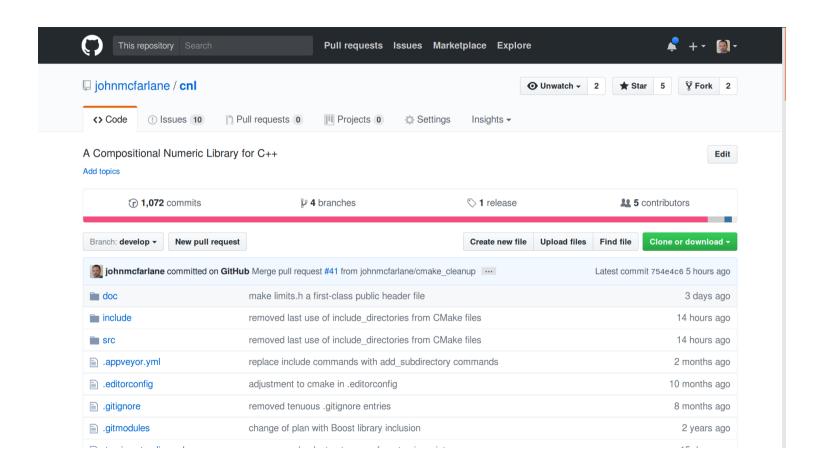
CNL: A Compositional Numeric Library

John McFarlane

A9.com

Background

Background



1. Low resolution

- 1. Low resolution
- 2. Limited range

- 1. Low resolution
- 2. Limited range
- 3. 'Interesting' behavior of arithmetic operations

1. Complicated

- ± 1.significand * 2 exponent
- special values, denormalized values, -0

- 1. Complicated
 - ± 1.significand * 2 exponent
 - special values, denormalized values, -0
- 2. Occasional weirdness can surprise:
 - o determinism, associativity, commutativity and ordering

- 1. Complicated
 - ± 1.significand * 2 exponent
 - special values, denormalized values, -0
- 2. Occasional weirdness can surprise:
 - determinism, associativity, commutativity and ordering
- 3. <cmath> functions lack constexpr

- 1. Complicated
 - ± 1.significand * 2 exponent
 - special values, denormalized values, -0
- 2. Occasional weirdness can surprise:
 - determinism, associativity, commutativity and ordering
- 3. <cmath> functions lack constexpr
- 4. Variable resolution

- 1. Complicated
 - ± 1.significand * 2 exponent
 - special values, denormalized values, -0
- 2. Occasional weirdness can surprise:
 - o determinism, associativity, commutativity and ordering
- 3. <cmath> functions lack constexpr
- 4. Variable resolution
- 5. Costly in energy and silicon

• Floating-point problems are not so bad.

- Floating-point problems are not so bad.
- Integers are a powerful abstraction over registers.

- Floating-point problems are not so bad.
- Integers are a powerful abstraction over registers.
- But we can do a lot better.

- Floating-point problems are not so bad.
- Integers are a powerful abstraction over registers.
- But we can do a lot better.
- so ...

"Do for int what the STL did for []."

"Do for int what the STL did for []."

• Provide zero-cost abstractions over language-level features:

```
std::array<T, N> a; // T a[N]
std::array<T, N>::iterator i = std::begin(a); // T* i
```

"Do for int what the STL did for []."

• Provide zero-cost abstractions over language-level features:

```
std::array<T, N> a; // T a[N]
std::array<T, N>::iterator i = std::begin(a); // T* i
```

• Maintain a familiar interface:

```
auto const& third = a[2];
for (auto const& element : a) { /* ... */ }
```

"Do for int what the STL did for []."

Provide zero-cost abstractions over language-level features:

```
std::array<T, N> a; // T a[N]
std::array<T, N>::iterator i = std::begin(a); // T* i
```

• Maintain a familiar interface:

```
auto const& third = a[2];
for (auto const& element : a) { /* ... */ }
```

Allow users to opt in to positive-cost functionality:

```
std::array<T, N> a{n};
auto const& bad_element = a.at(n+1); // gotcha!
```

"Do for int what the STL did for []."

Provide zero-cost abstractions over language-level features:

```
std::array<T, N> a; // T a[N]
std::array<T, N>::iterator i = std::begin(a); // T* i
```

• Maintain a familiar interface:

```
auto const& third = a[2];
for (auto const& element : a) { /* ... */ }
```

• Allow users to opt in to positive-cost functionality:

```
std::array<T, N> a{n};
auto const& bad_element = a.at(n+1); // gotcha!
```

And most importantly...

"Do for int what the STL did for []."

• Compose!

```
using fs_cache = unordered_map<filesystem::path, vector<byte>>;
```

Non-Goal

"Don't do for int what STL doesn't do for []."

Non-Goal

"Don't do for int what STL doesn't do for []."

• Don't make the user pay for what they don't use.

Fixed-Point Arithmetic

Definition:

Fixed-Point Arithmetic

Definition:

Example usage:

```
using cnl::fixed_point;

void f() {
    auto n = fixed_point<int, -8>{ 0.25 };
    std::cout << n * 5; // prints "1.25"
}</pre>
```

```
bool foo(float f) {
    auto fixed = fixed_point<int, -16>{f};
    auto fixed_plus_one = fixed + 1;
    return fixed_plus_one > fixed;
}
```

```
bool foo(float f) {
    auto fixed = fixed_point<int, -16>{f};
    auto fixed_plus_one = fixed + 1;
    return fixed_plus_one > fixed;
}
```

```
bool foo(float) {
    return true;
}
```

```
bool foo(float f) {
      auto fixed = fixed point<int, -16>{f};
      auto fixed plus one = fixed + 1;
      return fixed plus one > fixed;
 bool foo(float) {
      return true;
 x86-64 gcc 4.8.1
                                -std=c++11 -O3
                        //
                                                     A▼
          .LX0:
                                         Demangle
  11010
                  .text
                             \s+
                                  Intel
   1 foo(float):
        mov eax, 1
   3
        ret
+++auto+fixed_plus_one+%3D+fixed+%2B+1%3B%0A++++return+fixed_plus_one+%3E+fixed%3B%0A%7D'),I:'5',n:'0',o:'C%2B%2B+source+%231',t:'0'),
purce:1),1:'5',n:'0',o:'x86-64+qcc+4.8.1+(Editor+%231,+Compiler+%231)',t:'0')),k:100,1:'4',m:99.9999999999999,n:'0',o:",s:1,t:'0')),version:4)
```

```
x86-64 acc 4.8.1
                                  -std=c++11 -O3
                                                                //
                                           Demangle
                                                        A▼
 11010
          .LX0:
                  .text
                              \s+
                                    Intel
   1 f():
       movsd xmm0, QWORD PTR .LC0[rip]
        mov edi, OFFSET FLAT:std::cout
       jmp std::basic_ostream<char, std::char_traits<char> >& std::basic_ostream<char, std::char_traits<char> >::_M_in
   5 i():
       movsd xmm0, QWORD PTR .LC0[rip]
        mov edi, OFFSET FLAT:std::cout
       imp std::basic ostream<char, std::char traits<char> >& std::basic ostream<char, std::char traits<char> >:: M in:
     GLOBAL sub I Z1fv:
  10
      sub rsp. 8
  11
       mov edi, OFFSET FLAT:std:: ioinit
      call std::ios_base::Init::Init()
  12
       mov edx, OFFSET FLAT:__dso_handle
  13
       mov esi, OFFSET FLAT:std:: ioinit
  14
       mov edi, OFFSET FLAT:std::ios_base::Init::~Init()
  15
  16
        add rsp. 8
       jmp __cxa_atexit
  17
  18 .LC0:
  19
       .long 0
  20
        .long 1072955392
tt+%3C%3C+float(n+*+5)%3B+//+prints+%221.25%22%0A%7D%0A%0Avoid+i()+%7B%0A++++int+n+%3D+.25f+*+256.f%3B%0A++++std::cout+%3C%3C+
fontScale:0.7464959999999999,libs:!(),options:'-std%3Dc%2B%2B11+-O3',source:1),l:'5',n:'0',o:'x86-64+gcc+4.8.1+
```

```
// undefined behavior
auto a = numeric_limits<fixed_point<int, -16>>::max() + 1;
```

```
// undefined behavior
auto a = numeric_limits<fixed_point<int, -16>>::max() + 1;
```

```
// compiles
static_assert(1 == 1, "this does compile");
// error: static assertion failed: this does not compile
static_assert(1 != 1, "this does not compile");
// error: left shift count >= width of type
static_assert(1 << 1000, "this does not compile");</pre>
```

```
// undefined behavior
auto a = numeric_limits<fixed_point<int, -16>>::max() + 1;

// compiles
static_assert(1 == 1, "this does compile");

// error: static assertion failed: this does not compile
static_assert(1 != 1, "this does not compile");

// error: left shift count >= width of type
static_assert(1 << 1000, "this does not compile");

static_assert(numeric_limits<fixed_point<int, -16>>::max() - 1, "OK");
static_assert(numeric_limits<fixed_point<int, -16>>::max() + 1, "error");
```

The Bad

```
static_assert(fixed_point<unsigned>{1} < fixed_point<signed>{-1}, "OK(!)");
```

The Bad

```
static_assert(fixed_point<unsigned>{1} < fixed_point<signed>{-1}, "OK(!)");

static_assert(numeric_limits<int>::max() + 1, "error");
static_assert(unsigned{1} < signed{-1}, "evaluates to true");</pre>
```

The Bad

```
static_assert(fixed_point<unsigned>{1} < fixed_point<signed>{-1}, "OK(!)");

static_assert(numeric_limits<int>::max() + 1, "error");
static_assert(unsigned{1} < signed{-1}, "evaluates to true");

static_assert(fixed_point{1u} < fixed_point{-1});</pre>
```

```
auto n = fixed_point<int, -8>{1.5};
auto nn = n * n;  // fixed_point<int, -16>;
static_assert(std::is_same_v<decltype(nn), fixed_point<int, -16>>);
```



```
5.5 * 5.5 = 30.25
55. * .55 = 30.25
```

```
5.5 * 5.5 = 30.25
55. * .55 = 30.25
```

```
5.5 * 5.5 = 30.25
55. * .55 = 30.25
```

```
1 / 100 = 0.01
10 / 5.5 = 1.8181818181...
```

```
5.5 * 5.5 = 30.25
55. * .55 = 30.25

1 / 100 = 0.01
10 / 5.5 = 1.8181818181...

template<typename Integer> class fraction { Integer numerator, denominator; ... };
```

```
5.5 * 5.5 = 30.25
55. * .55 = 30.25

1 / 100 = 0.01
10 / 5.5 = 1.8181818181...

template<typename Integer> class fraction { Integer numerator, denominator; ... };

AAA.BBBBB * CCCCCC.DD = AAACCCCCC.BBBBBDD
```

```
5.5 * 5.5 = 30.25
55. * .55 = 30.25

1 / 100 = 0.01
10 / 5.5 = 1.818181818181...

template<typename Integer> class fraction { Integer numerator, denominator; ... };

AAA.BBBBB * CCCCCC.DD = AAACCCCCC.BBBBBDD
AAA.BBBBB / CCCCCC.DD = AAADD.BBBBBCCCCCC
```

```
5.5 * 5.5 = 30.25
55. * .55 = 30.25
1 / 100 = 0.01
10 / 5.5 = 1.818181818181...
template<typename Integer> class fraction { Integer numerator, denominator; ... };
AAA.BBBBB * CCCCCC.DD = AAACCCCCC.BBBBBDD
AAA.BBBBB / CCCCCC.DD = AAADD.BBBBBCCCCCC
constexpr auto n = fixed_point<int, -8>{1.5};
constexpr auto d = fixed_point<int, -8>{2.25};
constexpr auto q = cnl::divide(n, d); // fixed point<long, -31>;
```

```
auto n = fixed_point<uint8_t, -8>{0.99609375};
auto nn = n * n;
```

```
auto n = fixed_point<uint8_t, -8>{0.99609375};
auto nn = n * n;  // fixed_point<int, -16>{0.9922027587890625};
```

```
auto n = fixed_point<uint8_t, -8>{0.99609375};
auto nn = n * n;  // fixed_point<int, -16>{0.9922027587890625};

auto n = fixed_point<int, -31>{0.99609375};
auto nn = n * n;  // fixed_point<int, -62>{?!?!?!?!?!};
```

```
template<int Digits, class Narrowest = int>
class elastic_integer { WideEnoughInteger r; /* other stuff */ };
```

```
template<int Digits, class Narrowest = int>
class elastic_integer { WideEnoughInteger r; /* other stuff */ };
```

```
#include <cnl/elastic_integer.h>
using cnl::elastic_integer;
auto e = elastic_integer<31>{0x7FFFFFFF}; // r has 31 or more digits
```

```
template<int Digits, class Narrowest = int>
class elastic_integer { WideEnoughInteger r; /* other stuff */ };

#include <cnl/elastic_integer.h>
using cnl::elastic_integer;
auto e = elastic_integer<31>{0x7FFFFFFF}; // r has 31 or more digits

auto ee = e * e; // elastic_integer<62>{INT64_C(0x3FFFFFFF000000001)}
```

```
template<int Digits, class Narrowest = int>
class elastic_integer { WideEnoughInteger r; /* other stuff */ };

#include <cnl/elastic_integer.h>
using cnl::elastic_integer;
auto e = elastic_integer<31>{0x7FFFFFFF}; // r has 31 or more digits

auto ee = e * e; // elastic_integer<62>{INT64_C(0x3FFFFFFF000000001)}

auto _2ee = ee + ee; // elastic_integer<63>{INT64_C(0x7FFFFFFE000000002)}
```

```
template<int Digits, class Narrowest = int>
class elastic_integer { WideEnoughInteger r; /* other stuff */ };

#include <cnl/elastic_integer.h>
using cnl::elastic_integer;
auto e = elastic_integer<31>{0x7FFFFFFF}; // r has 31 or more digits

auto ee = e * e; // elastic_integer<62>{INT64_C(0x3FFFFFFFF000000001)}

auto _2ee = ee + ee; // elastic_integer<63>{INT64_C(0x7FFFFFFE000000002)}

auto fpe = fixed_point<elastic_integer<31>, -31>{0.99609375};
```

```
template<int Digits, class Narrowest = int>
class elastic integer { WideEnoughInteger r; /* other stuff */ };
#include <cnl/elastic integer.h>
using cnl::elastic integer;
auto e = elastic integer<31>{0x7FFFFFFF}; // r has 31 or more digits
auto ee = e * e; // elastic integer<62>{INT64 C(0x3FFFFFFF000000001)}
auto 2ee = ee + ee; // elastic integer<63>{INT64 C(0x7FFFFFE000000002)}
auto fpe = fixed point<elastic integer<31>, -31>{0.99609375};
auto sq = fpe * fpe; // fixed_point<elastic_integer<62>, -62>{0.9922027587890625}
```

```
template<int Digits, class Narrowest = int>
class elastic integer { WideEnoughInteger r; /* other stuff */ };
#include <cnl/elastic integer.h>
using cnl::elastic integer;
auto e = elastic integer<31>{0x7FFFFFFF}; // r has 31 or more digits
auto ee = e * e; // elastic integer<62>{INT64 C(0x3FFFFFFF000000001)}
auto 2ee = ee + ee; // elastic integer<63>{INT64 C(0x7FFFFFE000000002)}
auto fpe = fixed point<elastic integer<31>, -31>{0.99609375};
auto sq = fpe * fpe; // fixed point<elastic integer<62>, -62>{0.9922027587890625}
#include <cnl/auxiliary/elastic fixed point.h>
auto q = sq / sq; // fixed point<elastic integer<124>, -62>{1}, q), "");
```

```
#include <cnl/safe_integer.h>
using cnl::safe_integer;
auto i = safe_integer<uint8_t>{255};
```

```
#include <cnl/safe_integer.h>
using cnl::safe_integer;
auto i = safe_integer<uint8_t>{255};

auto j = i + 1;
```

```
#include <cnl/safe_integer.h>
using cnl::safe_integer;
auto i = safe_integer<uint8_t>{255};

auto j = i + 1; // safe_integer<int>{256}
```

```
#include <cnl/safe_integer.h>
using cnl::safe_integer;
auto i = safe_integer<uint8_t>{255};

auto j = i + 1; // safe_integer<int>{256}

safe_integer<uint8_t> k = i + 1; // throw std::overflow_error
```

```
#include <cnl/safe integer.h>
 using cnl::safe integer;
 auto i = safe integer<uint8 t>{255};
 auto j = i + 1; // safe integer<int>{256}
 safe integer<uint8_t> k = i + 1; // throw std::overflow error
 constexpr safe integer<uint8 t> k = i + 1;
 static_assert(cnl:: impl::identical(safe integer<int>{256}, k));
[ 29%] Building CXX object src/test/CMakeFiles/fp test.dir/cppcon2017.cpp.o
/home/john/cnl/src/test/cppcon2017.cpp:151:37: fatal error: constexpr variable 'k' must be i
    constexpr safe integer<uint8 t> k = i + 1;
/home/john/cnl/include/cnl/overflow.h:52:40: note: subexpression not valid in a constant exp
            return condition ? value : throw std::overflow error("");
and so on...
```

Deduction and UDLs

Deduction and UDLs

```
auto x = fixed_point{42ul}; // fixed_point<unsigned long, <math>0>{42}
```

```
auto x = fixed_point{42ul}; // fixed_point<unsigned long, 0>{42}
auto y = fixed_point{128}; // fixed_point<int, 0>{1}
```

```
auto x = fixed_point{42ul}; // fixed_point<unsigned long, 0>{42}

auto y = fixed_point{128}; // fixed_point<int, 0>{1}

using cnl::literals;
auto z = fixed_point{128_c}; // fixed_point<int, 7>{128}
```

```
auto c = elastic_integer{2017_c}; // elastic_integer<11>{2017}
```

```
auto c = elastic_integer{2017_c}; // elastic_integer<11>{2017}

auto e = 0x7f000_elastic; // fixed_point<elastic_integer<7>, 12>{0x7f000}
```

Interoperability - Boost.Multiprecision

Interoperability - Boost.Multiprecision

Fixed-Point + Boost.Multiprecision:

Interoperability - Boost. Multiprecision

Fixed-Point + Boost.Multiprecision:

• googol (10¹⁰⁰) ✓

Interoperability - Boost.Multiprecision

Fixed-Point + Boost.Multiprecision:

- googol (10¹⁰⁰) ✓
- googolth (1 / googol) ✓

Interoperability - Boost. Multiprecision

Fixed-Point + Boost.Multiprecision:

- googol (10¹⁰⁰) ✓
- googolth (1 / googol) ✓
- googolplex (10^{googol}) **X**

Interoperability - Boost.SIMD

```
#include <cnl/auxiliary/boost.simd.h>
using boost::simd::pack;

template < class T, std::size_t N, int Exponent>
using fixed_point_pack = fixed_point < pack < T, N>, Exponent>;
```

Interoperability - Boost.SIMD

```
#include <cnl/auxiliary/boost.simd.h>
using boost::simd::pack;

template < class T, std::size_t N, int Exponent>
using fixed_point_pack = fixed_point < pack < T, N > , Exponent>;
```

```
using fpp = fixed_point_pack<int, 4, -16>;
using initializer = initializer<fpp>;

auto expected = fpp{initializer{7.9375+-1, -8.+.125, 0+-5, 3.5+-3.5}};
auto augend = fpp{initializer{7.9375, -8., 0, 3.5}};
auto addend = fpp{initializer{-1, .125, -5, -3.5}};
auto sum = augend + addend;
```

Interoperability - Boost.SIMD

```
#include <cnl/auxiliary/boost.simd.h>
using boost::simd::pack;

template < class T, std::size_t N, int Exponent>
using fixed_point_pack = fixed_point < pack < T, N > , Exponent>;
```

```
using fpp = fixed_point_pack<int, 4, -16>;
using initializer = initializer<fpp>;

auto expected = fpp{initializer{7.9375+-1, -8.+.125, 0+-5, 3.5+-3.5}};
auto augend = fpp{initializer{7.9375, -8., 0, 3.5}};
auto addend = fpp{initializer{-1, .125, -5, -3.5}};
auto sum = augend + addend;
```

github.com/johnmcfarlane/cnl

github.com/johnmcfarlane/cnl

• Arbitrary width

github.com/johnmcfarlane/cnl

- Arbitrary width
- Full Support for Rounding and Overflow

```
template < class Rep = int, class RoundingTag = closest_rounding_tag>
class precise integer;
```

github.com/johnmcfarlane/cnl

- Arbitrary width
- Full Support for Rounding and Overflow

```
template<class Rep = int, class RoundingTag = closest_rounding_tag>
class precise_integer;
```

Full complement of operators for safe_integer and precise_integer

github.com/johnmcfarlane/cnl

- Arbitrary width
- Full Support for Rounding and Overflow

```
template<class Rep = int, class RoundingTag = closest_rounding_tag>
class precise_integer;
```

- Full complement of operators for safe_integer and precise_integer
- Full complement of free functions

```
add(saturated_overflow, UINT32_C(0xFFFFFFFF), UINT32_C(0x12345678))
divide(closest_rounding_tag, 2, 3);
```

github.com/johnmcfarlane/cnl

- Arbitrary width
- Full Support for Rounding and Overflow

```
template<class Rep = int, class RoundingTag = closest_rounding_tag>
class precise_integer;
```

- Full complement of operators for safe_integer and precise_integer
- Full complement of free functions

```
add(saturated_overflow, UINT32_C(0xFFFFFFFF), UINT32_C(0x12345678))
divide(closest_rounding_tag, 2, 3);
```

Better Literals

```
auto a = 0b1111.1111_elastic; // fixed_point<elastic_integer<8>, -4>
```