Rocket

Mauro Bianco, Carlos Osuna, Paolo Crosetto, Stefan Moosbrugger, Vassilis Karakasis, Jacob Progsch

Problem trying to solve

- GridTools Libraries implementation:
  - For solving PDEs in weather and climate.
  - For "generic" numerical algorithms on structured grids

Prior Profile

- Focus on an horizontal and a vertical stencils used in weather codes:

  - Fourth order horizontal diffusion operator (finite differences)

  - Vertical advection (implicit tridiagonal solve)

Evolution and Strategy

- Baseline:
  - smem for intermediate results
  - Global load through texture
- Experiments
  - K-parallel (vertical)=>(5% to 60%)
  - Prefetching in smem=>(no)
  - Prefetching in regs=>(no)
  - Cache in registers=>(30%)
  - Using double2=>(0% to 21%)
  - Const mem for strides=>(17%)
  - Warp assignment=>(no)
  - Cache bank size=>(13%)

Results and Final Profile

- Accomplished overall speedup of more than 50% on a specific case
- How easy and how general will be the implementation inside GridTools?
- How will these result evolve on newer architectures?

What problems you encountered

- We were expecting the compiler to do some of these things (register cache, efficient unrolling)
- The baseline concerning the warp assignment was already optimal
- Instruction Level Profiler is already available (CUDA 7.5) but not with Kepler

Wishlist

- CUDA compiler:
  - Why __device__, __host__ on user functions?
  - Support for more STL header functions
  - Support recent standards (recently quite some improvements)
  - Public roadmap: when will feature become available?
  - Kill warnings for calls to "__host__ __device__"
  - Faster compilation?