Software-First, Software-Mostly: Fast-Starting with Parallel Programming for Processor Array Overlays

(Audience: makers? erstwhile performance software engineers now compute accelerator developers)

Jan Gray
jan@fpga.org
“Quick! Make Our Workload Run Much Faster on these CPU+FPGA servers”

• Lessons from Catapult Bing Ranking
  – Classic FCCM speedups of key bottlenecks but!
  – Host ➔ code ➔ code ➔ FCCM ➔ host (many LOCs!)
  – Initial port: “big bang” of RTL + overlays + compilers
  – Software evolves ➔ maintenance and agility
  – So much infrastructure! Such skillsets
Software-First Software-Mostly

• Run the same (parallel) software on host or FPGA
• OpenCL!! OpenCL?
• Parallel models target parallel processor overlays
  – 1-1000 soft processors
  – Big bang
  – Up and running in minutes/hours not weeks
  – Custom FUs and standalone accelerators? Vendor libraries
  – FPGA dev as incremental performance engineering
  – Fewer 1 hour PARs; more 10 second recompiles
• Frameworks, libraries, ready to fly IP
• Many more can do this = much larger talent pool
Example

• Scenario: let us encrypt our data in flight!

• Approach
  – git clone OpenSSL and extract AES C code “as is”
  – Add parallelism i.e. use message passing to distribute work items, collect results
  – Compile and run on host – it works
  – Compile and run on FPGA – it works. Ship it!
  – Too slow? Profile, map hot spots to new FUs/FCCMs

• So much infrastructure!
HW Infrastructure? GRVI Phalanx W.I.P.
[talk tomorrow]

Maker scale
32 RISC-Vs. Digilent Arty $99

Industrial scale
400 RISC-Vs. KCU105 $2500