

GAME OPTIMIZATION WITH THE AMD RADEON™ DEVELOPER TOOL SUITE

CHRIS HESIK (AMD)

AMIT MULAY (AMD)

AMD together we advance_



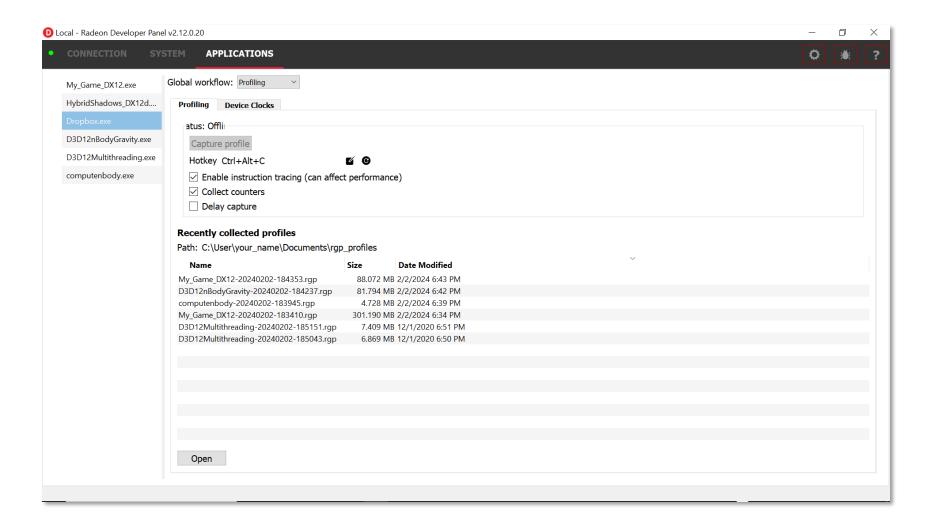
Agenda

GDC 2024

- Overview of the tools and new features
- Introduction to GPU Reshape

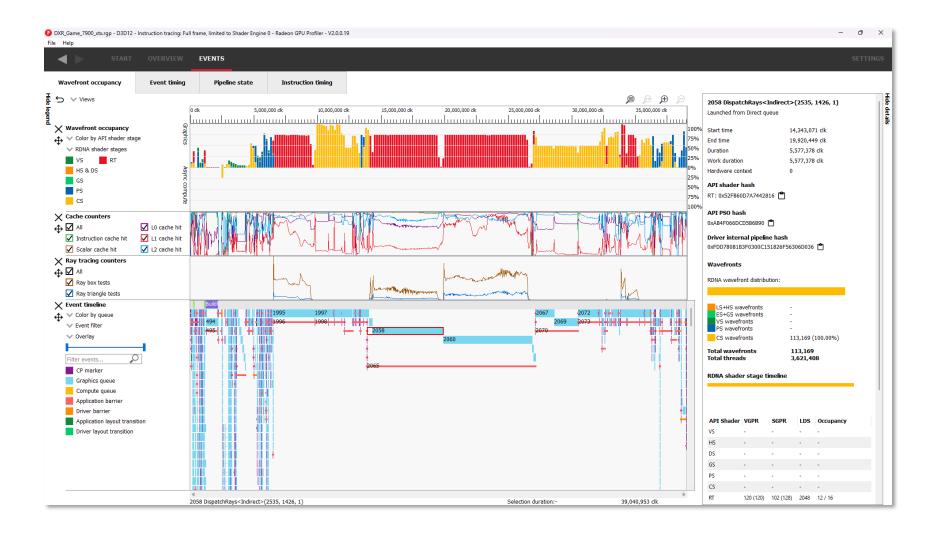


AMD TARADEON Developer Panel





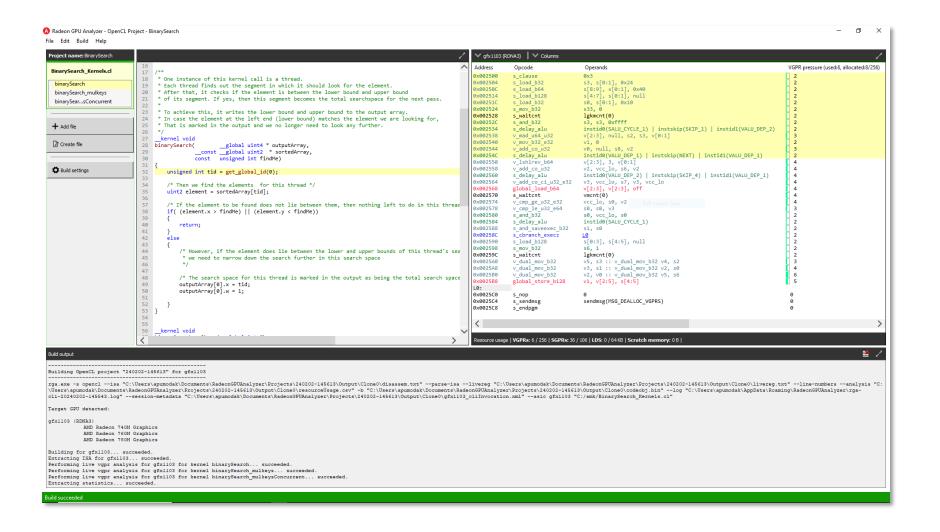
AMD RADEON GPU Profiler





GDC 2024

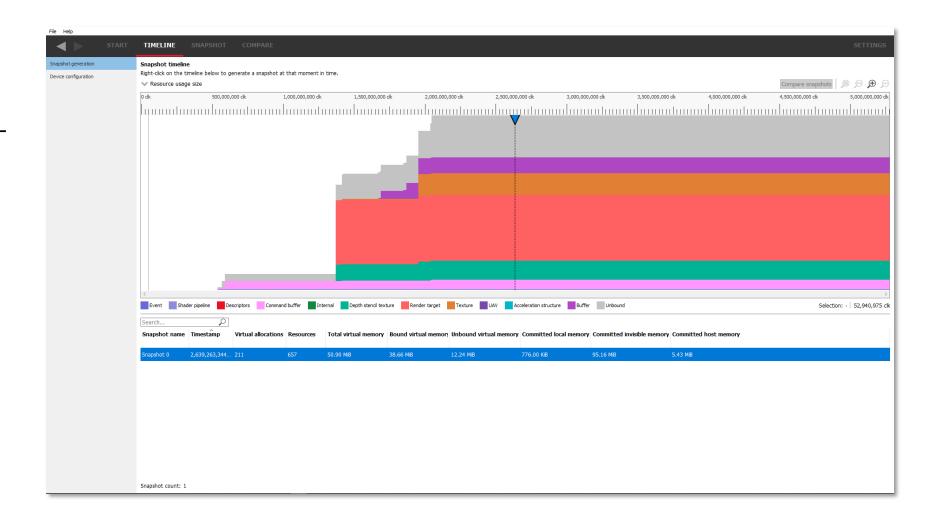
AMD TRADEON GPU Analyzer





GDC 2024

AMDI RADEON Memory Visualizer



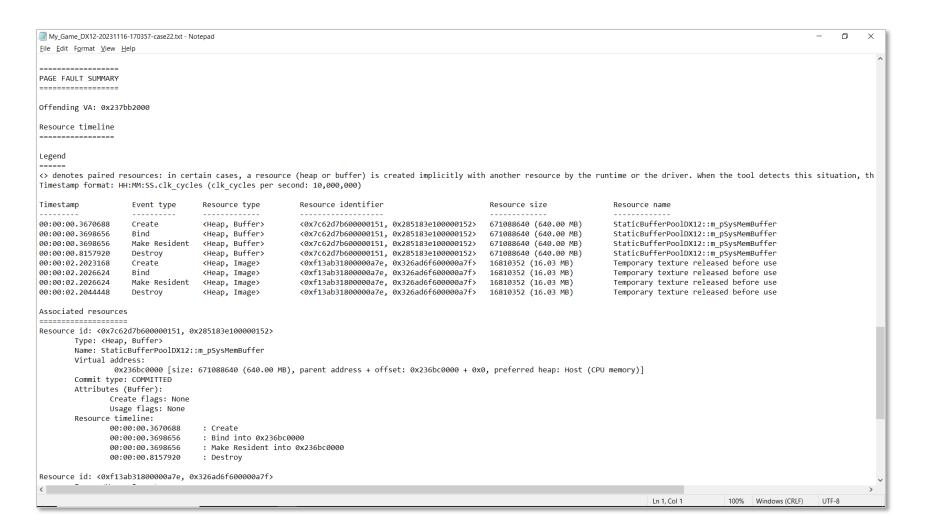


AMDI RADEON Raytracing Analyzer





AMDA RADEON GPU Detective





March 2024

GDC 2024

WHAT'S NEW?



RRA 1.2

UI Improvements to traversal view er, BLAS tab, TLAS view er pane, and traversal mode view er



RMV 1.6 Improved history resource table



GDC 2023

RGP 1.15 Redesigned ISA disassembly view Initial workgraph support

March

April

May

June

July

Aug

Sept

RRA 1.3

Ray Visualization feature Persistent UI state



RGA 2.8 Support for AMD Radeon™ RX 7800/7700 XT Resource name and implicit buffer fixes VGPR pressure GUI AMD offline LLVM pipeline

compiler



RGP 2.0 Redesigned Wavefront occupancy UI Dark mode support Raytracing pipeline thread divergence

mmand Buffer ID: 0x107c

⊢[x] ├[x] ├[x]



RRA 1.4 Ray Direction Visualization



RGD 1.1

Vulkan® Support

RMV 1.8 Improvements to Resource Usage size timeline



RGD 1.0 DirectX® 12 Support

RGP 1.16

ISA disassembly view improvements

Quality of life improvements

RMV 1.7

Aliased Resource improvements

Loading of RGD crash dump files



RGA 2.9 Analyze pre-compiled Code Object binaries. Analyze pre-compiled HIP binaries for the MI-200 architecture

GDC 2024

Oct Nov Dec



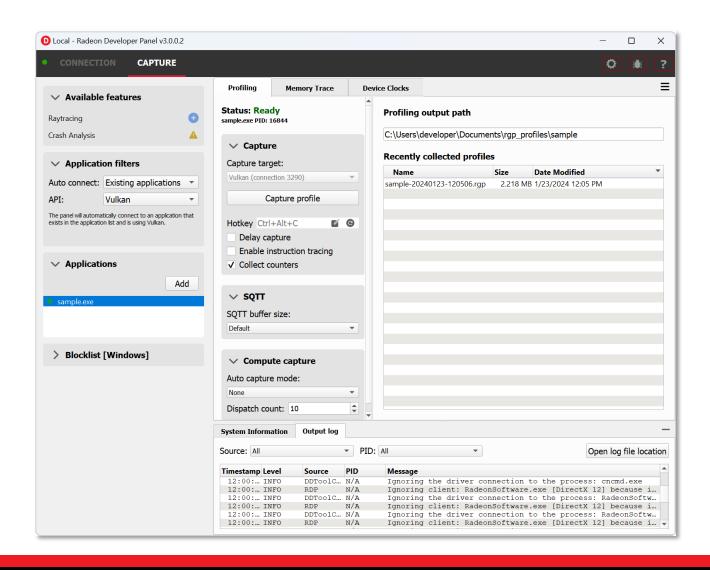




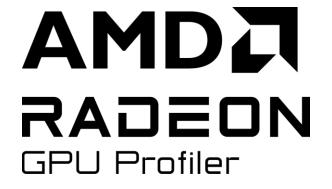
PREVIEW: AMD RADEON DEVELOPER PANEL 3.0

Redesigned user interface offers

- Improved experience for new users
- Simplified user workflows for setting capture options
- Persistence of all settings across invocations of the panel
- Support for new features without increasing complexity





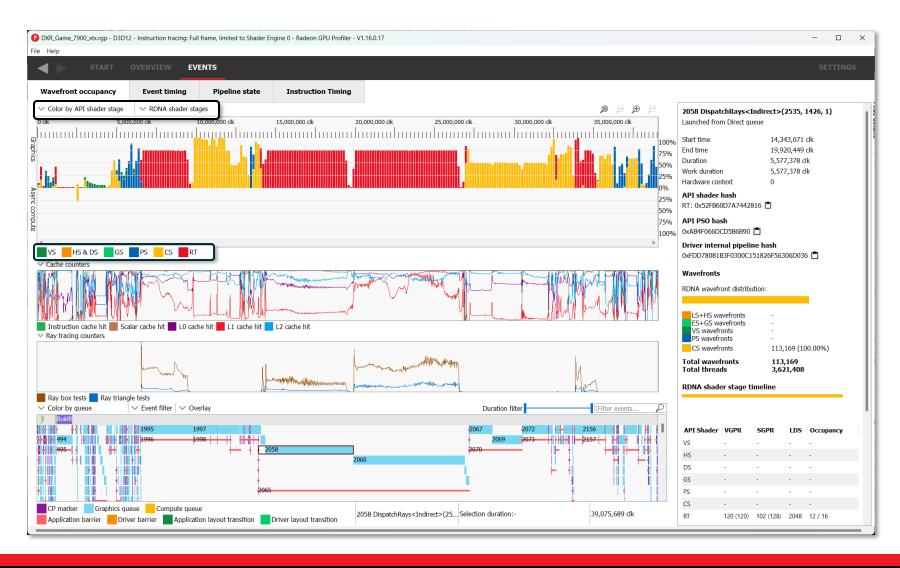




RGP 1.16 WAVEFRONT OCCUPANCY LAYOUT

Previous layout

- UI controls above each row
- Legends below each row

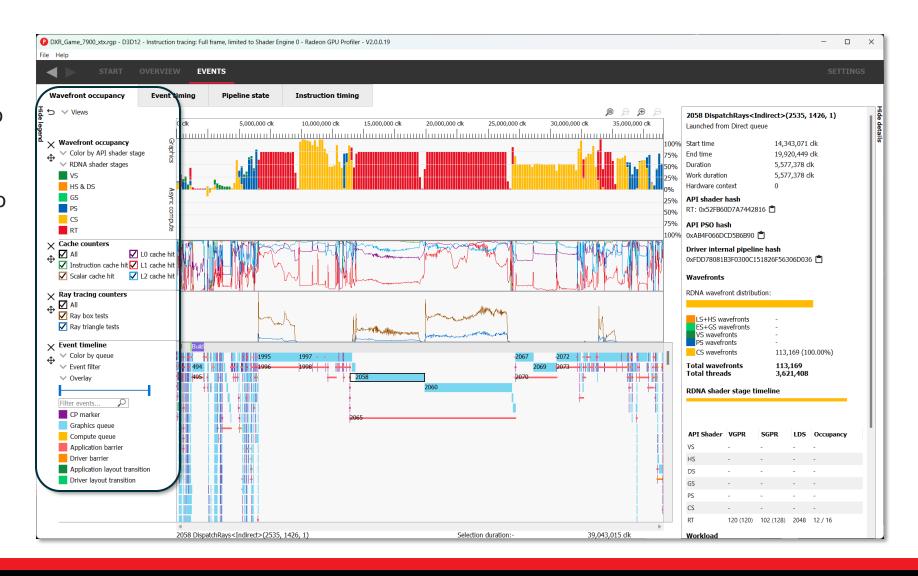




RGP 2.0 WAVEFRONT OCCUPANCY LAYOUT

New layout

- UI controls and Legends are moved to the left of each row
- More vertical screen real estate allocated to the data views

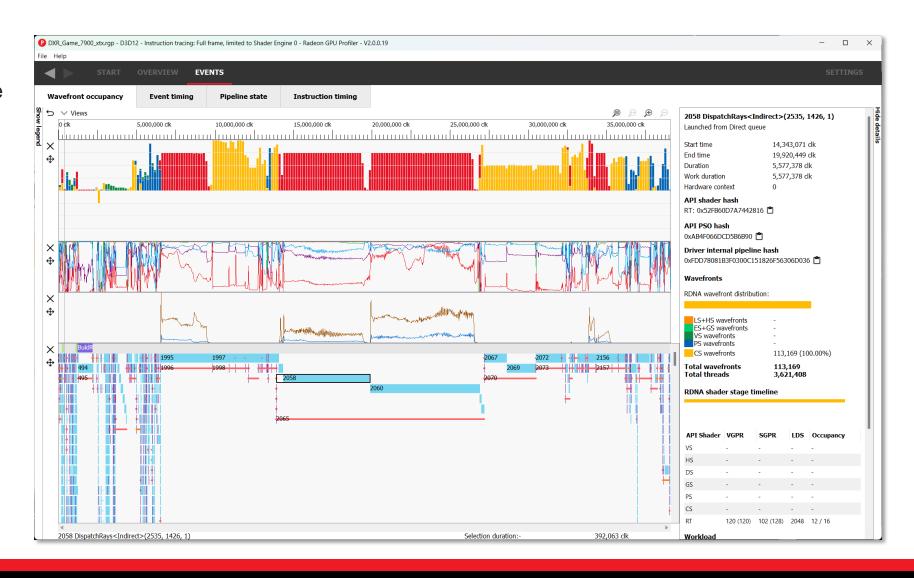




RGP 2.0 WAVEFRONT OCCUPANCY LAYOUT

New layout

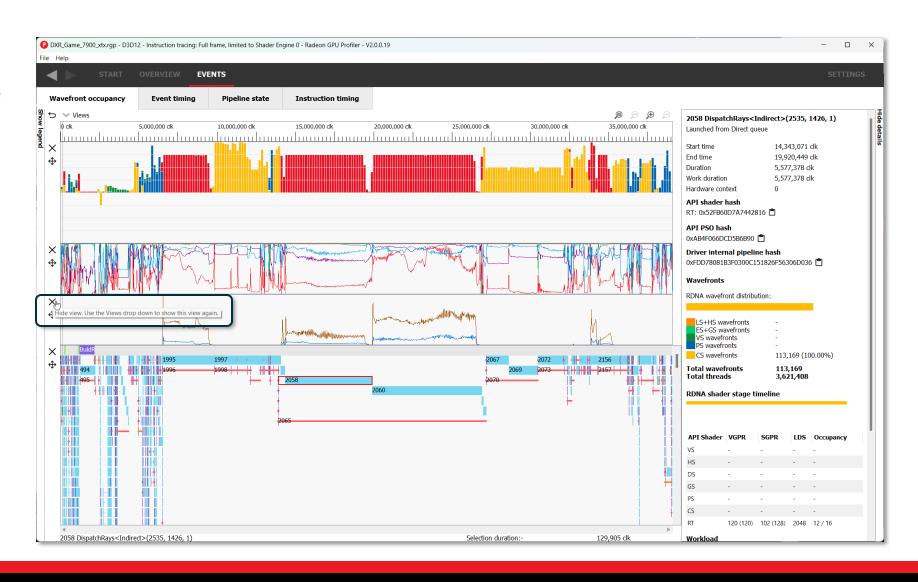
 The new left hand side panel can be hidden





New layout

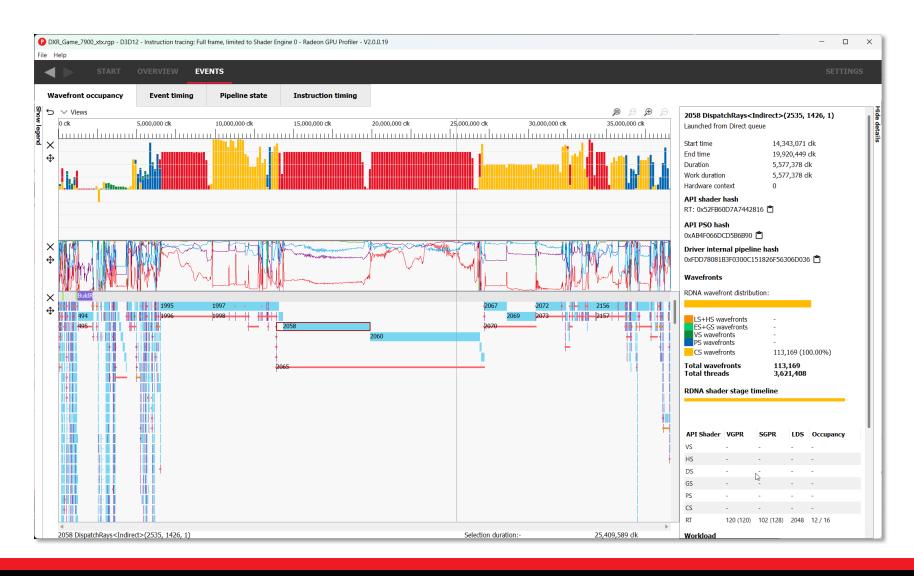
 Individual rows can be hidden





New layout

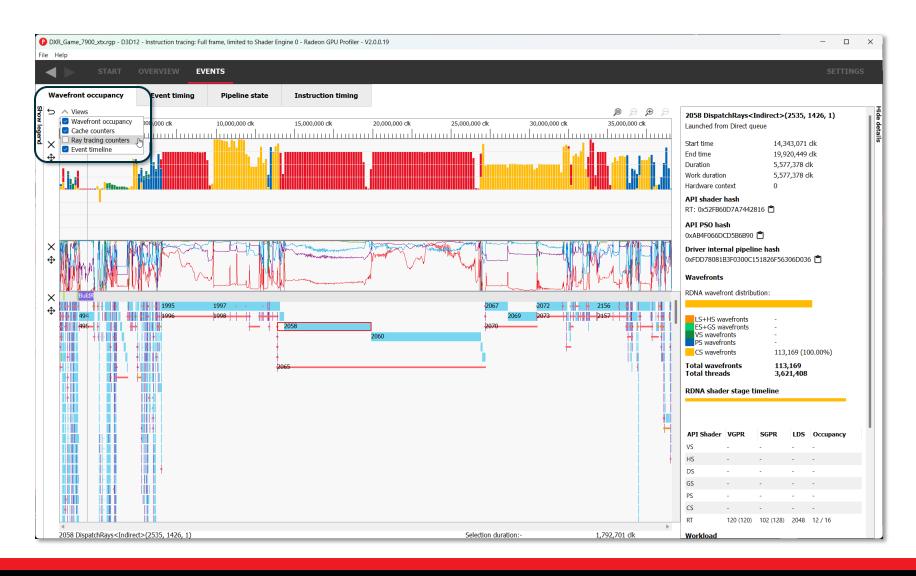
- Individual rows can be hidden
- Raytracing counters hidden





New layout

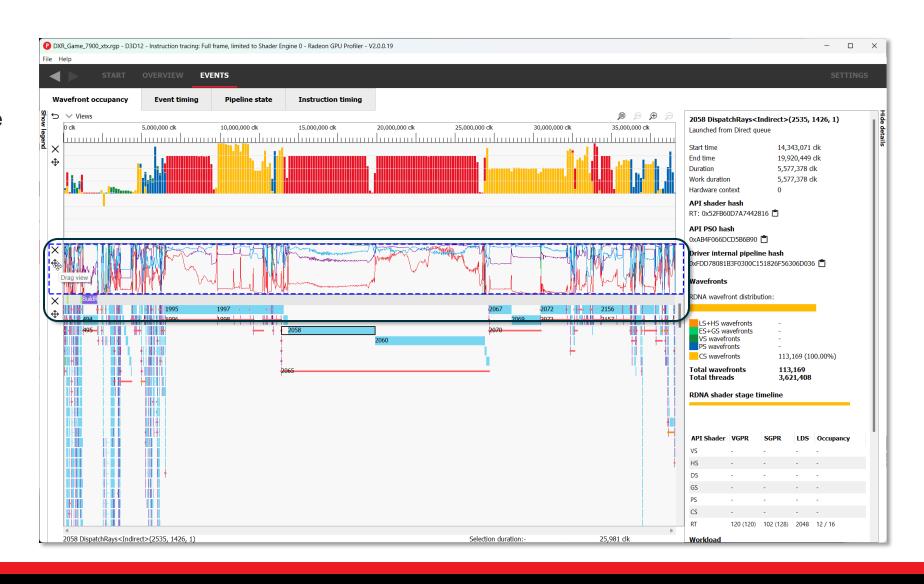
- Individual rows can be hidden
- Raytracing counters hidden
- Hidden rows can be shown again





New layout

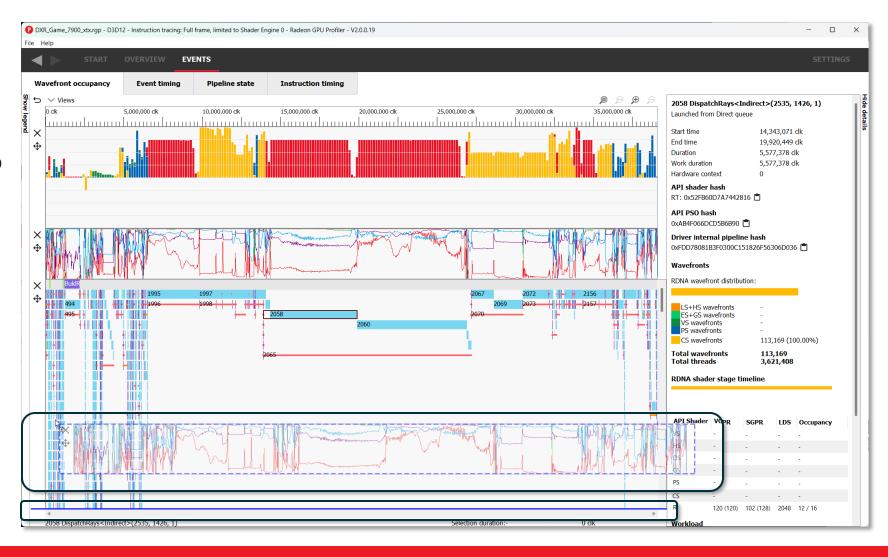
 The position of individual rows can be changed





New layout

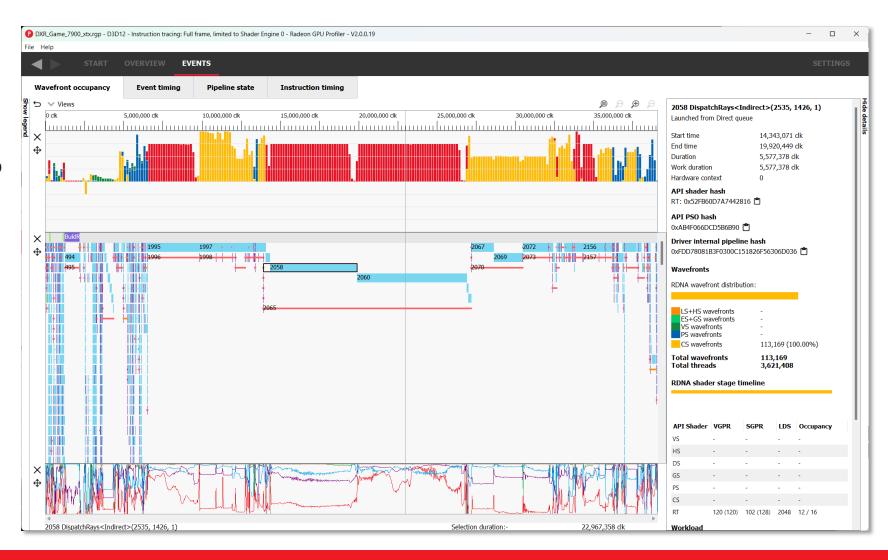
- The position of individual rows can be changed
- Click and drag a view to reposition it





New layout

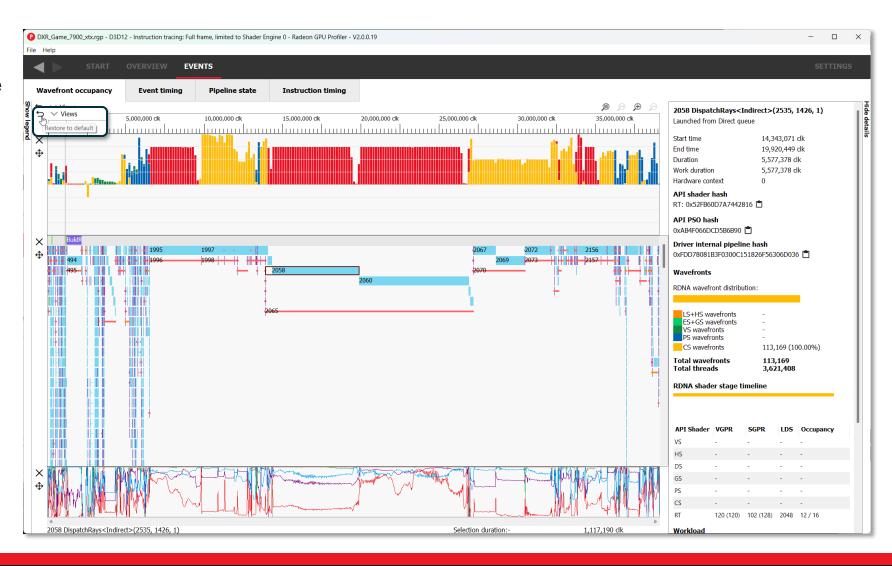
- The position of individual rows can be changed
- Click and drag a view to reposition it
- Drop it in the new position





New layout

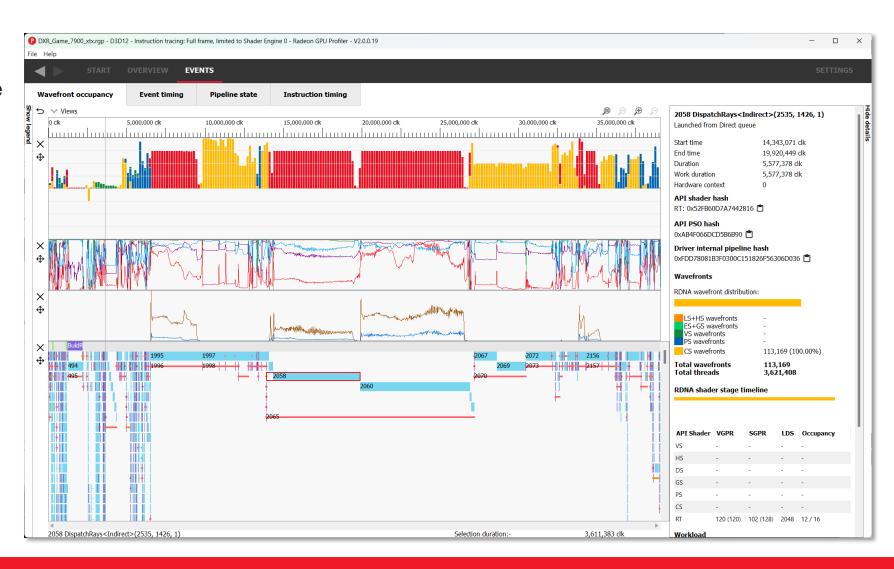
The default view can be restored





New layout

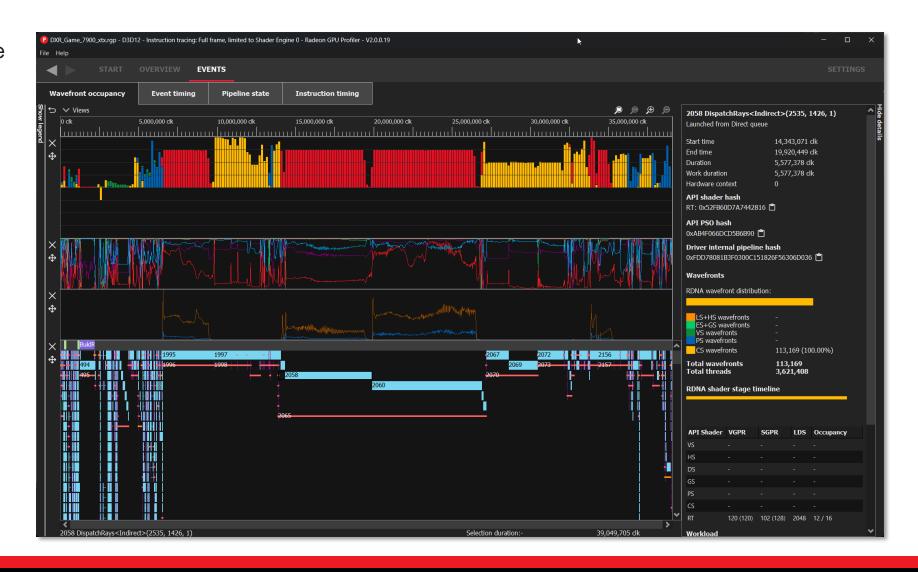
The default view can be restored





RGP – DARK MODE

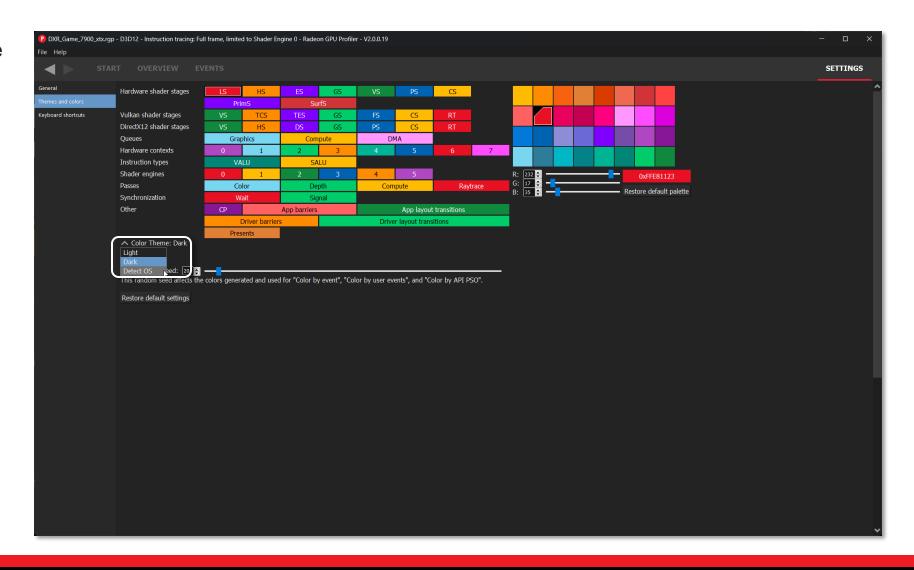
- RGP can be set to use Dark Mode or Light mode
- Or it can be set to follow the host operating system





RGP – DARK MODE

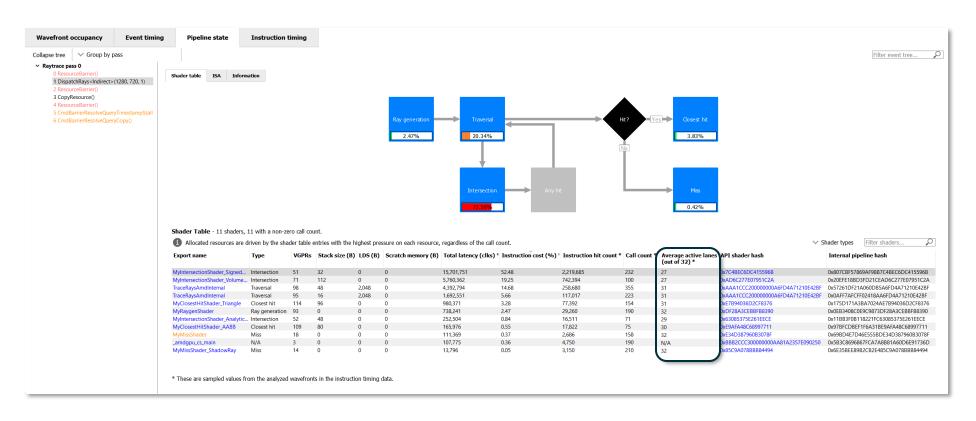
- RGP can be set to use Dark Mode or Light mode
- Or it can be set to follow the host operating system
- Use the new "Color Theme" setting on the "Themes and Colors" page





RGP: RAY TRACING THREAD DIVERGENCE

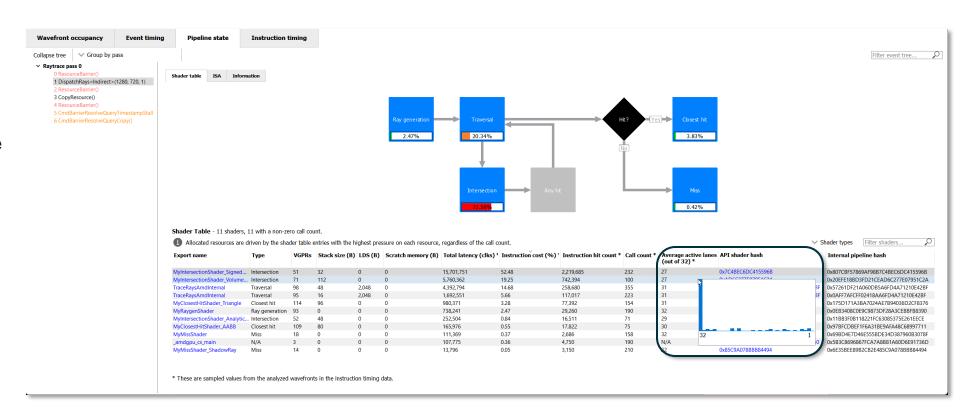
- Learn about thread divergence in your ray tracing pipelines
- RGP reports the average number of active lanes upon entry of each function in the ray tracing pipeline





RGP: RAY TRACING THREAD DIVERGENCE

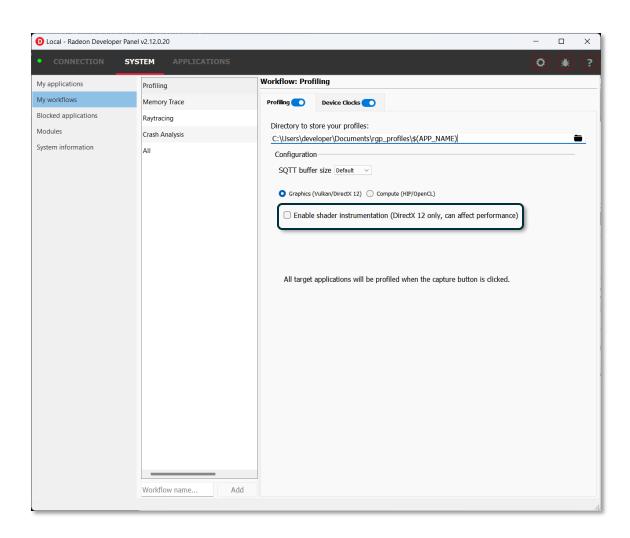
- Learn about thread divergence in your ray tracing pipelines
- New column reports the average number of active lanes upon entry of each function in the ray tracing pipeline
- Histogram shows the distribution across all invocations





RGP: RAY TRACING THREAD DIVERGENCE – RDP SUPPORT

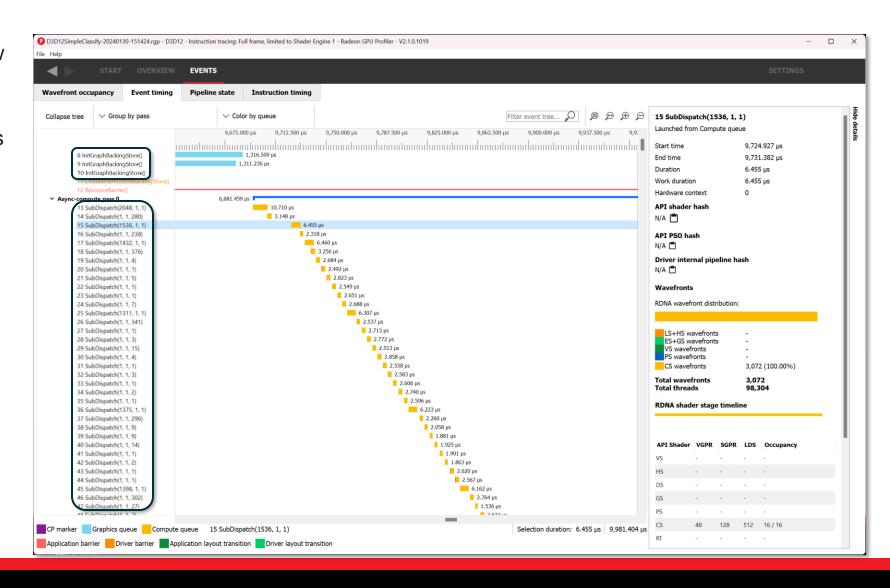
- "Enable shader instrumentation" checkbox in RDP
- May add extra overheard which can affect runtime performance





RGP: WORK GRAPHS

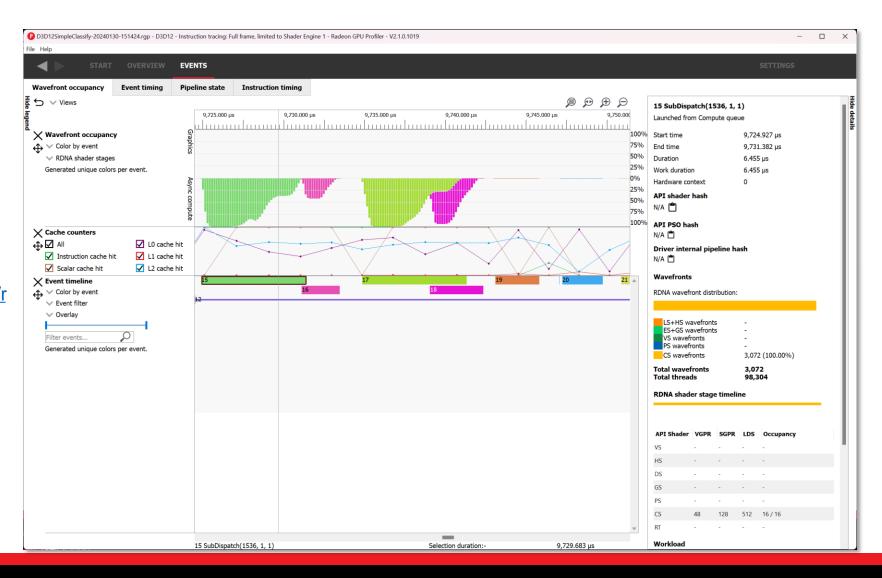
- RGP's event lists show individual subdispatches
- Shows how the work is broken down during graph execution



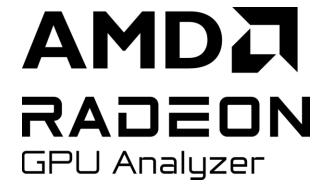


RGP: WORK GRAPHS

- RGP's event lists show individual sub-dispatches
- Shows how the work is broken down during graph execution
- Coloring by event shows which waves come from which graph subdispatches
- More info on GPUOpen: https://gpuopen.com/learn/rgp-work-graphs/





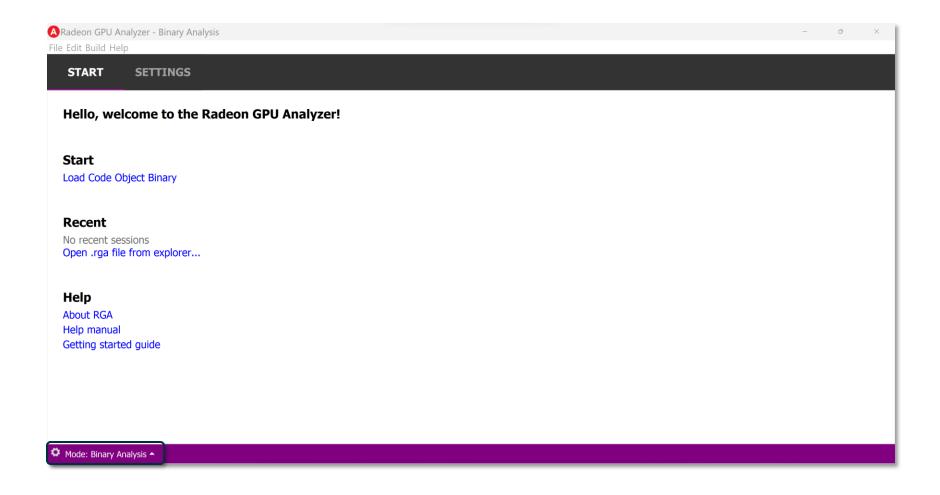




AMD RADEON™ GPU ANALYZER

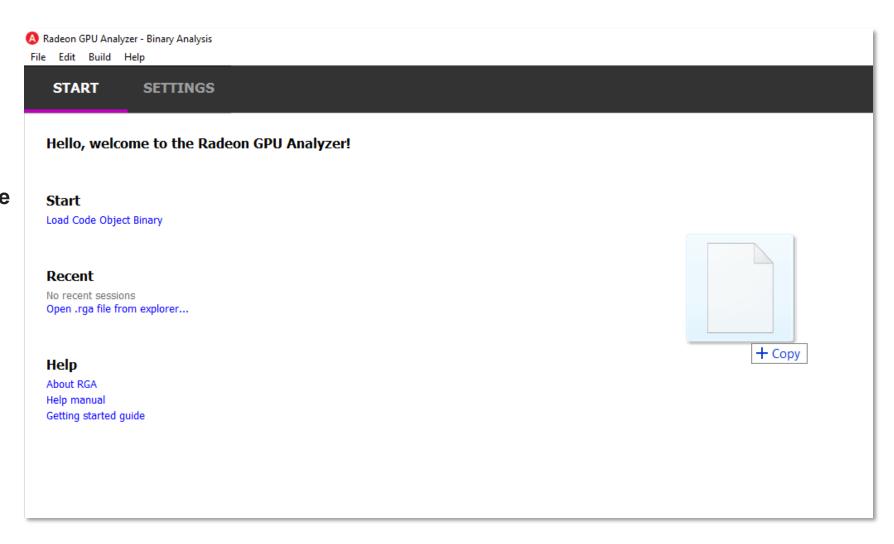
New mode

Binary Analysis



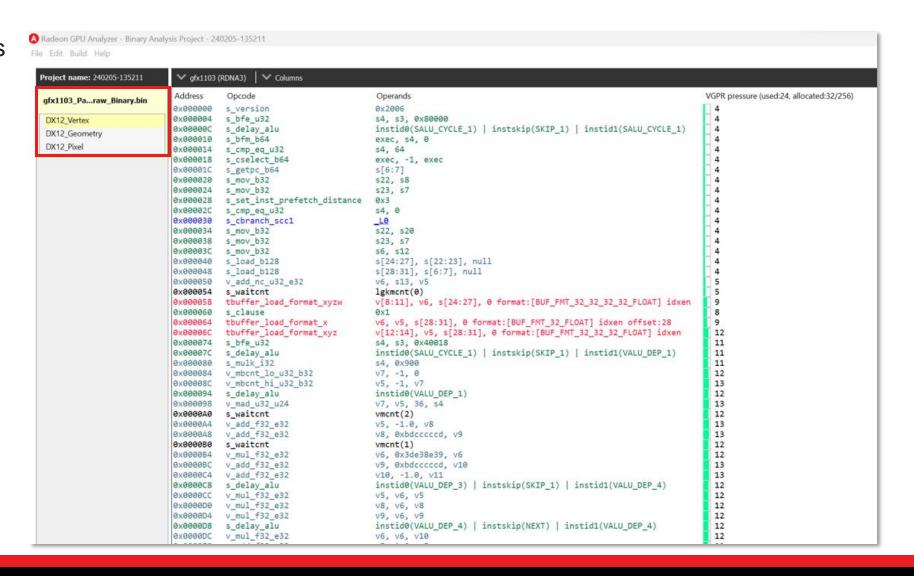


- Drag & drop any precompiled AMD GPU Code Object binary
- Unlike other RGA modes, here we start with a pre-compiled Code Object binary file



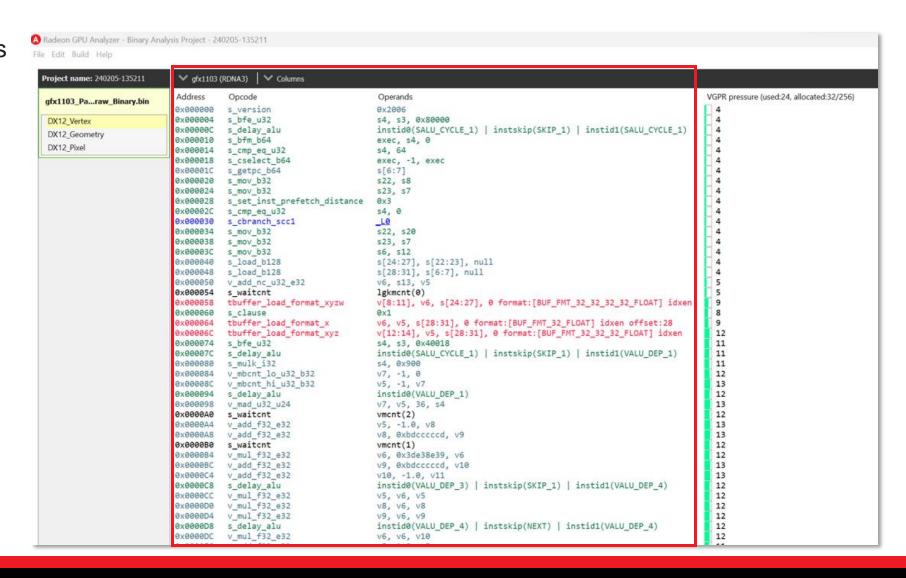


 Shaders and kernels appear on the left pane



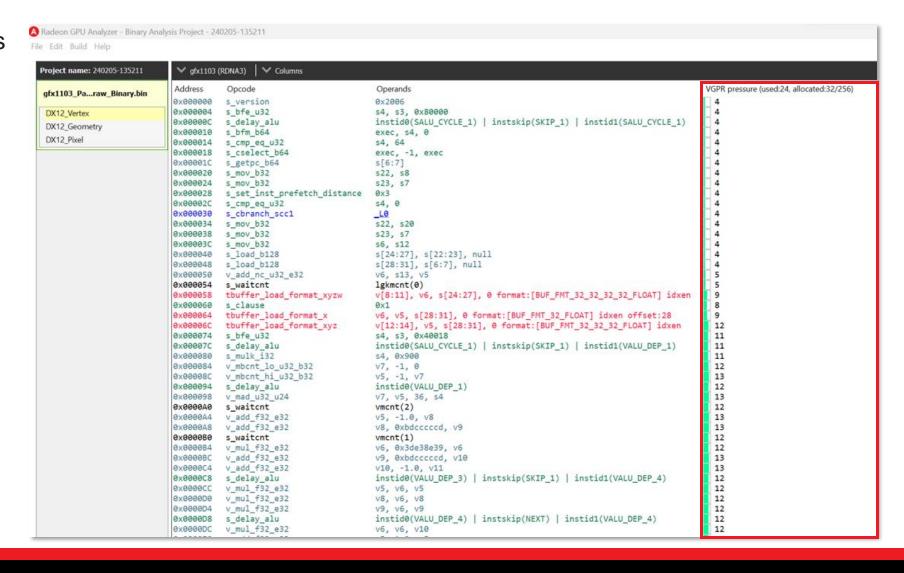


- Shaders and kernels appear on the left pane
- Binary gets disassembled





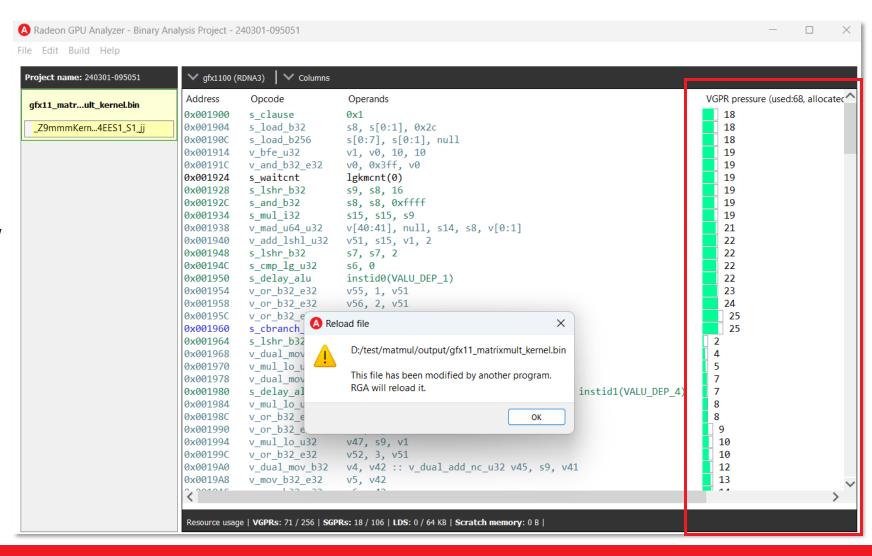
- Shaders and kernels appear on the left pane
- Binary gets disassembled
- VGPR pressure visualized





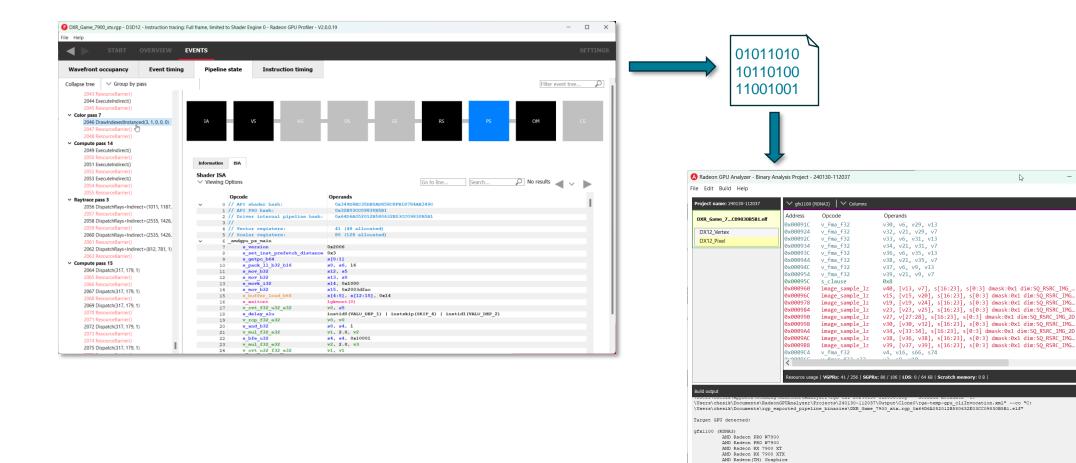
RGA BINARY ANALYSIS MODE

- RGA will detect modified binaries and reload the contents automatically
- You can continue to use your normal workflows to edit and compile binaries – RGA will always show the latest updates





PREVIEW: RGP/RGA INTEROP





AMD PUBLIC | GDC 2024 | AMD GPU DEVELOPER TOOLS | March 2024 38

Extracting ISA for gfx1100... succeeded. Performing live vgpr analysis for gfx1100...

succeeded.

- D X

VGPR pressu

33

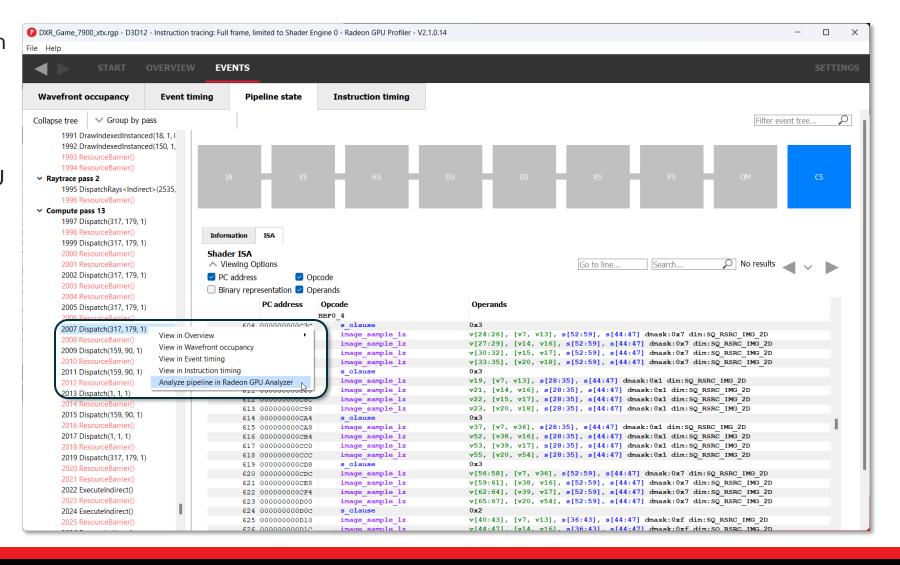
35

35 34

33

PREVIEW: RGP/RGA INTEROP

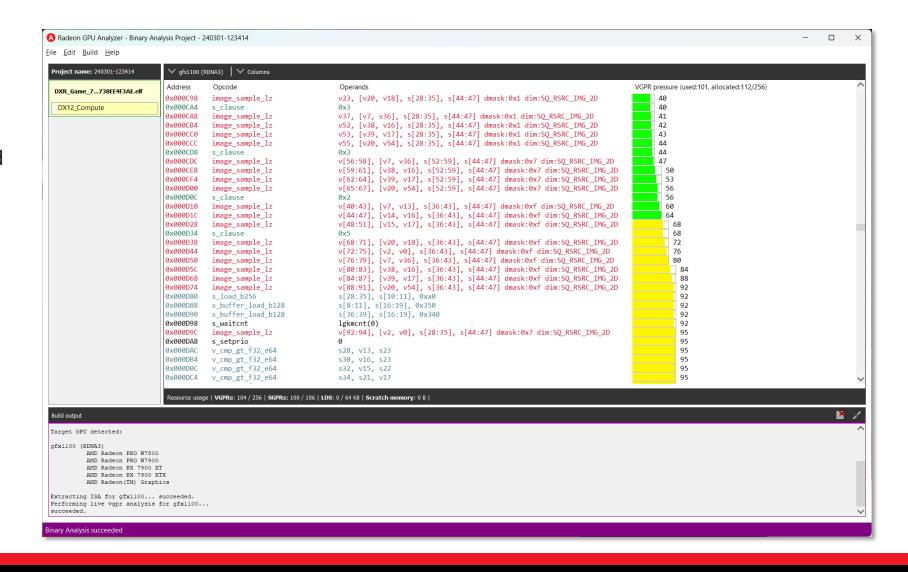
- Any pipeline available in RGP can be automatically extracted and analyzed in RGA
- Use the new "Analyze pipeline in Radeon GPU Analyzer" menu item





PREVIEW: RGP/RGA INTEROP

- The selected pipeline will be extracted
- RGA will be launched. and the extracted pipeline will be loaded and analyzed





- Currently RGA works with complete pipelines
 - For graphics, that means you would need at least the following:
 - Vertex shader
 - Pixel shader
 - Graphics pipeline state object
 - Root signature
- What if you don't have a complete pipeline?
 - You are missing one of the shaders
 - Missing the graphics pipeline state object
 - Missing the root signature
- RGA will have a new mode that can autogenerate the missing pieces of the pipeline!



- Let's consider an example:
- You have a single pixel shader, but do not have a vertex shader, a root signature or a graphics pipeline state object

```
...

≡ classic_ps.hlsl X

      C: > RGA-2.9.1 > dxc > single_shaders > ≡ classic_ps.hlsl
             struct VsOutput
                 float4 pos : SV Position;
ည
                 float2 tex coord : TEXCOORD0;
             };
             Texture2D<float4> texture0 : register(t0);
             SamplerState sampler0 : register(s0);
品
             float4 PsMain(VsOutput i) : SV_Target
        11
        12
                 return texture0.Sample(sampler0, i.tex coord);
        13
```



 Compiling as usual in DX12 mode, providing only the pixel shader as an input

```
Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows
PS C:\RGA-2.9.1> .\rqa.exe -s dx12 --autogen-dir "C:\RGA-2.9.1\Generated" -c gfx1100 --all-model 6_0 --ps-entry "PsMain"
 --isa "C:\RGA-2.9.1\Isa\Out.isa" --ps "dxc\single_shaders\classic_ps.hlsl"
Building for gfx1100...
Auto-generating root signature using reflection into C:\RGA-2.9.1\Generated ... success.
Auto-generating graphics pipeline state using reflection into C:\RGA-2.9.1\Generated ... success.
Auto-generating vertex shader using reflection into C:\RGA-2.9.1\Generated ... success.
Performing front-end compilation of vertex shader through DXC...
Front-end compilation success.
Performing front-end compilation of pixel shader through DXC...
Front-end compilation success.
Compiling graphics pipeline...
Compiling root signature defined in HLSL file C:\RGA-2.9.1\Generated\rga_autogen_20240131_105240_.rootsig in macro named
 RGA_ROOT_SIGNATURE...
Compiling root signature defined in HLSL file C:\RGA-2.9.1\Generated\rga_autogen_20240131_105240_.rootsig in macro named
RGA_ROOT_SIGNATURE...
Extracting vertex shader disassembly...
vertex shader disassembly extracted successfully.
Extracting pixel shader disassembly...
pixel shader disassembly extracted successfully.
succeeded.
PS C:\RGA-2.9.1>
```



 The vertex shader is auto generated

```
a_autogen_20240305_111334_vs.hlsl ×
ga_autogen_20240305_111334_vs.hlsl
   // Auto-generated with Radeon GPU Analyzer (RGA).
   struct VsInput
       float4 attribute0: POSITION0;
   };
   struct VsOutput
       float4 attribute0: SV_POSITION;
       float2 attribute1: TEXCOORD0;
   };
   void main(VsInput input, out VsOutput output)
       float4 result = float4(0.0, 0.0, 0.0, 0.0);
       result += float4(input.attribute0.xyzw);
       output.attribute0 = float4(result.xyzw);
       output.attribute1 = float2(result.xy);
```

44



 The root signature is auto generated

```
≡ rga autogen 20240131_105240_.rootsig ×
// Auto-generated with Radeon GPU Analyzer (RGA).
      #define RGA_ROOT_SIGNATURE \
          "RootFlags( ALLOW_INPUT_ASSEMBLER_INPUT_LAYOUT | DENY_HULL_SHADER_ROOT_ACCESS "
          "| DENY_DOMAIN_SHADER_ROOT_ACCESS | DENY_GEOMETRY_SHADER_ROOT_ACCESS ), " \
          "DescriptorTable(Sampler(s0), visibility=SHADER_VISIBILITY_PIXEL), " \
          "DescriptorTable(SRV(t0), visibility=SHADER VISIBILITY PIXEL)"
```



```
≡ rga_autogen_20240305_111334_.qpso ×
≡ rga_autogen_20240305_111334_.gpso
      # Auto-generated with Radeon GPU Analyzer (RGA).
      # schemaVersion
      1.0
      # InputLayoutNumElements (the number of D3D12_INPUT_ELEMENT_DESC elements in the D3D12_INPUT_LAYOUT_DESC structure - must match the following "InputLayout" section)
      # InputLayout ( {SemanticName, SemanticIndex, Format, InputSlot, AlignedByteOffset, InputSlotClass, InstanceDataStepRate } )
      { "POSITION", 0, DXGI_FORMAT_R32G32B32A32_FLOAT, 0, 0, D3D12_INPUT_CLASSIFICATION_PER_VERTEX_DATA, 0 }
      # PrimitiveTopologyType (the D3D12_PRIMITIVE_TOPOLOGY_TYPE value to be used when creating the PSO)
      D3D12 PRIMITIVE TOPOLOGY TYPE TRIANGLE
      # NumRenderTargets (the number of formats in the upcoming RTVFormats section)
      # RTVFormats (an array of DXGI_FORMAT-typed values for the render target formats - the number of items in the array should match the above NumRenderTargets section)
      { DXGI_FORMAT_R8G8B8A8_UNORM }
```

The graphics pipeline state object is auto generated



GDC 2024

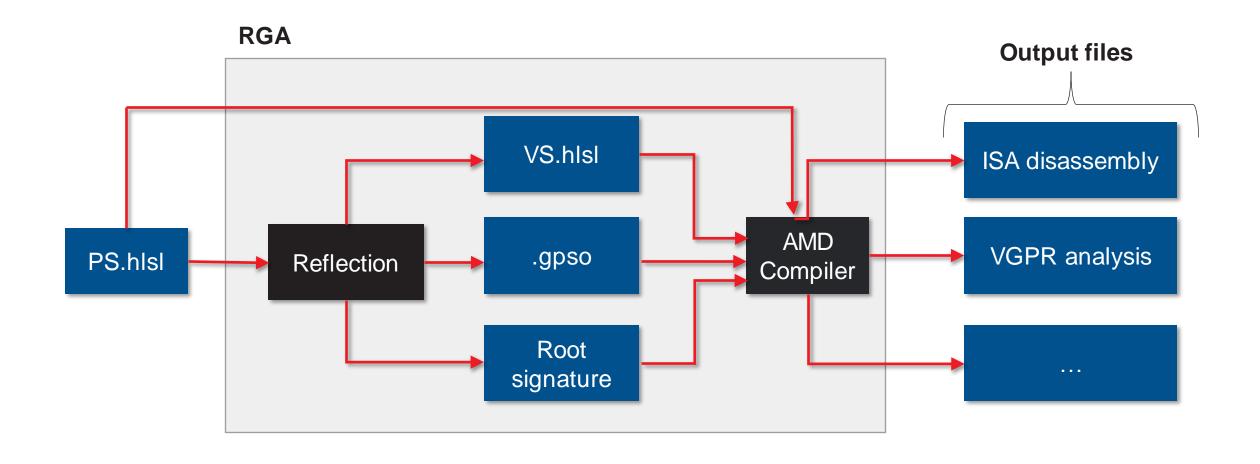
 RGA will output the pixel shader's disassembly

```
≡ qfx1100_Out_pixel.isa ×

    gfx1100_Out_pixel.isa

     : D3D12 Shader Hash 0xa63e3e834124e93ee28366807a0926af
      ; API PSO Hash 0xdd7f0f8f79875091
      ; Driver Internal Pipeline Hash 0x2774d5f37b5943a0
      ; ----- Disassembly -----
      shader main
        asic(GFX11)
       type(PS)
        sgpr_count(14)
        vgpr_count(8)
        wave_size(64)
                                                                // s ps state in s0
                     UC_VERSION_GFX11 | UC_VERSION_W64_BIT // 000000000000: B0802006
        s set inst prefetch distance 0x0003
        s_mov_b32
                     m0, s4
        s_mov_b64
                     s[12:13], exec
        s wqm b64
                     exec, exec
        s_getpc_b64 s[0:1]
        lds_param_load v2, attr0.x wait_vdst:0
        lds_param_load v3, attr0.y wait_vdst:0
        s_mov_b32 s4, s3
        v_interp_p10_f32 v4, v2, v0, v2 wait_exp:1
        v_interp_p10_f32 v0, v3, v0, v3 wait_exp:0
        s_mov_b32 s5, s1
                                                            // 000000000034: BE850001
        s_mov_b32
                    s0, s2
                                                            s delay alu instid0(VALU DEP 2) | instskip(SKIP 1) | instid1(VALU DEP 2) // 00000000003C: BF870122
       v_interp_p2_f32 v2, v2, v1, v4 wait_exp:7
                                                            // 000000000040: CD010702 04120302
       s_load_b256 s[4:11], s[4:5], null
                                                             // 000000000048: F40C0102 F800000
       v_interp_p2_f32 v0, v3, v1, v0 wait_exp:7
        s_load_b128 s[0:3], s[0:1], null
        s_and_b64
                     exec, exec, s[12:13]
                                                            // 0000000000060: 8BFE0C7E
        s waitcnt
                     lgkmcnt(0)
        image_sample v[0:3], [v2,v0], s[4:11], s[0:3] dmask:0xf dim:SQ_RSRC_IMG_2D // 000000000068: F06C0F05 00010002 000000000
                     vmcnt(0)
        s_waitcnt
        v_cvt_pk_rtz_f16_f32 v0, v0, v1
                                                            // 000000000078: 5E000300
        v_cvt_pk_rtz_f16_f32 v2, v2, v3
                     exec, s[12:13]
                     mrt0, v0, v2, off, off done
        exp
```









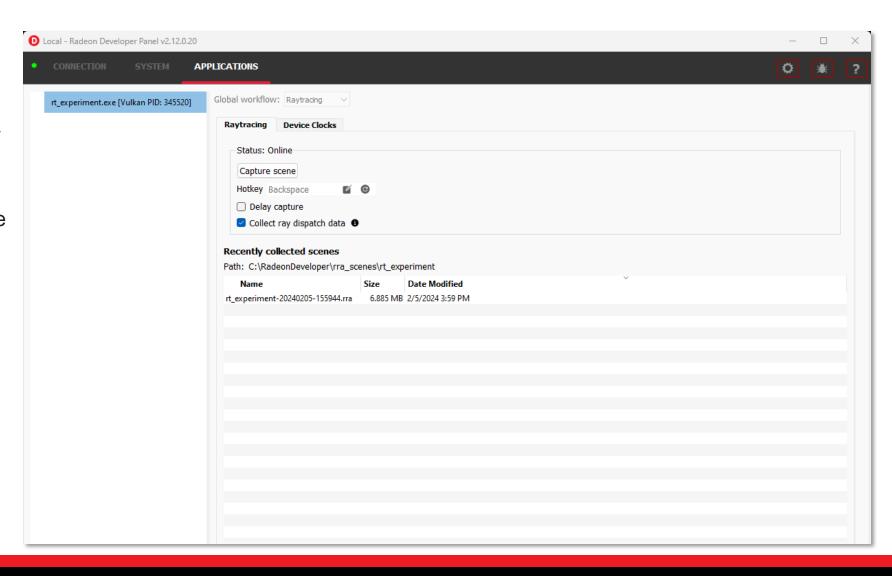


49

GDC 2024

RRA 1.3 - NEW RAY FEATURES

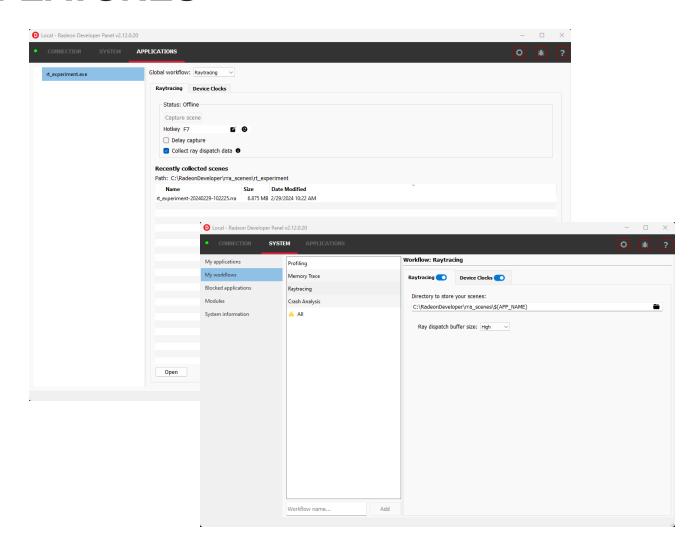
- New ability to capture raytracing dispatches
- Captured through RDP by enabling "Collect ray dispatch data" option
- A buffer size option is needed to ensure all the data can be captured





RRA 1.3 - NEW RAY FEATURES

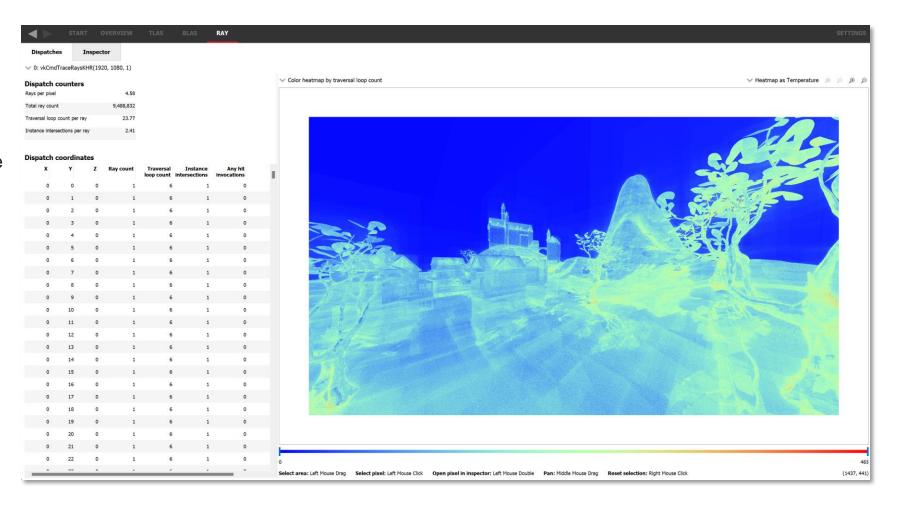
- New ability to capture raytracing dispatches
- Captured through RDP by enabling "Collect ray dispatch data" option
- A buffer size option is needed to ensure all the data can be captured





RRA 1.3 - DISPATCH

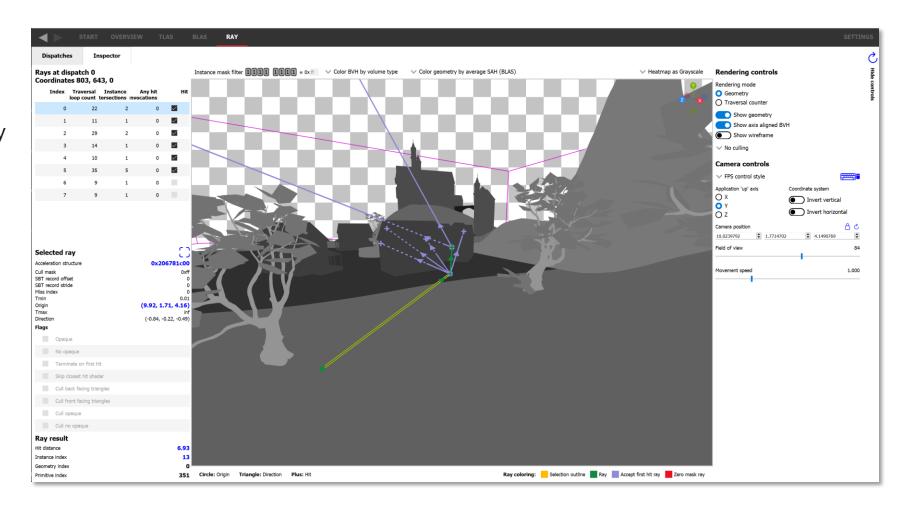
- Dispatch can be visualized into heatmap of various types if the dispatch shape is 2D or 3D
- Can also be mapped from 1D dispatches, must not be sparse
- Allows for exact traversal cost analysis since all the shadow and reflection rays will be included
- Select a "pixel" to inspect it





RRA 1.3 - RAY INSPECTOR

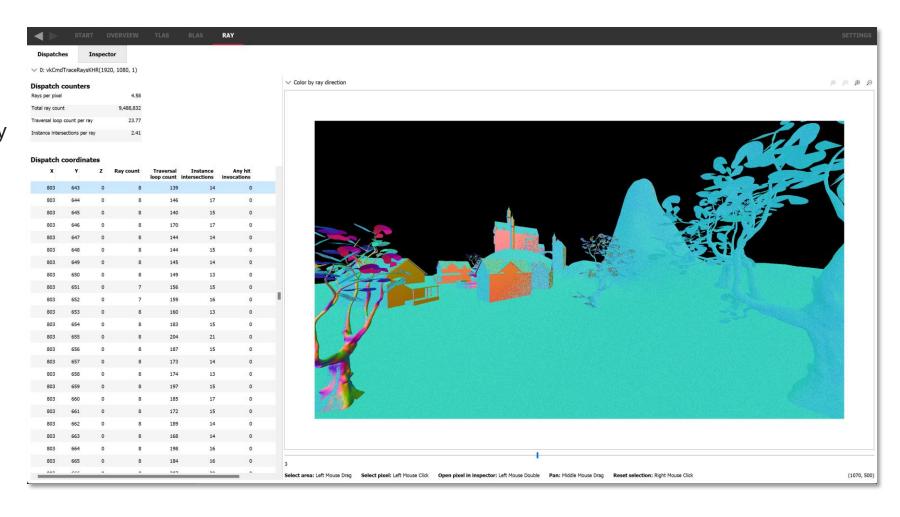
- View each ray for a given dispatch index
- See the cost and arguments for each ray





RRA 1.4

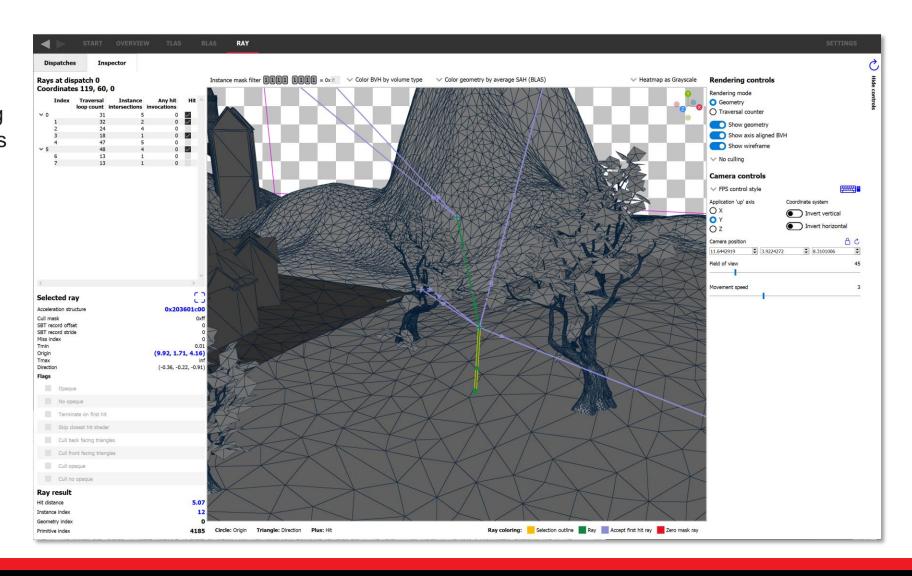
- New ray directions feature on the ray dispatches
- Quickly find and identify shadow and reflection areas
- Bugfixes and quality of life updates



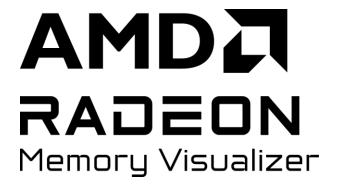


PREVIEW: RRA 1.5

- New ray hierarchy in Ray Inspector
- Quickly verify sampling rate and recursive calls





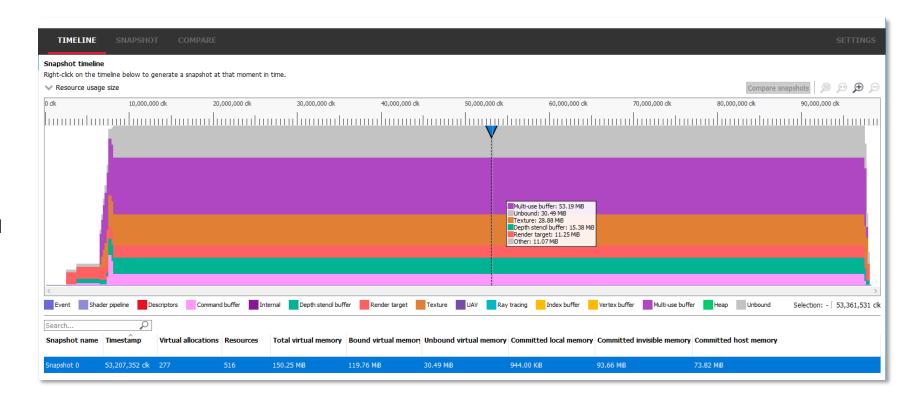




56

RMV TIMELINE IMPROVEMENTS

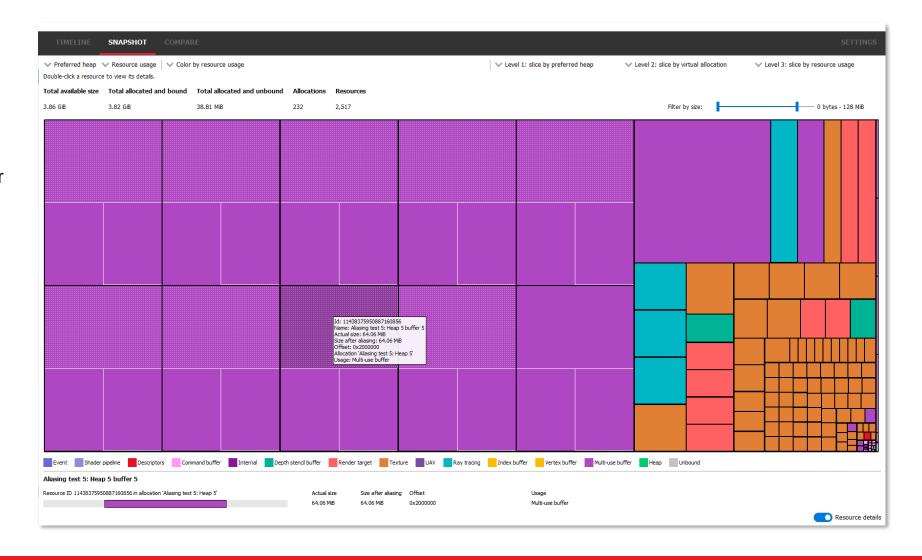
- Better memory visualization
- Unbound memory is now shown on timeline
- Usage sizes for aliased resources properly calculated
- Implicit resources filtered from calculations





RMV RESOURCE OVERVIEW IMPROVEMENTS

- Usage sizes for aliased resources properly calculated and shown in tooltip
- Named allocations
- Improved size range filter





OTHER RMV IMPROVEMENTS

- Heap overview pane contains additional information
- Device configuration pane shows system memory and driver info
- Support added for file format which supports compression
- New time unit format
- Expanded history resource columns







AMD PUBLIC | GDC 2024 | AMD GPU DEVELOPER TOOLS | March 2024

60

AMD RADEON™ GPU DETECTIVE (RGD)

Newest member of AMD Radeon[™] Developer Tool Suite (https://gpuopen.com/tools/)

Overview:

- Tool for post-mortem analysis of GPU crashes
- Sets driver to Crash Analysis mode before reproducing crash
- Developers capture AMD GPU Crash Dump files upon crash
- Produces concise crash analysis report in Text/JSON formats
- Report helps narrow down the search for the crash root cause



AMD RADEON™ GPU DETECTIVE (RGD)

Newest member of AMD Radeon™ Developer Tool Suite (https://gpuopen.com/tools/)

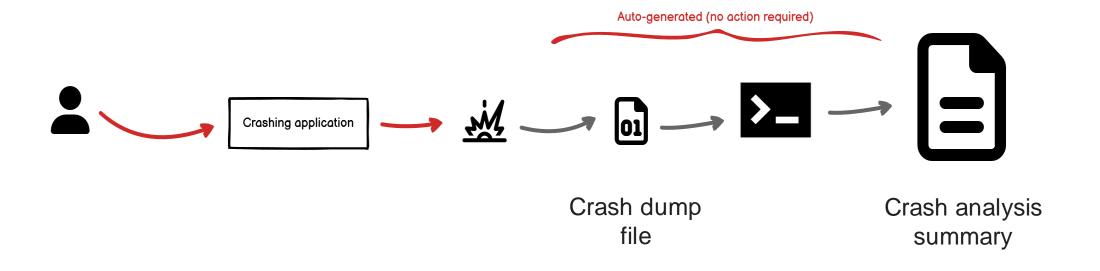
Requirements:

- OS: Windows 10 or 11
- GPU: AMD Radeon™ RX 7000 Series (RDNA™ 3) or AMD Radeon™ RX 6000 Series $(RDNA^{TM} 2)$
- Driver: AMD Software: Adrenalin Edition 23.12.1 or newer
- Graphics API used by the crashing application: DirectX 12 or Vulkan

GDC 2024



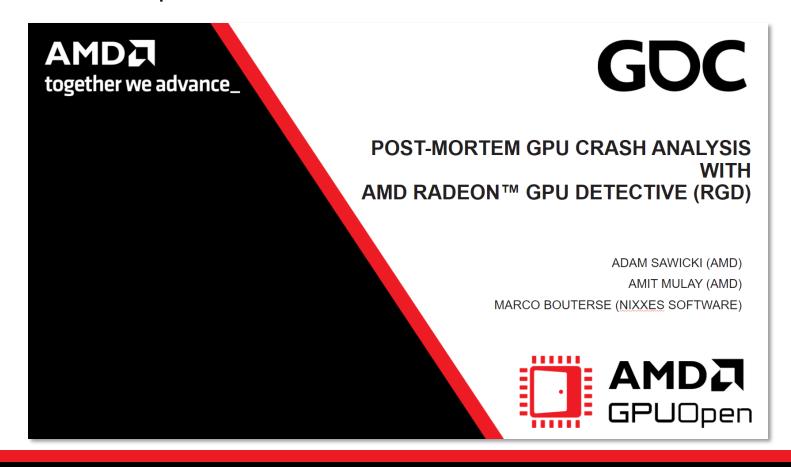
UNVEILING THE TOOL: WORKFLOW AND FUNCTIONAL INSIGHTS





UNVEILING THE TOOL: WORKFLOW AND FUNCTIONAL INSIGHTS

- Curious about the newest addition to our tool suite?
- Stay tuned for the next presentation





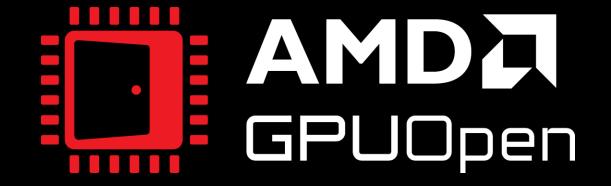
DISCLAIMER

GENERAL DISCLAIMER

The information contained herein is for informational purposes only and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD's products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale. GD-18

© 2024 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Radeon and combinations thereof are trademarks of Advanced Micro Devices, Inc. DirectX is a registered trademark of Microsoft Corporation in the US and other jurisdictions. Linux is a registered trademark of Linus Torvalds. OpenCL is a trademark of Apple, Inc. used by permission from The Khronos Group. LLVM is a trademark of LLVM Foundation. SPIR, SPIR-V and the SPIR logo are trademarks of the Khronos Group Inc. Vulkan and the Vulkan logo are registered trademarks of the Khronos Group Inc. Windows is a registered trademark of Microsoft Corporation in the US and other jurisdictions. Other product names used in this publication are for identification purposes and used may be trademarke of their respective companies. in this publication are for identification purposes only and may be trademarks of their respective companies.





AMD together we advance_

AMD TRYZEN

