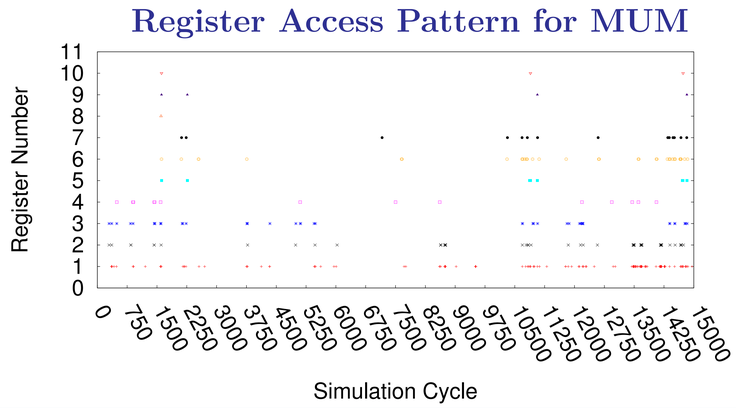
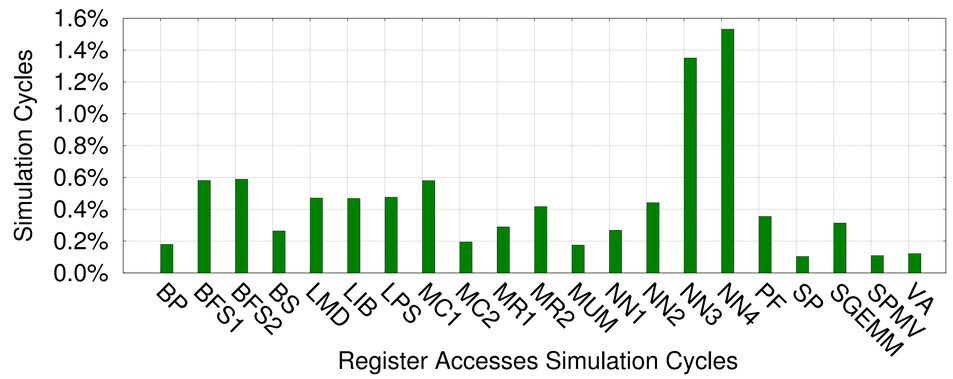
GREENER: A Tool for Improving Energy Efficiency of GPU Register File

Vishwesh Jatala (IIT Kanpur), Jayvant Anantpur (Mentor Graphics), Amey Karkare (IIT Kanpur)

Problem: Leakage Power Dissipation in GPUs

1. With decrease in the feature size of the semi-conductor devices, sub-threshold leakage power became a crucial factor in the manufacturing process.
2. Leakage power of GPU constitutes more than 50% of the total power [2].
3. Register files in the GPU dissipate about 15% of the total power [1].
4. Registers in a GPU dissipate leakage power throughout the entire execution of its warp irrespective of their access pattern.



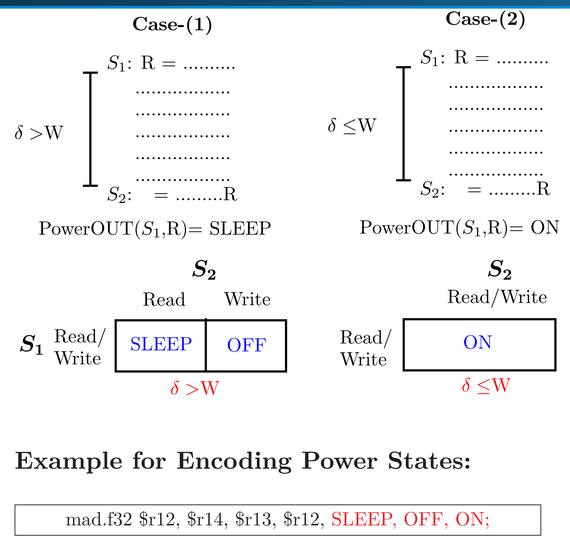


Solution: GREENER

Idea: Reduce the leakage power by putting the registers into low power states based on their access patterns.

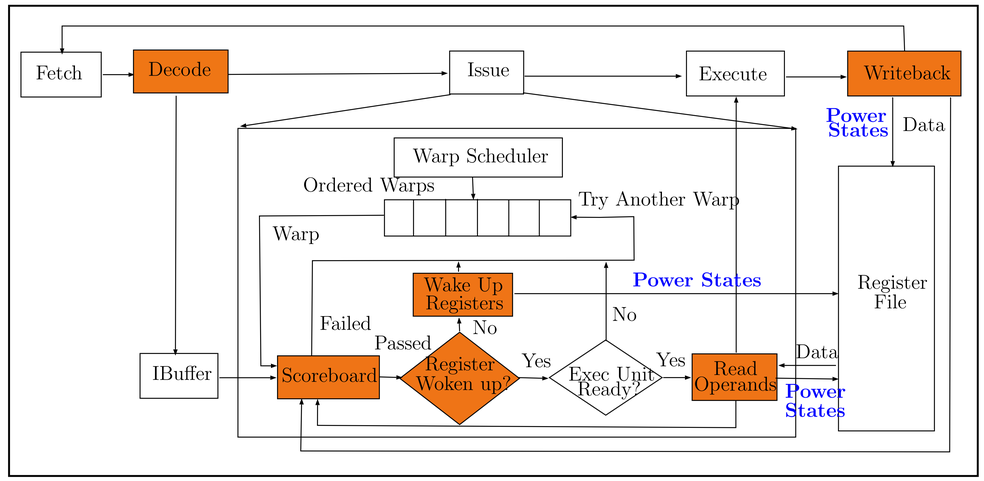
Strategy:

1. Employ a compile-time analysis to estimate the run-time register access information.
2. Use result of the analysis to determine the power states (ON, SLEEP, or OFF) after each instruction.
3. Modify the instructions to encode power state of the registers with the instruction, and transform an input assembly code to a power optimized assembly code.



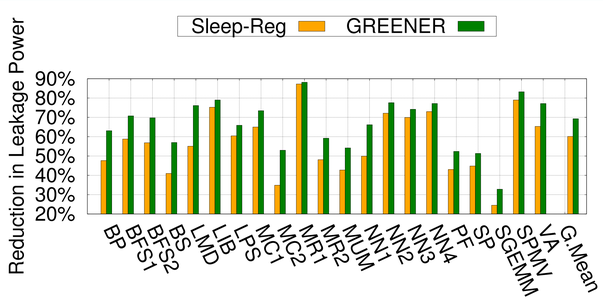
Example for Encoding Power States:

```
mad.f32 $r12, $r14, $r13, $r12, SLEEP, OFF, ON;
```

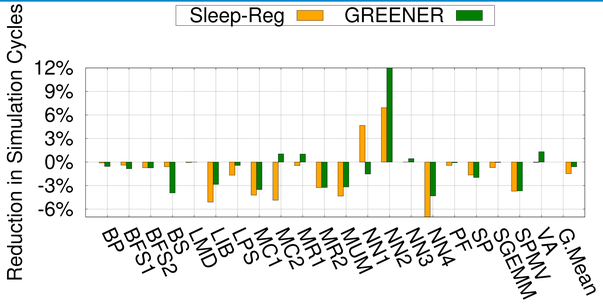


Modifications to GPU Pipeline

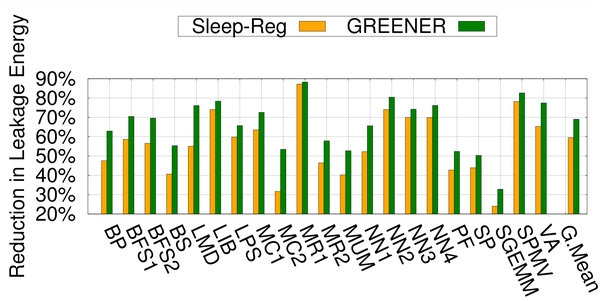
Experimental Results



Comparing Register Leakage Power



Comparing Performance



Comparing Register Leakage Energy

Experimental Setup

Resource	GPU Configuration
Architecture	NVIDIA Tesla K20x
No of SMs	14
Register File Size per SM	256KB
Technology Node	32nm

References

[1] H. Jeon, G. S. Ravi, N. S. Kim, and M. Annavaram. GPU Register File Virtualization. In *MICRO*, 2015.

[2] J. Lim, N. B. Lakshminarayana, H. Kim, W. Song, S. Yalamanchili, and W. Sung. Power Modeling for GPU Architectures Using McPAT. *ACM TODAES*, June 2014.