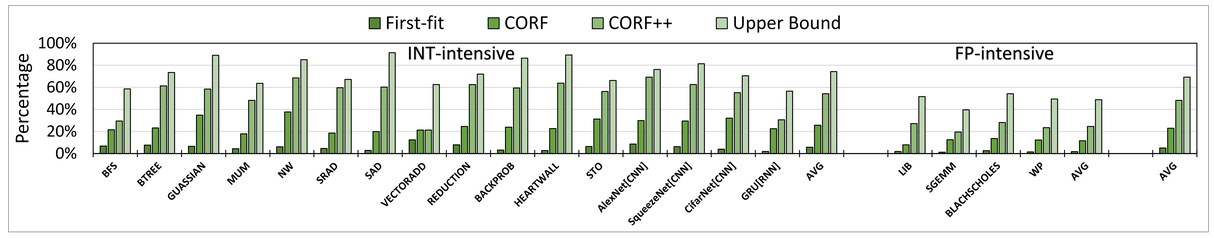


Figure 5: Coalesced instructions: CORF and CORF++ significantly increases the amount of coalescing opportunities.

CORF: Coalescing Operand Register File

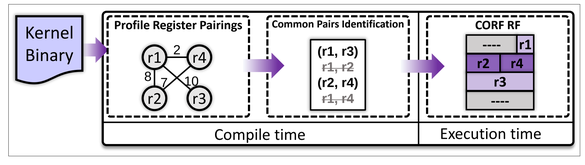


Figure 6: CORF overview.

- Compiler *statically* identifies *register pairs*.
- During run-time, register pairs will be dynamically packed together.

Main limitation: Each register may only be coalesced **exclusively** with one other register.

CORF++: Re-architected Register File

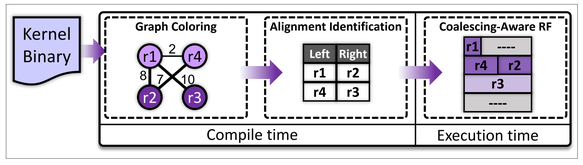


Figure 7: CORF++ overview.

- Compiler solves a variant of graph coloring problem (Figure 4) to simplify the allocation to left or right-aligning assignment.
- During runtime, any left-aligned register is coalescable with any right-aligned register providing they don't overlap.

Evaluation

- Simulator: GPGPU-Sim modeling NVIDIA Fermi architecture.
- Workloads: Rodinia, Parboil, CUDA-SDK, and Tango.

Coalescing opportunity: As it is shown in Figure 5, CORF and CORF++ significantly increase the amount of coalescing opportunity.

Performance improvement: CORF and CORF++ improve IPC by 5% and 9%, respectively.

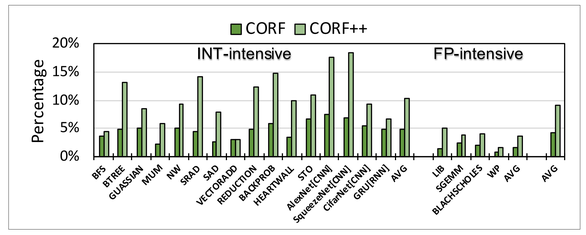


Figure 8: IPC Improvement.

RF access reduction: CORF and CORF++ reduce RF accesses by 10% and 23%, respectively.

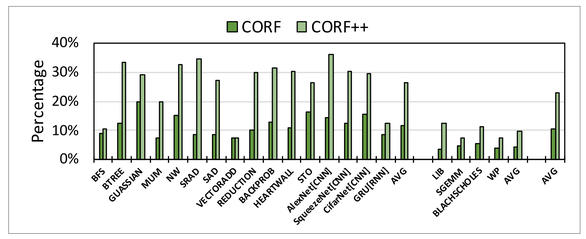


Figure 9: RF Access Reduction.

Also In the Paper...

- Combined CORF++ and Register File Virtualization [Jeon'15], resulting in the smallest known effective register file size among register compression proposals. The benefit of the two techniques add up, but their overhead do not.
- Detailed hardware/software overhead analysis.

Talk

Session 2H, Tuesday April 16th 4:10 pm