ROCm on AMD GPUs in Fedora

Tim Flink

AMD
together we advance_

Outline

- Introduction
- ROCm in Fedora
- Will my GPU work with ROCm?
- Less Traditional Accelerators
- Questions

Introduction

Who am I and why am I talking about this?

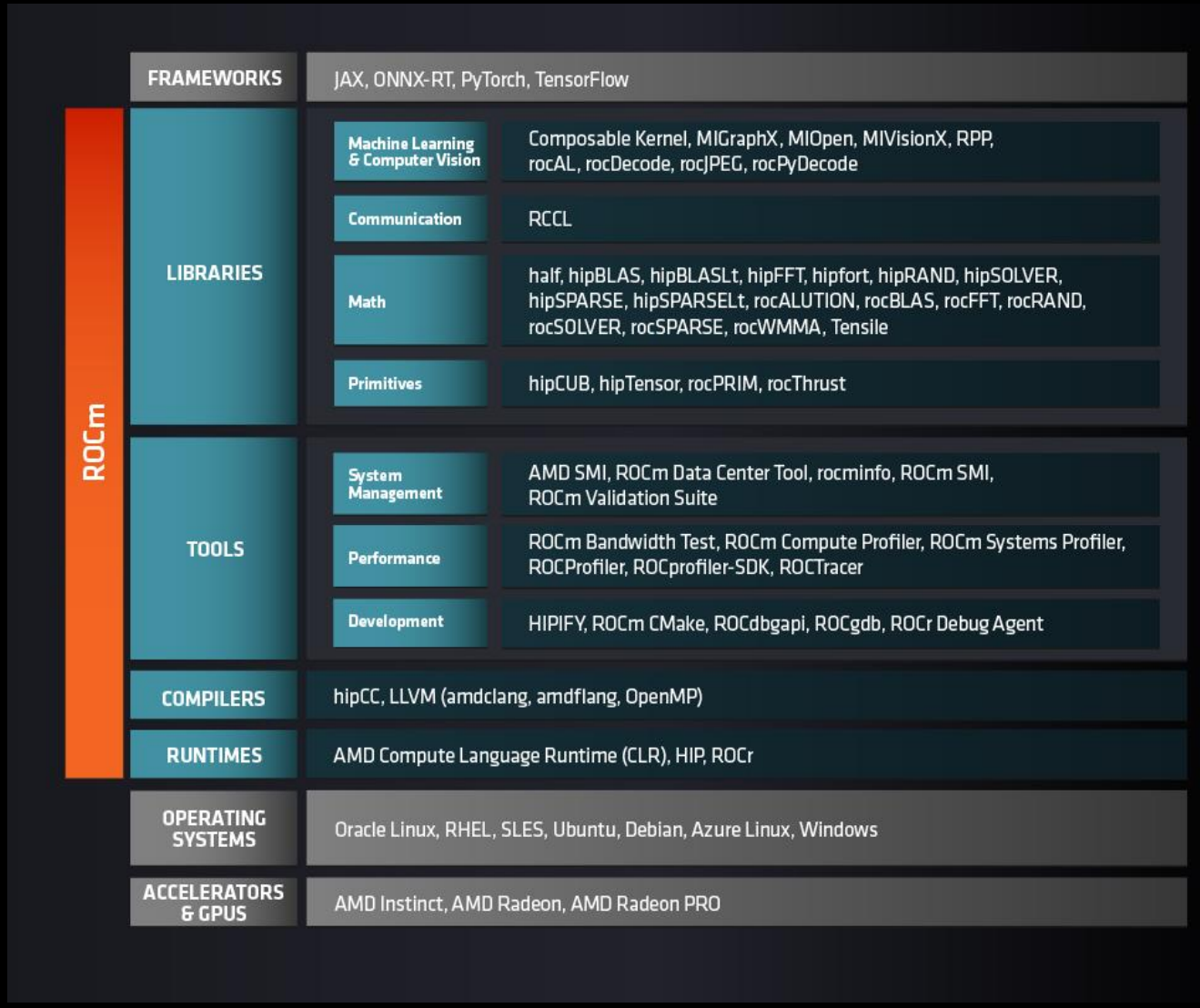
- Worked for Red Hat for over 14 years, most of that was as part of Fedora Quality
- Got into helping get open source AI/ML tooling into Fedora after grad school
- Founding member of the ROCm packaging SIG
- Started working for AMD in 2025 to improve ROCm's presence in upstream Linux distros

What is ROCm?

"ROCm is a software stack, composed primarily of open-source software, that provides the tools for programming AMD Graphics Processing Units (GPUs), from low-level kernels to high-level end-user applications."

- Official ROCm Documentation

What is ROCm?



ROCm in Fedora

What is ROCm In Box?

- "In Box" is what AMD calls efforts to have software available directly from the distro vendor's repositories
  - To have a good out-of-box experience, all of the bits need to be "in box"
  - Easily enabled repos like EPEL are also included if there is no other option
- Examples of AMD In Box
  - Upstream amdgpu kernel module
  - Mesa packages
  - ROCm packages

What is AMD doing for ROCm in Fedora?

- 3 of the 4 Fedora ROCm packagers are AMD employees working on Fedora as part of their job
  - Members of the rocm-packagers-sig dist-git group have commit access to all the ROCm packages
- In practice, almost all of the ROCm package updates in Fedora are done by AMD employees
- Early phases of internal testing for new ROCm package releases
  - Will help with overall quality but won't be directly visible by end users

How to find Fedora's ROCm packagers?

- Fedora ai/ml room on matrix
- #ai-ml-sig tag on Discourse
- rocm-packagers-sig@lists.fedoraproject.org

How much of ROCm is in Fedora?

- Almost the entire ROCm stack is available in Fedora
- Still missing a few "long tail" packages that will be packaged
  - composable_kernel
- Some less common bits have not been packaged due to licensing issues, unsuitability for distribution as a package in an upstream distro or just low perceived value to the Fedora community
  - ROCm Validation Suite, ROCm gdb, rocprofiler
  - These may be packaged if their issues are resolved or there is need for them

What can I do with ROCm?

- Applications In Fedora
  - ollama
  - Blender
- Libraries in Fedora
  - HW accelerated math libraries
    - BLAS, FFT, etc.
  - Computer vision
    - MIVisionX
  - AI/ML
    - pytorch
    - onnx

Will ROCm work with my GPU?

What does "supported" mean?

- In Box packages are a community effort at AMD
  - Any support is "best effort"
  - Packagers will help as they can but there are no contractual promises around fixes, speed of updates, etc.
- Any official support from AMD requires using the officially supported builds
  - See official documentation for ROCm

Will ROCm work with my GPU?

- ROCm is built for specific, individual GPU targets
  - There is no general target at this time
- "Is compiled for" is not always the same as "will work without any trouble"
- It's better to break this down into multiple questions
  - What is officially supported in ROCm upstream?
  - What targets are Fedora packages built for?
  - What targets see the most testing in Fedora?

What is officially supported upstream?

- The list of officially supported accelerators is available in the ROCm Documentation
  - https://rocm.docs.amd.com/projects/install-on-linux/en/latest/reference/system-requirements.html
- Non-Datacenter accelerators:
  - gfx1100
    - Radeon RX 7900 XTX
    - Radeon RX 7900 XT
    - Radeon Pro W7900 (DS)
    - Radeon Pro W7800 (48GB)
  - gfx1030
    - Radeon Pro W6800

Cards vs. ISAs

- Accelerators are commonly referred to by the name of the board
  - Radeon 7900 XTX et al.
- ROCm components will categorize accelerators by their Instruction Set Architecture (ISA)
  - gfx1201
    - AMD Radeon 9070 XT
  - gfx1103
    - AMD Radeon 780M
  - gfx1100
    - AMD Radeon 7900 XTX, AMD Radeon Pro W7900
  - gfx942
    - AMD Instinct MI300X
- Similar numbers for ISA generally imply similar cards
  - gfx1100 (Radeon Pro W7900) and gfx1101 (Radeon Pro W7700) are very similar
  - gfx1100 and gfx1103 (Radeon 780M) are less similar but there is still a lot of overlap
  - gfx1100 and gfx906 (Radeon Pro VII) are not very similar

What ISA is my GPU?

- rocminfo from the rocminfo package will show you a GPU's ISA

```
$ rocminfo

# trimmed for space

*******
Agent 2
*******
  Name:                    gfx1101
  Uuid:                    GPU-f432de0f0fab9c1f
  Marketing Name:          AMD Radeon Pro W7700
  Vendor Name:             AMD
```

What ISAs are Fedora packages built for?

- The default ISA list is controlled by the rocm-rpm-macros package
  - rocm.macros in rawhide is the most up-to-date-list
- A few packages are more restricted
  - Example: there are a limited number of ISAs that hipblaslt can be built for
- The current default list is:
  - `gfx900;gfx906:gfx908;gfx90a;gfx942;gfx1010;gfx1012;gfx1030;gfx1031;gfx1035;gfx1100;gfx1101;gfx1102;gfx1103;gfx1150;gfx1151;gfx1152;gfx1200;gfx1201`

Which ISAs see the most testing in Fedora?

- The Fedora ROCm packagers do much of the Fedora-specific testing and focus on desktop GPUs
- gfx1100 is currently considered to be the "reference" target for Fedora's ROCm packages
  - AMD Radeon 7900 XT/XTX, AMD Radeon Pro W7900
- gfx1201 will likely become the "reference" target for Fedora's ROCm packages once it is officially supported upstream in a future ROCm release
  - AMD Radeon 9070 XT

Related: What AMD GPU should I get for ROCm?

- If you already have an AMD GPU, try it to see if it works with Fedora's ROCm packages
- If you want the most seamless experience, choose something that is already supported upstream
- If the officially supported GPUs are not an option, try to get as close as possible
  - The ISA for any given card is available publicly on sites like TechPowerUp

Less Traditional Accelerators

Less Traditional Accelerators

- Traditionally, HC has been the domain of workstations, servers and large clusters using discrete GPUs
- Targeting smaller accelerators has been a recent development that is still in progress
  - Mobile GPUs
  - NPUs

Mobile GPUs

- Integrated graphics have become much more powerful
- No mobile GPUs are officially supported for ROCm
  - In practice, things generally work for newer ISAs and are getting better
- Fedora's ROCm packages are compiled for some recent mobile GPUs
  - gfx1103, gfx1150, gfx1151
  - AMD Radeon 740M, AMD Radeon 890M, AMD Radeon 8060S
- Please contact the Fedora ROCm maintainers or file bugs if mobile GPUs do not work

NPUs

- Neural Processing Units (NPUs) are a relatively new arrival
- AMD NPUs use a different driver than AMD GPUs and have a different software stack that is not ROCm
- NPU support is still under rapid development and things are changing quickly

Wrapping Up

Are you interested in helping?

- Help us test GPUs that aren't the primary focus
  - The Fedora packagers have limited amounts of time and hardware
- File bugs if something doesn't work
- Ask questions if something doesn't make sense

Questions?

Copyright and disclaimer

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. The information contained herein is subject to change and may be rendered inaccurate releases, for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. Any computer system has risks of security vulnerabilities that cannot be completely prevented or mitigated. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

THIS INFORMATION IS PROVIDED 'AS IS." AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS, OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION. AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY, OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY RELIANCE, DIRECT, INDIRECT, SPECIAL, OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION


AMD