Mesa Continuous Integration at Intel

Mark Janes
Clayton Craft

Free Software Spectum at Intel

Zune was great	SurfacePro for development	Likes Cortana	No Android	Distro tattoo	Open FW	Open Hardware	Tin Hat

1×10^{-14} 1×10^{-12} 1×10^{-8} 1×10^{-4} 1×10^{-2} 1×10^{2} 1×10^{4}

Mark Clayton

4×10^{-7} 5×10^{-7} 6×10^{-7} 7×10^{-7}

WIndows phone superior	Outlook encrypted emails	Edge/Bing for web	Ubuntu VM	Dual Boot	Command Line Tools	Linux only	Distro bumper sticker	Text-based email client	Tiling window manager	Declines Android services

Ballmer

Stallman

# Why is continuous integration valuable for Mesa?

Summary of Mesa CI at Intel

- ~200 systems with full Intel hardware coverage going back to 2007
- dEQP, Piglit, VulkanCTS, Crucible, OpenGL CTS, OpenGLES CTS
- Build-tests for Android and non-Intel platforms
- Millions of tests per run for every commit
- Target execution time of 30min.
- False positives ~0.0001%
- Generates performance trend lines for common benchmarks
- Open source CI implementation (https://gitlab.freedesktop.org/Mesa_CI)

Mesa Performance

gfxbench5_tess
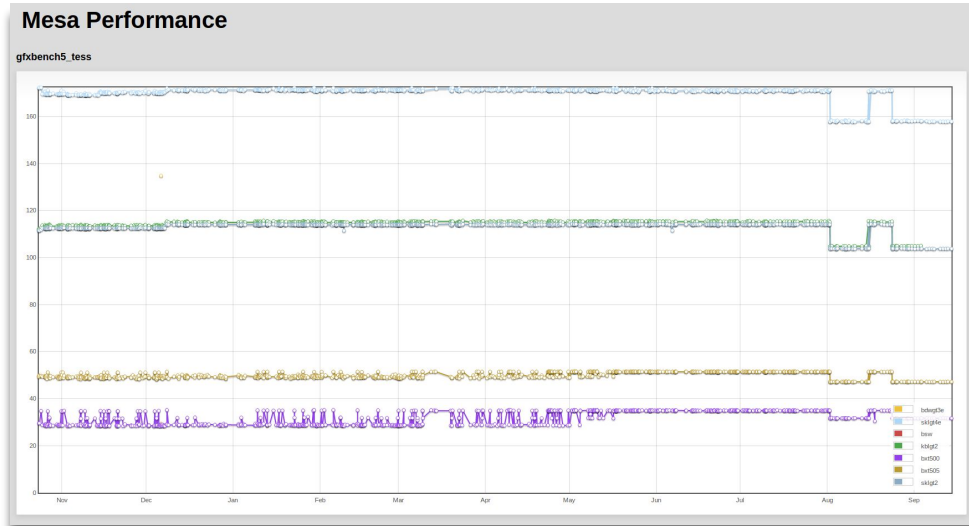
Phoronix: making more whitespace commits
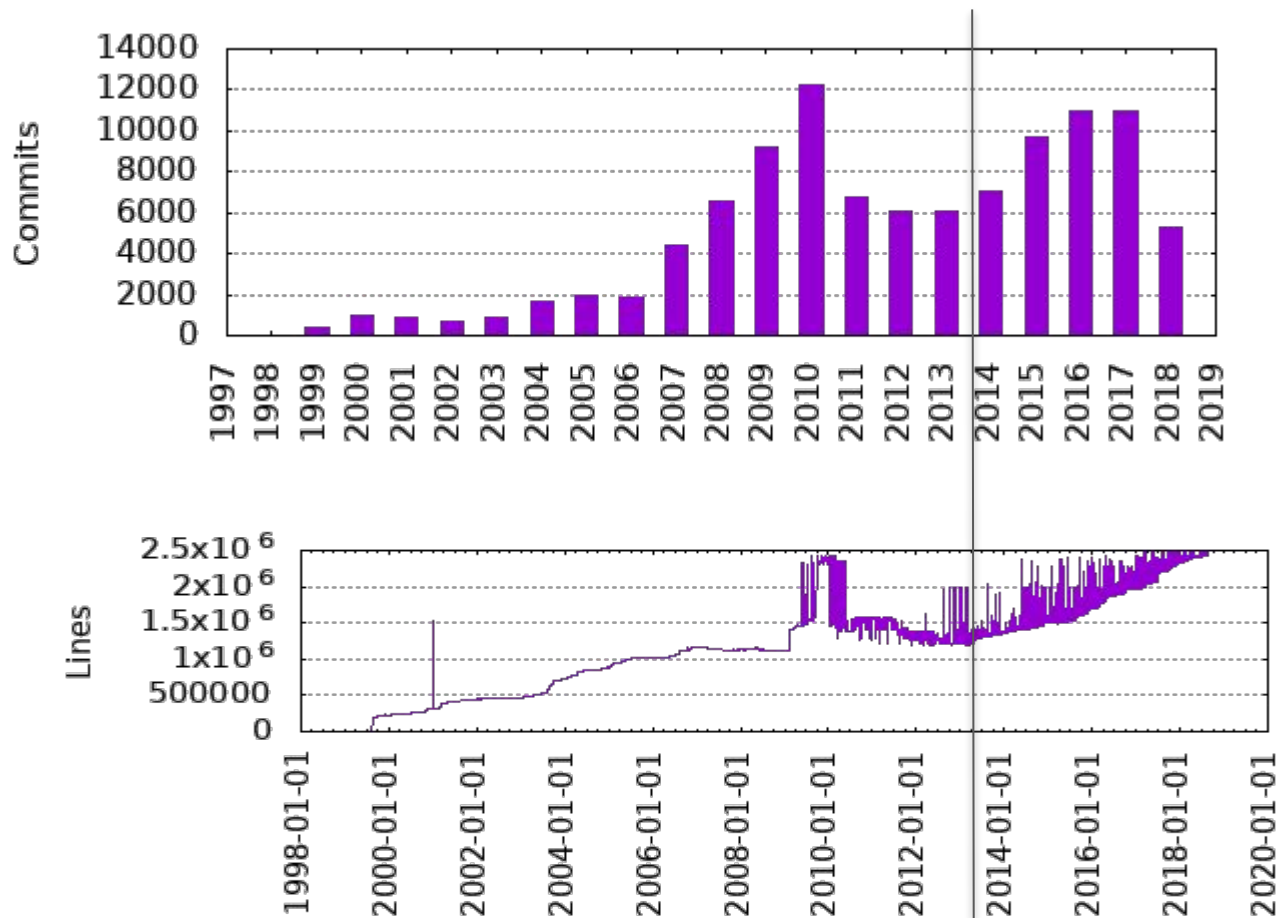
CI Builds vs Commits since 8/2014, Developers with > 30 commits

■ CI Builds ■ Commits
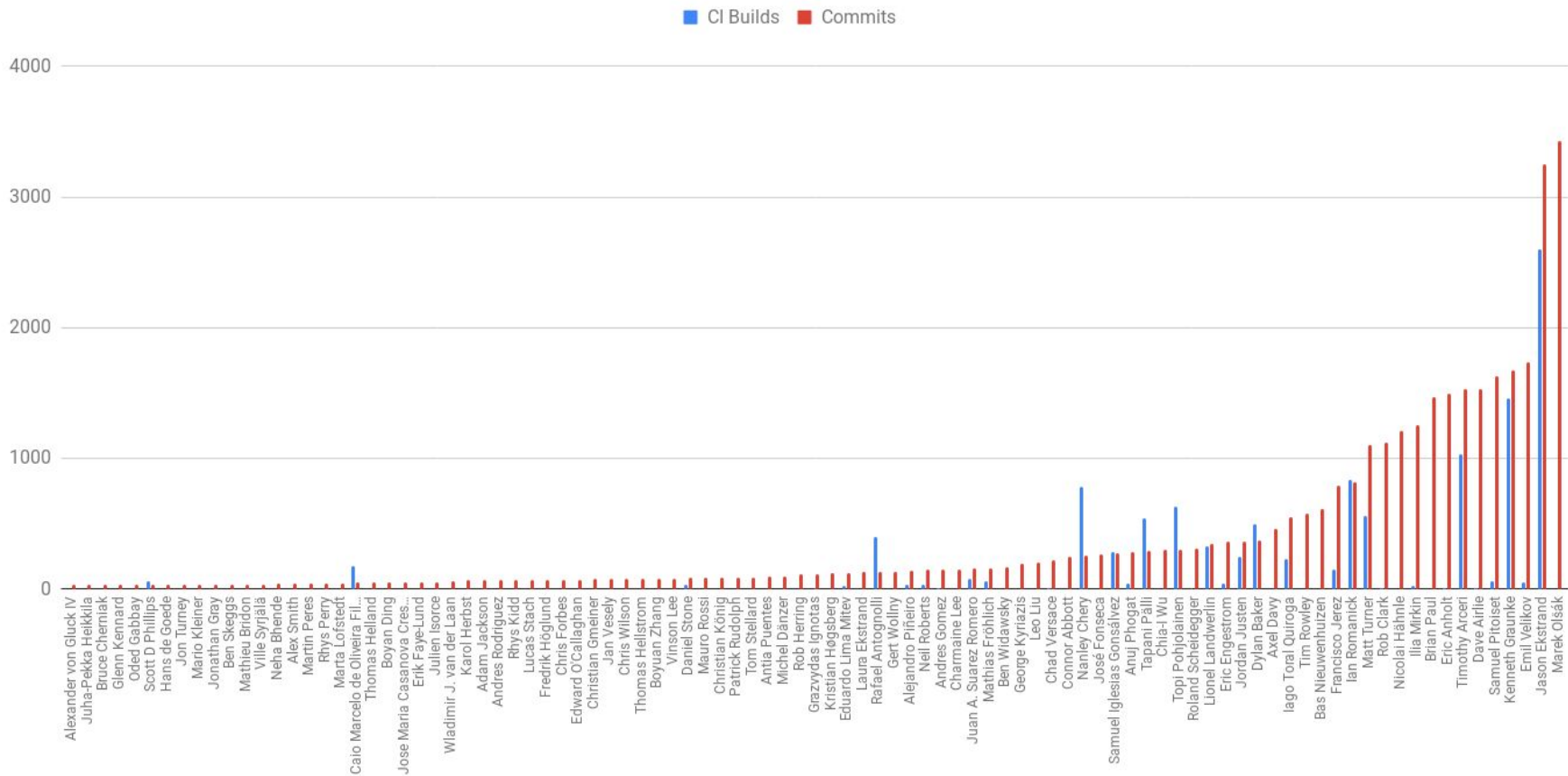
Reverts and Fixes tags

6521d4a659b911bb86d979564de03665616a671e
Author: Samuel Pitoiset <samuel.pitoiset@gmail.com>
Commit: Samuel Pitoiset <samuel.pitoiset@gmail.com>

Revert "radv: Optimize rebinding the same descriptor s

This introduces random GPU hangs on Vega, at lea

This reverts commit
02a43edf186cb9998741ba765cb948bb238a122d.

02a43edf186cb9998741ba765cb948bb238a122d
Author: Bas Nieuwenhuizen <bas@basnieuwenhuizen.nl>

radv: Optimize rebinding the same descriptor set.

This makes it cheaper to just change the dynamic offsets with
the same descriptor sets.

Suggested-by: Philip Rebohle <philip.rebohle@tu-dortmund.de>
Reviewed-by: Samuel Pitoiset <samuel.pitoiset@gmail.com>

c75a4e5b465261e982ea31ef875325a3cc30e79d
Author: Dylan Baker <dylan@pnwbakers.com>

meson: Check for actual LLVM required versions

Currently we always check for 3.9.0, which   pretty
safe since everything except radv work   th >= 3.9 and
3.9 is pretty old at this point.   ever, radv actually
requires 4.0, and there is a   n for radeonsi to do
the same.

Fixes: **673dda833076** ("meson: build "radv" vulkan driver
for radeon hardware")
Signed-off-by: Dylan Baker <dylan.c.baker@intel.com>
Reviewed-by: Bas Nieuwenhuizen <bas@basnieuwenhuizen.nl>

673dda8330769309a319d3e7f24a029cd72a1caf
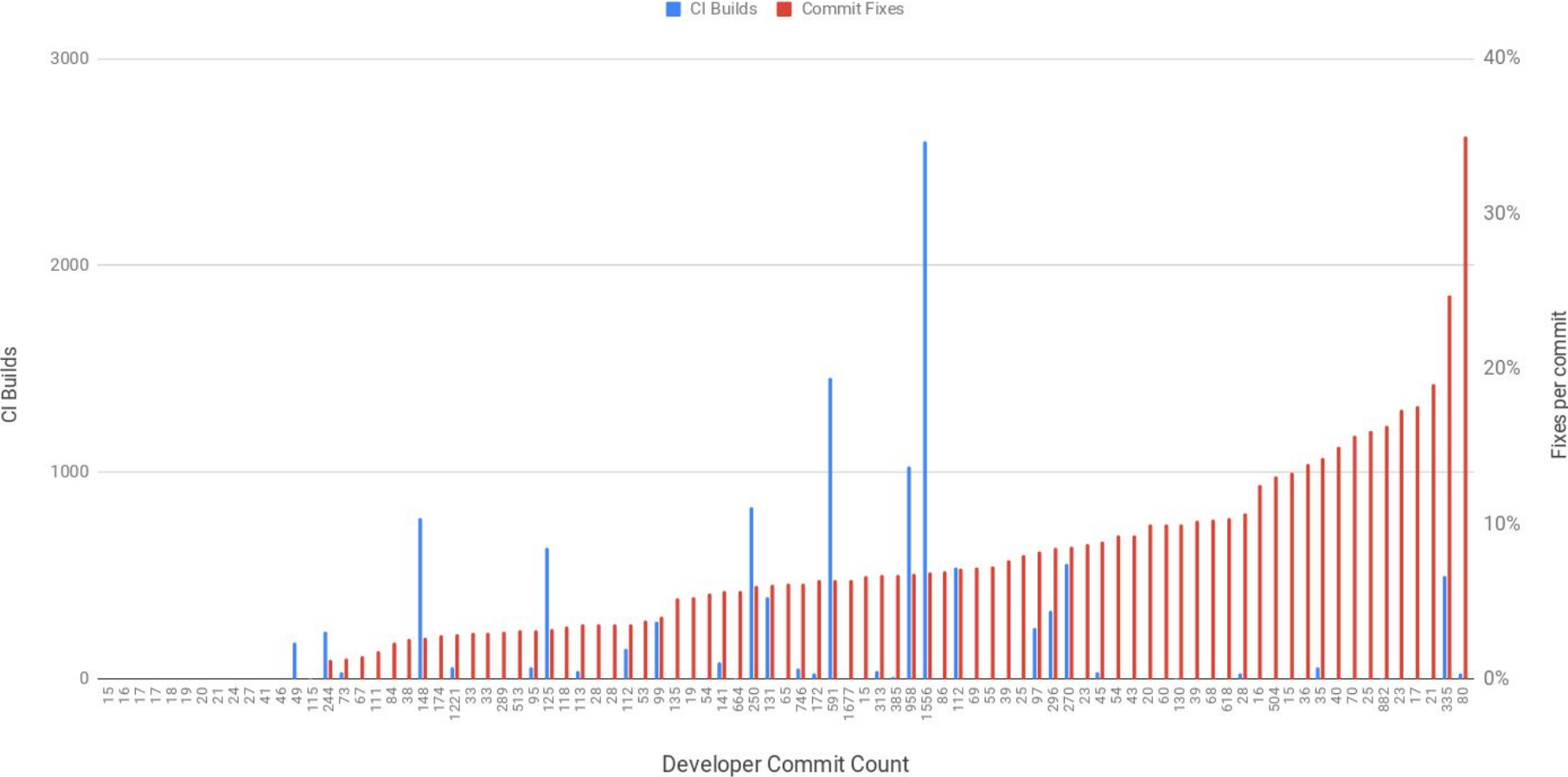Author: Dylan Baker <dylan@pnwbakers.com>

meson: build "radv" vulkan driver for radeon hardware

This builds, installs, and has been tested on a r290x
(Hawaii) with the Vulkan CTS. It dies horribly in a fire at
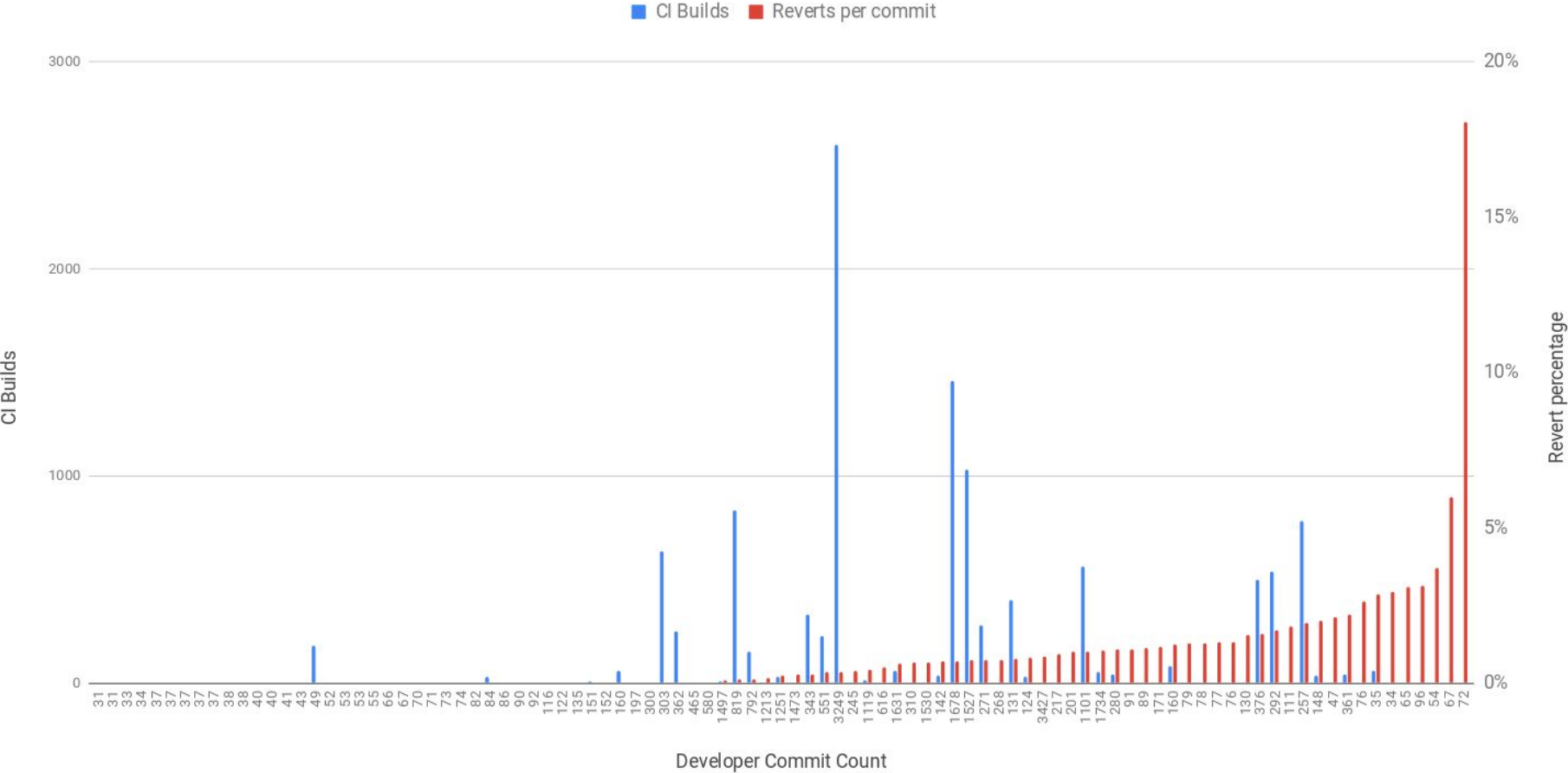the same point for the meson build as the autotools build.

Signed-off-by: Dylan Baker <dylanx.c.baker@intel.com>
Reviewed-by: Bas Nieuwenhuizen <bas@basnieuwenhuizen.nl>

Fixes per commit since 2017, Developers with > 15 commits

■ CI Builds ■ Commit Fixes

CI Builds / Fixes per commit

Developer Commit Count

Reverts per commit since 9/2014, Developers with > 30 commits

Legend: ■ CI Builds ■ Reverts per commit

X-axis label: Developer Commit Count

Left Y-axis: CI Builds

Right Y-axis: Revert percentage

Tracking test status

i965 CI tracks all test status changes in configuration files.

Known issues are filtered from the results, to make new regressions obvious in results:

- Developer pushes a broken commit.
- CI regresses at least one test.
- i965 CI staff investigate, write FDO bug, where resolution is tracked.
- Regressed tests are filtered to "skip" status by CI config.
- Developer fixes bug.
- Test status changes, creating unexpected result in CI.
- I965 CI staff investigate.
- CI configs updated to reflect fix.

CI Config file: SandyBridge Piglit results

```
[expected-failures]
piglit.shaders.arb_texture_gather-miplevels
piglit.shaders.point-vertex-id
...
piglit.shaders.glsl-deriv-varyings = piglit cd62eff8e5
Piglit.spec.ext_texture_compression_s3tc = piglit_d05448d06f

[expected-crashes]
piglit.fast_color_clear.fcc-front-buffer = mesa 880573e7
piglit.spec.egl 1_4.egl-copy-buffers = piglit 85e3b32b32

[fixed-tests]
piglit.spec.glsl-es-1_00.linker.fface-invariant = mesa 9b5c0c520
piglit.spec.glsl-es-1_00.linker.fcoord-invariant = mesa 9b5c0c520
```

Filtering results based on blamed commit

```
[expected-failures]
piglit.spec.foo

[expected-crashes]

[fixed-tests]
```

e1623da8185 * idr test bug fix 1

49e4248a93a * i965/nir: export nir_optimize

Filtering results based on blamed commit

When test status changes, i965 CI staff triage results, close FDO bugs, and update CI configuration.

```
[expected-failures]
piglit.spec.foo

[expected-crashes]

[fixed-tests]
piglit.spec.foo
```
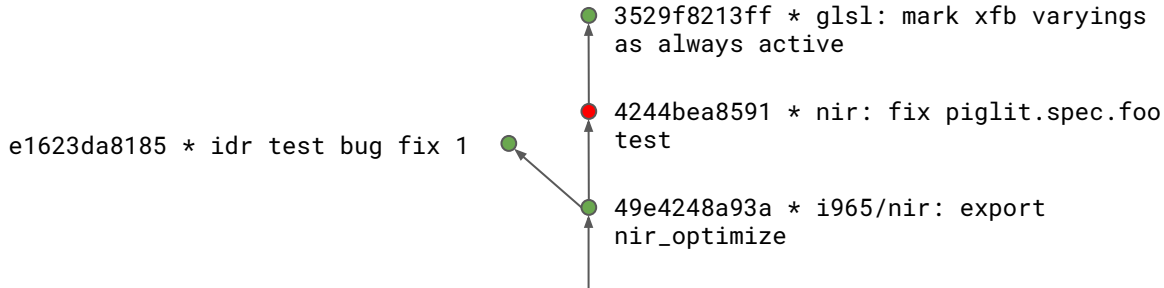
e1623da8185 * idr test bug fix 1

4244bea8591 * nir: fix piglit.spec.foo test

49e4248a93a * i965/nir: export nir_optimize
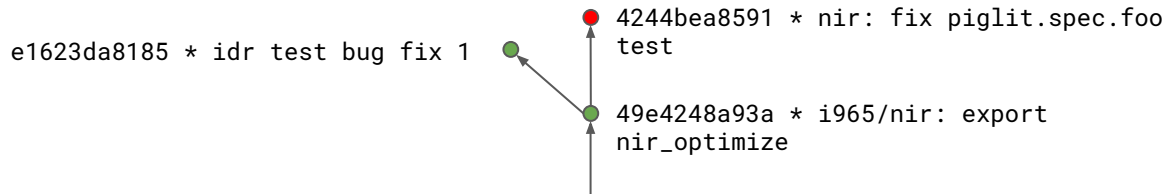
Filtering results based on blamed commit

Updated CI configuration files will report success for subsequent CI builds.

Test failure patterns are hardware specific. Each platform needs a separate configuration file. Some test suites require separate configuration for 32-bit builds.

```
[expected-failures]

[expected-crashes]

[fixed-tests]
piglit.spec.foo
```

3529f8213ff * glsl: mark xfb varyings as always active

4244bea8591 * nir: fix piglit.spec.foo test

e1623da8185 * idr test bug fix 1

49e4248a93a * i965/nir: export nir_optimize

Filtering results based on blamed commit

Branches will report spurious test status changes as CI tracks progress in the master branch.

Ian's branch does not contain the fix for piglit.spec.foo. Ian's is test run will fail that test. The failure state does NOT match CI expectations.

```
[expected-failures]

[expected-crashes]

[fixed-tests]
piglit.spec.foo
```

962cc1bd17c * idr fix comment
WTF?! Why does this fail now?

e1623da8185 * idr test bug fix 1

3529f8213ff * glsl: mark xfb varyings
as always active

4244bea8591 * nir: fix piglit.spec.foo
test

49e4248a93a * i965/nir: export
nir_optimize
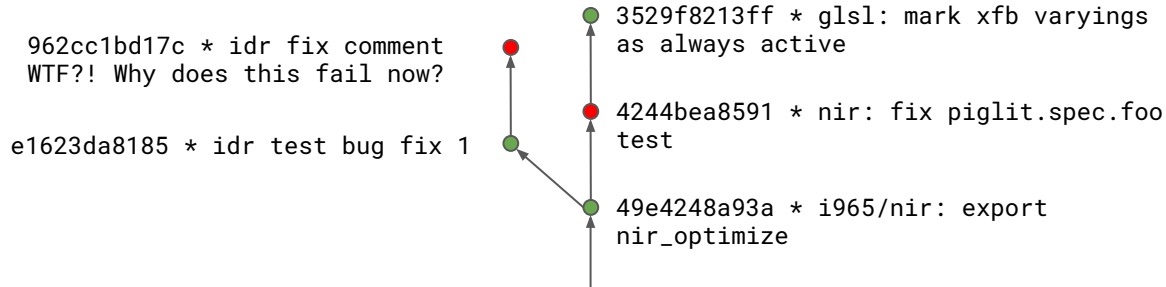
Filtering results based on blamed commit

Mesa CI records the blamed commit for every test status change.

For every unexpected test result, Mesa checks to see if the target branch contains the commit blamed by the CI config.

Ian's branch does not contain 4244bea8591, so CI comprehends that the test status ought to be wrong for the branch.

```
[expected-failures]

[expected-crashes]

[fixed-tests]
piglit.spec.foo = mesa 4244bea8591
```

```
majanes@giraffe:~/src/mesa$ git branch -a --contains
4244bea8591
* master
* remotes/curro/wip/test
majanes@giraffe:~/src/mesa$
```

```
962cc1bd17c * idr fix comment
piglit.spec.foo is filtered out

e1623da8185 * idr test bug fix 1
```

```
3529f8213ff * glsl: mark xfb varyings
as always active

4244bea8591 * nir: fix piglit.spec.foo
test

49e4248a93a * i965/nir: export
nir_optimize
```
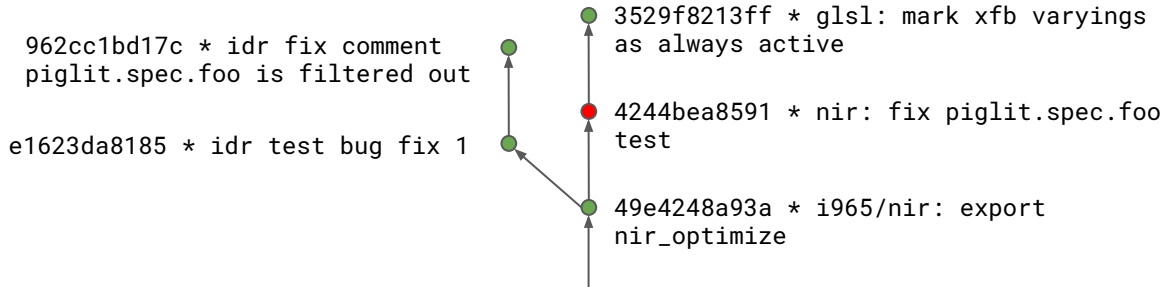
Filtering results based on blamed commit

Over time, CI configuration allows testing of releases for Mesa stable branches and old test suites.

Automated tests are fixed on a daily basis. Over time, this represents thousands of test results. Testing a stable point release is non-trivial.

Typically, CI systems fork the entire CI to test a stable branch. This is incompatible with hardware updates and other changes that affect all branches.

```
[expected-failures]
dEQP-GLES3.functional.shaders.preprocessor.builtin.line_expression_fragment = deqp 55988830
dEQP-GLES3.functional.shaders.preprocessor.builtin.line_expression_vertex = deqp 55988830

[expected-crashes]

[fixed-tests]
dEQP-GLES3.functional.state_query.integers.stencil_value_mask_getfloat = mesa 37d63b50
dEQP-GLES3.functional.state_query.integers.stencil_back_value_mask_getfloat = mesa 37d63b50
dEQP-EGL.functional.color_clears.multi_thread.gles1_gles2.rgba8888_window = deqp 89c3844c
dEQP-EGL.functional.color_clears.multi_context.gles1_gles2.rgb888_pbuffer = deqp 89c3844c
dEQP-GLES3.functional.state_query.integers.stencil_value_mask_separate_both_getfloat = mesa 37d63b50
dEQP-GLES3.functional.state_query.integers.stencil_back_value_mask_separate_both_getfloat = mesa 37d63b50

<many more fixed tests ... >

EQP-EGL.functional.create_context.no_config = mesa 5e2909e7
dEQP-GLES31.functional.debug.negative_coverage.log.tessellation.single_tessellation_stage = mesa e6e8475b
dEQP-GLES3.functional.negative_api.texture.teximage3d = deqp 9e51a954
```

9e51a954 * fix dEQP test

49e4248a93a * 17.2.2

e1623da8185 * 17.2.1

3529f8213ff * 17.2 branchpoint

962cc1bd17c * 17.1.2

< **many** commits >

55988830 * new dEQP test

e1623da8185 * 17.1.1

49e4248a93a * 17.1 branchpoint
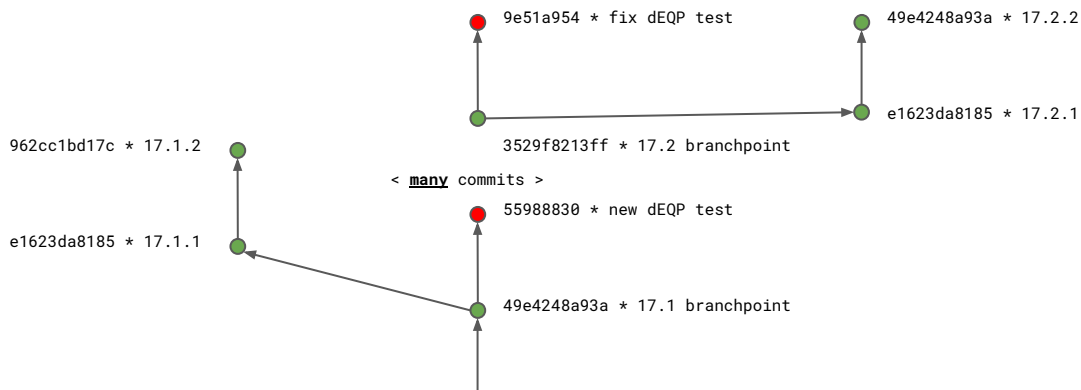
Commits regressing i965 tests since 8/2014, developers with >15 commits

Legend: ■ CI Builds ■ Commits causing i965 Regressions

Y-axis (left): CI Bulids — 0, 1000, 2000, 3000
Y-axis (right): Commits regressing tests — 0%, 3%, 5%, 8%, 10%
X-axis: Developer Commit Count
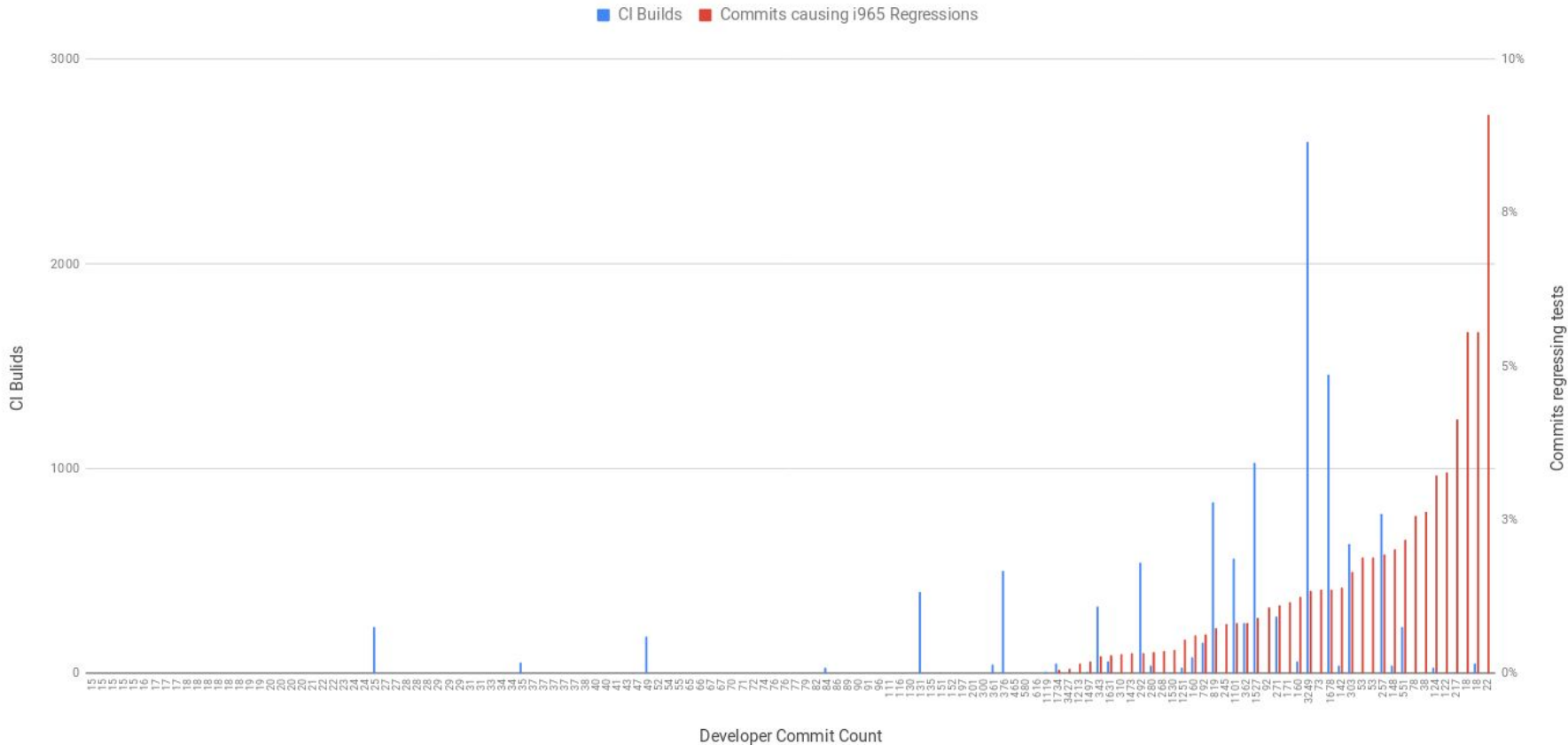
Tests regressed per commit since 8/2014, Developers with >15 commits
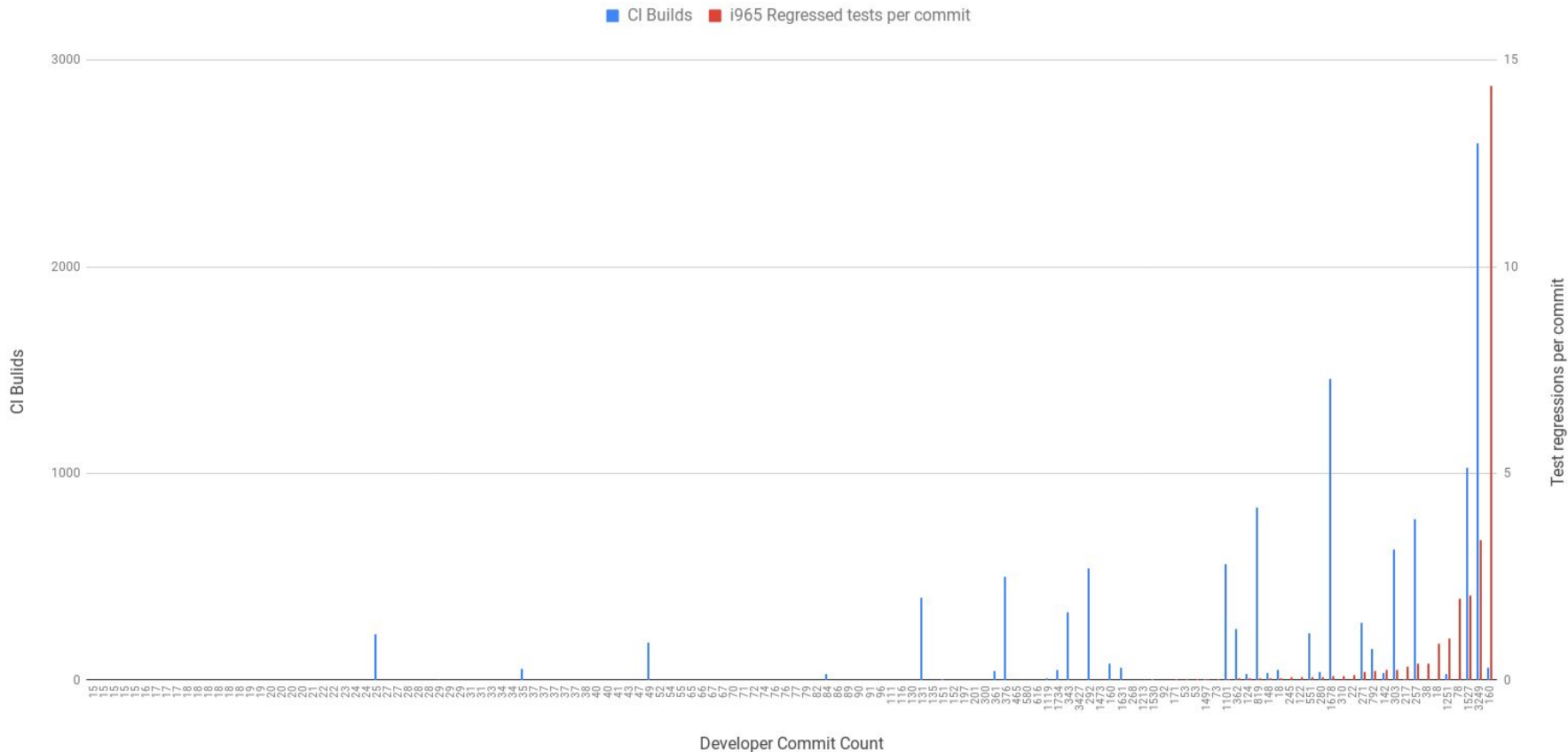
Mesa CI at Intel

Mesa CI at Intel

Mesa CI at Intel

High density storage

Summary of Mesa CI at Intel

High density storage
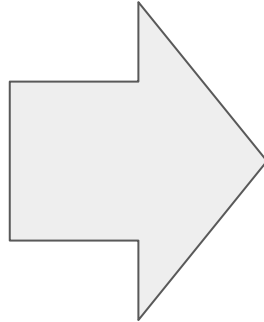
How are we improving it?

New public CI results site!

Features:

- Results for other jobs (eg master, kwg, jekstrand)
- Broken tests counts
- Revisions / transparency of sources
- Logs for broken components
- Test history
- Browse results by test suite
- It's fast!

Internal results site:	1 minute 24 seconds
New public results site:	0.62(ish) seconds

Time to load a CI build result page

https://mesa-ci.01.org

How are we improving it?

New public CI results site!

https://mesa-ci.01.org

How are we improving it?

Future:

- Support new use cases for the results website:
    - Show the logs and the status of components during execution
    - Allow developers to trigger custom builds
        - Queue up a set of jobs, with names
        - Especially useful for piglit
    - Allow developers to do A/B comparisons of builds
- Collaborate with other GPU vendors or Distros.

Got ideas? Let us know!

- https://www.pivotaltracker.com/n/projects/1471364
- https://gitlab.freedesktop.org/Mesa_CI/mesa_ci_results

Backup

Caveats

Rebased branches (e.g. i915 kernel trees) break methodology for tracking regressions

- Blamed commit SHAs may no longer exist after a rebase

Can hide regressions at release time

- Regressions are accepted failures in CI, so i965 CI staff has to manually compare results to N-1 release to find regressions attributed to the release candidate

But... A/B testing is more costly and only shows data for a single delta!

Test Regressions per KLOC, contributors with > 2 KLOC

Why is continuous integration valuable for Mesa?

Mesa CI at Intel is currently used for

      … Developer (Intel and community) testing

      … Mesa release verification

      … Intel pre-silicon testing in simulation

      … Performance testing

      … Validation of upstream test suites (dEQP, Vulkancts, etc)