FPGA-based Accelerated Cloud Computing with AWS EC2 F1 and SDAccel

FPGA 2018, 02/25/2018 – Monterey (CA)

Parimal Patel (parimalp@xilinx.com)
XUP Senior Systems Engineer
Tutorial Objectives

- Overview on cloud computing and Xilinx FPGA-based acceleration using Amazon AWS EC2 F1 instance.

- Introduction to FPGA acceleration using high level languages via Xilinx SDAccel.

- Provide hands-on experience on using AWS EC2 F1 and Xilinx SDAccel.
Agenda

• Overview of AWS EC2 F1 and SDAccel
• Get started on AWS EC2 F1
  • Lab 1
  • Lab 2
  • Lab 3
• Wrap-up
Introducing AWS EC2 F1

- Amazon AWS EC2 F1 is a compute instance with Xilinx FPGAs which can be programmed to create custom hardware accelerated applications.

- F1 instances are easy to program and come with everything needed to develop, simulate, debug, and compile hardware accelerators.

- Once a FPGA design is complete, it can be registered as an Amazon FPGA Image (AFI), and deployed to F1 instance in just a few clicks.
Benefits of the AWS EC2 F1 Cloud Compute Platform

- Makes FPGA acceleration available to a large community of developers, and to millions of potential AWS users
- Provides dedicated and large amounts of FPGA logic with elasticity to scale to multiple FPGAs
- Simplifies the development process by providing cloud-based FPGA development tools
- Provides a Marketplace for FPGA applications, giving more choice, secure and easy access to millions of AWS users
The SDAccel Development Environment

- **C/C++ Application**
- **RTL, C/C++, OpenCL Accelerators**

Host CPU

Xilinx FPGA

PCle
The SDAccel Development Environment

- Fully integrated Eclipse-based environment
- Develop, profile and deploy applications accelerated with Xilinx FPGAs
- Concurrent programming of the host application and FPGA kernels
- Automatic hardware execution flows
- Built-in debug, profiling and performance analysis tools
How does FPGA Acceleration Work on AWS?

CPU handles the rest

FPGA handles compute-intensive, deeply pipelined, hardware-accelerated operations
AWS EC2 F1 Platform Model

CPU

- Custom Application
- OpenCL API
- OpenCL Runtime
- Drivers

Xilinx FPGA

- Custom Kernels
- AXI Interfaces
- DDR
- DMA

Shell

PCIe

User Application Code

AWS EC2 F1 Platform
It’s Time to Get started on AWS EC2 F1
Labs

Instructions at

https://github.com/Xilinx/AWS-F1-Developer-Labs/blob/master/README.md

Timeline

– Connecting to your F1 instance (~10 minutes)
– Experiencing F1 acceleration (~15 minutes)
– Developing and optimizing F1 application with SDAccel (~30 minutes, may be a stretch goal)
– Wrap-up (10:15 – 10:25 AM)
Lab 1: Logging In on AWS EC2 F1

- Click on the provided link to your preconfigured instance
- Use the provided Account ID
  - Xilinx-aws-f1-developer-labs
- Enter provided user name
  - user
- Enter the following password
  - <will be provided> (same for centos user on RDP)

- You will run “hello world” application
- Precompiled FPGA binaries
Setting up RDP Screen Color Depth And Other Tips

▶ You can copy and paste commands between the instructions window and terminal window of RDP using Ctrl+C (copy) and Ctrl+Shift+V (paste)
▶ If RDP crashes/freezes, you may have to reboot the instance
  – Go to the local web browser where you started the instance, select it, click on Actions, and select Instance State, and then click Reboot
Lab 2: Experiencing F1 Acceleration

Step 1

X86 CPU

libx265

SW Encoder

Raw YUV Video

HEVC Encoded Video

Step 2

X86 CPU

FFmpeg

Custom Plugin

OpenCL API

Raw YUV Video

HEVC Encoder

F1 FPGA
Lab 3:

- Understanding software and hardware emulation flows
  - Kernel calls
  - Application timeline
- Profiling performance and identifying how to optimize host
- Kernel code
  - Inverse Discrete Cosine Transform (IDCT)
- Kernel optimization
  - HLS reports
- Use pre-compiled AFI to run on the F1 instance
It’s Time to Wrap-Up (10:15 AM)

➢ Please close RDP, stop instance, and logout
  – Click the ‘X’ icon to close your RDP client
  – In your local browser, return to the tab showing the EC2 Console
  – In the EC2 Console, make sure you have selected your instance
  – Click the Actions button, select Instance State, and then click Stop
  – Make sure to Sign Out

➢ PLEASE RETURN THE EVALUATION FORM
Summary

- F1-optimized HEVC encoder showed acceleration over pure software driven ffmpeg encoder
- SDAccel environment enabled application profiling and optimization
- Hardware acceleration on VU9P FPGA using pre-built AFI
- Labs conducted using pre-generated AMI and pre-configured EC2 F1 instance
Create Your Amazon Web Services (AWS) Account

https://aws.amazon.com/
Want Support to Continue to Work on AWS F1?

- Submit academic research proposal that justifies F1 usage
  - Please submit a one page proposal by March 12, 2018 to parimal.patel@Xilinx.com with subject “Research Proposal for AWS F1 Work” stating what you propose to work on
- A team will evaluate your proposal
- Ten proposals will be selected and each will be awarded $250 research credit (~150 hours of compute time on F1.2x or ~625 hours on C4.2x)
- Outstanding follow on research may entitle additional research credit
Thank You