

ECE 555 Group Presentation

Igor Semyonov Jordan Carnes Robert Laverne Griffin
George Macon University, Department of Electrical and Computer Engineering

April 15, 2025

Why not Just Use C?

It should not be used for production.
See https://veresov.pro/cmustdie/

Why Rust Type System

Listing 1: Cats

```
struct FakeCat {
   alive: bool,
   hungry: bool,
}

enum RealCat {
   Alive {
   hungry: bool,
   },
```

Impossible states can be encoded in the type system, which leads to compile time errors, instead of runtime checks.

Listing 2: Unphysical (zombie) Cat

```
let fake_cat = FakeCat {
    alive: false,
    hungry: true,
};
```

Why Rust Type System

```
#® main.rs > ® cats.rs
                                                                          main
  5
        -let real cat1 = RealCat::Dead:↓
  3
        let real cat2 = RealCat::Alive {↓
  2
             hungry: true.↓
  1
        -};↓
23
        let real_cat3 = RealCat::Dead{hungry: true};
  1
     rust-analyzer: no such field [E0559]
     rustc: variant `RealCat::Dead` has no field named `hungry`
     all struct fields are already assigned [E0559]
      cats.rs 🛭 🗷 5 0 1
NOR
                                                utf-8 < 🔞 rust 🦪
                                                                  88%
                                                                         23:1
```

Figure 1: Error when declaring an unrealistic cat

Why Rust Safety

- Result: Errors are values, i.e., no nidden control flow
- Functions that can fail return Result and this must be explicitly handled.

- No manual memory management
- Values are dropped once they go out of scope

Why Rust Ergonomics

High level abstractions that enable multithreading without the error-prone approach of pthreads in ${\sf C}.$

Why Rust Single threaded bsort

Listing 3: Single threaded bsort

- a is a vector of numbers
- chunks_mut slices the vector into non-overlapping slices of the given length
- for_each iterates over the chunks, running the provided function with each chunk as the input.

Why Rust Concurrency is easy while avoiding race conditions

Listing 4: Example of easy parallelization

```
a.chunks_mut(split_length)
par_bridge()
for_each(bsort);
```

- Chunks is known to split a vector into distinct slices
- Hence we can safely send each chunk to a different thread

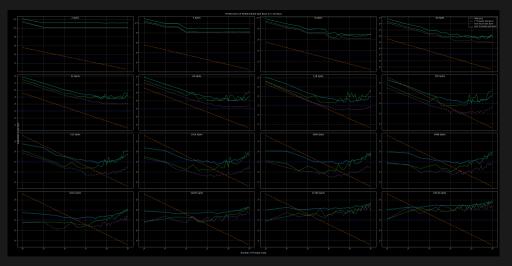


Figure 2: Preformacne results from bsort project. Note that the Rust reference sort is not plotted as it was under 1ms, compared to 8ms for the C qsort.

Why Rust Zero Cost Abstraction

All these features are provided at compile time and are optimized away when compiling so that they do not affect performance at run time.

Rust Ownership and Borrowing Problems with shared memory access in Rust

- Ownership Rules[1, Ch04-01]
 - o Each value has an owner
 - There can only be one owner at a time
 - When the owner goes out of scope, the value will be dropped.

- Borrow Rules[1, Ch04-02]
 - At any given time, you can have either one mutable reference or any number of immutable references.
 - References must always be valid.

C, and especially CUDA, style shared mutable memory access violates the ownership and borrow rules.

At this time, we therefore need to use unsafe Rust when writing and calling GPU kernels.

Rust and CUDA Current options

- Rust GPU
- Rust CUDA

Maybe mention other options for GPU programming in rust, like any rust support for ROCm, HIP, intel gaudi, or other.

Rust and CUDA Description of how the Rust-CUDA crate works

Rust and CUDA Focusing on Rust CUDA

Currently being rebooted and is in active development. Uses nvidia's nvvm tool which is built on LLVM 7.



Description of problems when going straing from source code to machine code.

The problem LLVM solves

Description of LLVM and it's intermediate representation and how this has enabled much easier language development.

NVVM

How NVVM works

Rust CUDA and NVVM

How NVVM is used in Rust CUDA

Fractals Mandelbrot and Burning Ship

 $c \in \mathbb{C}$ is in the mandelbrot set if the sequence $\{z_n\}$ converges.

$$z_{n+1} := z_n^2 + c$$
 $z_0 = 0$

The Burning Ship fractal is defined similarly but the sequence is

$$z_{n+1} := (|\operatorname{Re}(z_n)| + |\operatorname{Im}(z_n)|i)^2 + c$$
 $z_0 = 0$

Mandelbrot Kernel

```
pub unsafe fn mandelbrot(
        n re: usize.
        n im: usize,
        re min: f32,
        re max: f32.
 6
        re range: f32,
        im min: f32.
 8
        im max: f32.
 9
        im range: f32.
        zn limit: u32.
        out: *mut u8,
12
13
         let idx linear = thread::index() as usize:
14
         let idx = thread::index 2d():
15
         let idx re = idx[1]:
         let idx im = idx [0]:
16
18
         let c_re = re_range
19
             * (idx_re as f32 / (n_re - 1) as f32)
20
            + re_min;
         let c im = im range
             * (idx_im as f32 / (n_im - 1) as f32)
23
            + im_min;
```

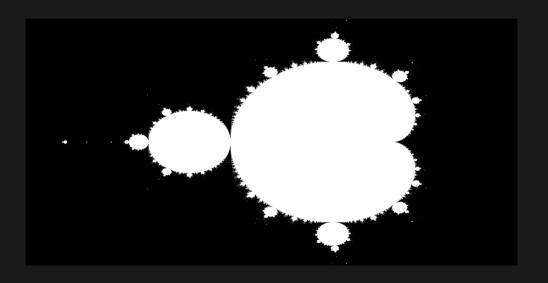
```
24
         let mut z re = c re;
25
         let mut z_im = c_im;
         for in 0..zn limit / ZN SKIP {
26
27
             for in 0..ZN SKIP {
28
29
                      z_re, z_im,
30
                  ) = (
                      z re * z re - z im * z im +
                            c_re,
                      2.0 * z re * z im + c im.
33
34
             if z_re * z_re + z_im * z_im > 4.0 {
35
                  let elem = &mut *out.add(
36
                       idx linear):
37
                  *elem = \overline{255}:
                 break:
38
39
40
41
```

Launching the Kernel

```
unsafe {
             launch!(
                  module.mandelbrot <<< grid_size, block_size,
                        0. stream >>>(
                      N RE.
4
                      N IM.
                      re min.
                      re_max,
                      re_range,
                      im_min,
10
                      im max.
                      im_range,
                      zn_limit,
13
                      out gpu.as device ptr().
14
15
16
```

- Fiarly similar to launching a kernel in C.
- Must be used inside an unsafe block since the kernel itself is an unsafe function.

Output Image



Fractals Timing Results for 1 frame on CPU and GPU

I will add the single, frame timing results here

Fractals Live Demo

I plan to be sharing my screen for the presentation and will switch to a live demo here.

This is a slide With a subtitle

This is some text in a column. Could be a figure instead.

- This is a list
- It is an itemized one
- Hence the bullets