# Communicating Sequential Processes in C++

John Skaller

April 21, 2021

# Contents

# Chapter 1

# Introduction

Blah.

# Chapter 2

# Overview

The Felix programming language uses a novel architecture which is related to Tony Hoare's communicating sequential process (CSP) model. We will document the design and implementation but our primary goal here is to allow the C++ programmer to write code conforming to the run time library (RTL) requirements to implement the abstractions.

However we will first demonstrate how the system works by progressively introducing features in pure C++, without reference to the Felix RTL. When the API is finally introduced the programmer will then be familiar with the motivation and design issues. The problem is all such architectures involve complex interactions between components which cannot be introduced sequentially; but humans learn in steps.

# Chapter 3

# Phase 1

## 3.1 Continuations

A *continuation* is an object which encapsulates the future of a program. What has already happened is called the *history,* it is represented by a sequence of significant events called a *trace.*

In a traditional system, when a subroutine is called, the caller is *suspended* with its state saved on the machine stack, and a value containing the address of the next instruction to be executed after the subroutine being called returns, is passed to that subroutine. This is called the *return address.*

When the subroutine is finished its work, it sets the program counter to the return address and the stack pointer back to the stack frame of the caller, thereby *resuming* the caller.

The callers data frame, together with the return address, is a *continuation,* and in particular, the *current continuation* at the point of the subroutine call.

## 3.2 Contracts

The caller and callee form *client-server* relationship, somewhat politically incorrectly also called a *master-slave* relationship. The relationship is established by a call with expectation a service will be perfomed based on the notion of a contract in which the duties of the parties are established, usually with *static typing* but also with *documentation.*

The client obligation is expressed by the argument types and additional constraints known as the pre-conditions, which are the expectations of the service provider. This is one side of the contract. On the other side the client expects a certain result known as the postconditions to be met and the service provider,

in accepting arguments conforming to the pre-condition agreement, accepts the responsibility to satisfy them. The postcondition usually consists of a return value of a particular type also meeting additional constraints, together with a performance expectation. For a functions, the minimal expectation is that the subroutine actually completes its task, which is known as *termination*.

## 3.3 Enforcement

In many systems machinery known as a static type system is used to partially enforce the terms of a contract as follows: if a violation of the contract is discovered by the language translator, transation is terminated with a violation report so that the programmer or programmers responsible for the encodings can share a coffee and discuss how to correct the deficiency.

It may the caller has passed the wrong arguments, but the contract itself may be inappropriate.

Enforcement of other constraints can be attempted with dynamic checks and the flow of control annotated with event information in debug traces.

### 3.3.1 Proof

A type system is implemented with an algorithm that attempts to prove correct typing. An contract violation is reported if the system is unable to construct a proof. In simple systems the proof is usually nothing more than a check that the argument types of a subroutine call match the specified types of the parameters; and, the type of the return value matches the specified return type. In the first case, *blame* is layed on the caller, and thus its author, whereas in the latter, it is layed on the subroutine, and thus its author.

In this way the system can be used to select the author who may be responsible for correcting the defect; remember, however, that the contract itself may be inappropriate: perhaps the caller was right and the subroutine should have accepted the given arguments after all?

In more advanced systems like C++, argument and parameter types do not have to be identical. It is acceptable that the argument be a subtype of the specified parameter type, due to the notion that a value which a type which is a subtype of another, is also a value of that wider type. Furthermore, with the notion of overloading, multiple functions accepting different argument types might be provided. If a suitable function is not found for given arguments, perhaps the caller should choose different arguments, but perhaps instead a new overload is required.

### 3.3.2 Design by proof

Type systems only provide a low level of contract enforcement. In particular, the absence of a type error does not prove the overall contract of the program is satisfied.

A situation where a high level contract is satisfied is known as *correctness.* If a program is incorrect, it is possible to provide a proof with a counterexample, that is, an input for which an acceptable answer is not provided. Often, this will be because no answer is provided at all because the program goes into an infinite loop or crashes.

Proving a program or subroutine is correct is often difficult. The key idea here, which cannot be overstated, is that your design, and coding, should be *driven by a the idea of a proof*, even if your proof is not written out. Programming by contract is a methodology in which you sketch a proof by breaking the proof down into a set of discrete but interrelated sub-proofs.

In mathematics, these sub-proofs are called lemmas or theorems.

## 3.4 Peer to Peer Design

Whilst subroutines and the correpsonding master-slave relationships provide concept of modularity which allows breaking down a problem into a set of service provides, each with its own contract, thereby facilitating *reasoning* about a program; complex relationships built entirely this way can be difficult to manage.

A good program also contains modules which have *peer-to-peer* relations, in which each party is a willing and voluntary participant. The parties both accept some responsibility and each has the authority and capability to carry out their part of the bargain. In return, each expects the other party to honour its committments.

The peer-to-peer unit of modularity corresponding to a subroutine is called a *coroutine.* There are various methods by which coroutines can cooperate but the one we will focus on is a device called a channel. The idea of a channel is that data can be read or written by coroutines on channels, and other coroutines sharing access to these channels can write data to service a reader, whilst a reader can read it, to service a write request. A coroutine can access multiple channels, and so can play the role of a reader at one point, and a writer at another.

Coroutines are notionally infinite loops that process streams of data, whilst channels connections determine the data flow topology. Along with this new model, we also need new kinds of contracts. Using type systems to an obvious approach, but this is a new and open research field; the types requires are often calles *session types* because they specify the contracts for a series of interactions.

We will start our development by showing coroutines *subsume* subroutines. This means that anything you can do with subroutines can also be done with coroutines. I will state now quite clearly that this does *not* all subroutines should be replaced. Contrarily, the programmer how has two models of computation that can be used together.

Indeed, starting with C++ our primary coding machinery uses the subroutine model and we will have to use that to implement coroutines.

## 3.5 Emulating Subroutines

We are going to develop a technique in which a heap allocated frame is used to hold local data instead of the machine stack, these frames will be linked with pointers forming a so-called *spaghetti stack*. These objects will be continuations. The reason for doing this will become apparent later. Suffice it to say that context switching between threads of control is done by stack swapping, and user space switching between threads using a spaghetti stack can be done extremely fast by simply swapping pointers. By contrast, swapping machine stacks during pre-emptions of hardware threads is extremely expensive, the number of allocable threads is strictly limited by the operating system, and the amount of memory used for a thread is high, even if it is not actually used, but the cost in address space reserved for each pthread is extreme. This is because with linear addressing, the address space for the maximum possible stack size must be pre-allocated, even if the actual memory is only assigned to it on demand. User space threads, called *fibres* using spaghetti stacks, impose no such demands: one pays, instead, an extra cost allocating and deallocating heap frames, in return for lightning fast context switching and minimal use of both memory and address space.

### 3.5.1 Emulating Subroutines

Our first task is to emulate subroutine operation. Consider the following class representing the base of a continuation:

```
struct con_t {
  con_t *caller; // caller continuation
  int pc;        // program counter
};
```

Now we will make a simple subroutine:

```
struct hello : con_t {
  int num;

  con_t *call(con_t *cc, int n) {
```

```
      caller = cc;
      num = n;
      return this;
    }

    con_t * resume() {
      ::std::cout << "Hello " << num << ::std::endl;
      auto tmp = caller;
      delete this;
      return tmp;
    }
  };
```

It is important to note that we have split control flow into two parts.

In the `call` function we initialise the object with the callers current continuation and the argument of the call. Passing the callers current continuation explicitly and saving it is essential to the construction of a spaghetti stack. We have created subroutine but we have *suspended* execution until later.

In the `resume` method we actually execute the suspension, resuming it where it left off, or, in this case where it needs to start. When we're finished, we return the callers continuation.

Now lets write another procedure which calls our `hello` routine. We're going to call it 10 times, passing a counter than ranges from 1 to 10.

```
  struct doit: public con_t {
    int counter;

    con_t *call(con_t *cc) {
      caller = cc;
      pc = 0;
      return this;
    }

    con_t *resume() {
      switch(pc){

      case 0:
        counter = 1; // init local variable

      case 1:
        pc = 2;
        return (new hello) -> call(this,counter);
```

```
      case 2:
        ++counter;
        if(counter <= 10) {
          pc = 1;
          return this;
        }
        {
          auto tmp = caller;
          delete this;
          return tmp;
        }
      } // end switch
    } // end resume
  } // end doit
```

The `call` function implements a subroutine call to the `doit` procedure but leaves it in a suspended state, ready to run, starting at its entry point. The entry point is always given by setting the program counter `pc` to 0. The caller's continuation `cc` is passed in, it will be `nullptr` if this is the top level procedure.

Local variables such as `counter` are always represented by a non-static member. Function local automatic variables are not used because these cannot be preserved when the procedure is suspended.

To actually use this system we need a driver. So here's out mainline:

```
#include <iostream>
// put code here
int main() {
  doit *program  = new doit;
  con_t *p = program->call(nullptr);
  while(p)p=p->resume(); // driver
  return 0;
}
```

We first construct a continuation object for our top level procedure `doit` and then initialise it with its `call` method by passing the current continuation, which is `nullptr` because this is the top level procedure. The `while` loop which is running our code will terminate when a `nullptr` is returned.

Inside the `doit::resume` method we want to start at the entry point, `case 0` by initialising the loop control variable `counter` to 1.

Then we call the `hello` subroutine, passing it the counter. We also have to pass it our `this` pointer and set our `pc` value to the next code to execute at `case 2`.

So the steps in a subroutine call require:

1. Set our `pc` to the code we want to execute after the call
2. Construct the target subroutine object
3. Initialise the target routine by passing it our `this` pointer as the caller continuation
4. Also pass any arguments of the call
5. Return the pointer of the subroutine suspension we just create to the driver

The driver now resumes the subroutine at its entry point and steps through it.

To return from a subroute we simply return the saved `caller` to the driver which resumes stepping through it at the case it saved in its `pc` variable, and also delete the object.

Importantly all local variables of the caller are preserved because they're all non-static members of the C++ structure.

In the code above we repeatedly heap allocate the subroutine, and delete it again, on each loop iteration. This emulates the pushing and popping of frames on the machine stack, using the heap allocated frames a spaghetti stack uses.

The performance overhead in this case can be reduced by only allocating the subroutine frame once, and that will work .. in this case. But then, you could also just use a standard C++ procedure on the machine stack, in this case.

The Felix compiler performs these optimisations by static analysis, the C++ programmer using the Felix run time can perform them by hand. Felix also use a garbage collector so that pointers to variables in frames of objects cannot dangle, even if the procedure has completed.

But by now you're asking, why should we program this way, ever?

### 3.5.2 Interleaving

In order to grasp the advantages of our system, suppose we would like to run 100 copies of our `doit` procedure concurrently. If you use the system Operating System machinery you can do this by creating 100 pre-emptive threads all running the code at once.

If you have only, say, 4 CPUs, the OS can only execute 4 routines simultaneously, so it occasionally pre-empts one of the running threads, saves its state, and starts the CPU running the code of some other thread from where it was previously pre-empted. This is called *pre-emptive multi-tasking*.

The state of a conventional thread is stored in the machine stack, machine registers, and heap objects for which there are pointers in these objects (recursively). So the OS has to save a machine stack pointer and other registers, and set new values to restart another thread.

This may be OK but such control exchanges are slow and expensive, and they occur at random times outside program control, and so require special synchronisation devices such as locks, which are also expensive. However by far the biggest problem is that machine stacks are allocated in pages which are typically 4K bytes. With linear addressing the problem is much worse: the maximum possible size of a machine stack must be calculated and address space reserved for all of it. This is fine for 100 threads.

But what if you have 100 million threads? Your operating system will probably crash because it cannot handle that number of threads, and your virtual memory system will be madly paging memory off disk, to reload the data of a machine stack on every swap. Although 64 bit machines have a lot of address space, a lot is wasted due to the 4K granularity.

By contrast, swapping spaghetti stacks can be done in user space by swapping a single pointer, and stacks only use the amount of space they need because they're allocated on the heap, which typically has 16 byte granularity, a lot less than 4K bytes. In addition with coroutine architecture, interleaving only occurs cooperatively, eliminating the need for a lot of expensive synchronisation devices.

We are going to first demonstrate interleaving in a rather silly way, as a prelude to introducing a real scheduler. Our silly scheduler will just execute things in a round robin order. We do a bit of each active fibre of control in turn until all are completed.

We will use a C++ queue object to do this.

```
int main() {
  auto q = ::std::queue<con_t*>;

  // initialise queue
  for (int i=0; i<100; ++i){
    doit *program  = new doit;
    con_t *p = program->call(nullptr);
    q.push(p);
  }

  // execute
  while (!q.empty()) {
    con_t *p = q.pop();
    p = p->resume();
    if(p) q.push(p);
  }
  return 0;
}
```

Wow! We just implemented a *cooerative multi-tasking* scheduler! It runs steps of a set of fibres until there is no work left to do. It uses a round-robin scheduling strategy which attempts to be "fair", executing one "step" of each thread in turn.

# Chapter 4

# Phase 2

### 4.0.1 Faster Please

The round robin scheduler above using a C++ standard library queue is very slow. The reason is the library routine will allocate nodes in a doubly linked list whenever you push onto the queue, and delete one, whenever you pop.

To fix this problem we will add a new object representing a fibre:

```
struct fibre_t {
  con_t *cc;
  fibre_t *next;
  fibre_t() : cc(nullptr), next(nullptr) {}
  fibre_t(con_t *ccin) : cc(ccin), next (nullptr) {}
};
```

Now here is our new mainline:

```
int main() {
  fibre_t *current  = nullptr;

  // initialise queue
  for (int i=0; i<100; ++i) {
    doit *program  = new doit;
    con_t *p = program->call(nullptr);
    current = fibre_t (p, current);
  }

  // execute
  while (current) {
```

```
      con_t *p = current->cc;
      while(p)p=p->resume();
      fibre_t *tmp = current;
      delete current;
      current = tmp;
    }
    return 0;
  }
```

Our code uses a stack protocol for the active fibres rather than the queue our previous example used. Why? The reason is simple: we don't care about ordering but we care about speed. Less instruuctions are required to push and pop from a singly linked list.

We have paid a small price, we have to allocate an extra object, the fibre, and find the top continuation of the fibre with an extra level of indirection. But then the processing loop is the same.

Lets encapsulate the fibre driver in a method:

```
  struct fibre_t {
    con_t *cc;
    fibre_t *next;
    fibre_t() : cc(nullptr), next(nullptr) {}
    fibre_t(con_t *ccin) : cc(ccin), next (nullptr) {}
    void run_fibre() { while(cc)cc=cc->resume(); }
  };
```

Lets encapsulate the operation in a new object, a scheduler:

```
  struct sync_sched {
    fibre_t * current;
    fibre_t *active;

    scheduler() : current(nullptr), active(nullptr) {}
    void push(fibre_t *fresh);
    fibre_t *pop();
    void sync_run();
  };
```

Here are the methods; push:

```
    void sync_sched::push(fibre_t *fresh) {
      fresh->next = active;
```

```
      active = fresh;
    }
```

pop:

```
    fibre_t *sync_sched::pop() {
      fibre_t *tmp = active;
      active = active->next;
      return tmp;
    }
```

and the scheduler itself:

```
    void sync_run() {
      while(active) {
        current = active;
        active = active-> next; // pop
        current->run_fibre();
        delete current;
      }
    }
  };
```

## 4.1 Channels

Finally we're about to discover the whole reason for the model we've developed: the use of a device called a synchronous channel for communication. A channel is nothing more than a list of fibres which is either empty, or all the fibres are waiting to read, or all waiting to write.

Here's the overall data type:

```
  struct channel_t {
    fibre_t *next;
    channel_t () : next (nullptr) {}
    ~channel_t()[
    void push_reader(fibre_t r);
    void push_writer(fibre_t w);
    fibre_t *pop_reader();
    fibre_t *pop_writer();
  };
```

and the destructor:

```
~channel_t::channel_t() {
  while (next) {
    fibre_t *f = (fibre_t*)*(unitptr_t)next & ~(unitptr_t)1u);
    fibre_t *tmp = f->next;
    delete f;
    next = tmp;
  }
}
```

Now the methods to pop a reader from the channel:

```
void channel_t::push_reader(fibre_t r) {
  r->next = next;
  next = r;
}
```

and push a writer onto the channel:

```
void channel_t::push_writer(fibre_t w) {
  w->next = next;
  next = (fibre_t*)((uintptr_t)w & (uintptr_t)1u);
}
```

to pop a reader:

```
fibre_t *channel_t::pop_reader() {
  fibre_t *tmp = next;
  if((uintptr_t)tmp & (unitptr_t)1u) return nullptr;
  next = next->next;
  return tmp;
}
```

and to pop a writer:

```
fibre_t *channel_t::pop_writer() {
  fibre_t *tmp = next;
  if(!((uintptr_t)tmp & (unitptr_t)1u)) return nullptr;
  next = (fibre_t*)next->next;
  return (fibre_t*)((uintptr_t)tmp & ~(uintptr_t)1u);
}
```

The casts there are because we're doing a hack! We're stealing the low bit of the fibre pointer and using it as a flag to determine if the channel is holding

readers or writers. If the pointer is not null, and the low bit is 0, the channel holds readers, if the low bit is 1, the channel holds writers.

If we try to pop a reader from a channel containing writers, we get a nullptr back. Similarly if we try to pop a writer from a channel containing readers we get a nullptr back. Of course if the channel is empty we also get a nullptr back.

Note that the push operations have preconditions: to push a reader, the channel must be empty or contain only readers; to push a writer, the channel must be empty or contain only writers. We should prove these conditions are met at every point these methods are invoked.

Now, we have to actually code the I/O operations. Unfortunately whilst the rules are simple, the implementation is a bit tricky!

We need to reserve a slot in a continuation for the I/O request: there are two requests, one is to read, the other to write.

The operation is that, if a write is requested, the data in the data slot we added will be written to a reader on the channel, if there is one, otherwise the writer suspended and is added to the channel.

If a read is requested, the data from the data slot of a writer on the channel is moved to the reader, if there is one, otherwise the reader is suspended and is added to the channel.

If the I/O request is satisfied, the reader or writer on the channel is removed from the channel and now both the reader and writer are added to the active list of the scheduler, because they're ready to proceed. The scheduler then picks a new suspended fibre to run, removing it from the active list.

In principle the scheduler can pick any active fibre to proceed with. As an optimisation and to make it more intuitive, we will always pick the reader to continue after an I/O operation. Since the reader is only getting a single machine word, it has a chance to cast it to a pointer to the actual data to be transfered and copy it if necessary, before the writer can clobber it with new data.

Now we need to design the service requests. We will use an enum to specify the request types:

```
enum svc_code_t {
  read_request_code_e,
  write_request_code_e,
  spawn_fibre_request_code_e
};
```

I have thrown in a third request: spawn. This is a request to push a new fibre on the scheduler active list. We need this request because the code in a continuation does not have direct access to the scheduler! In fact, it does not have access to the fibre it is top of either!

Now we need the data packets for the requests. An I/O request needs to tell the system two things: the channel on which to perform the request, and, the location into which to put, or from which to get, the word to be transfered:

```
struct io_request_t {
  svc_code_t svc_code;
  channel_t *chan;
  void **pdata;
};
```

For the spawn:

```
struct spawn_fibre_request_t {
  svc_code_t svc_code;
  fibre_t *tospawn;
};
```

Finally we have to write up these data structures into a variant:

```
union svc_req_t {
  io_request_t io_request;
  spawn_fibre_request_t spawn_fibre_request;
  svc_code_t get_code () const { return io_request.svc_code; }
};
```

Note: technically, the `get_code` method could violate the C++ Standard rules by accessing the wrong component. It would be more correct to separate the code out from the I/O request packet types, and make a union of them, then make a struct with the code as the first member and the union as the second. The problem is it is hard to construct such an object. It is easier to cheat, and construct a simple request object and cast it to the union type, which is again breaking the rules.

We have to extend the continuation to hold a request:

```
// continuation
struct con_t {
  con_t *caller; // caller continuation
  int pc;        // program counter
  union svc_req_t *svc_req; // request
  virtual con_t *resume()=0;
  virtual ~con_t(){}
};
```

and now the fibre has to handle it by returning it:

```
// fibre
struct fibre_t {
  con_t *cc;
  fibre_t *next;

  // default DEAD
  fibre_t() : cc(nullptr), next(nullptr) {}

  // construct from continuation
  fibre_t(con_t *ccin) : cc(ccin), next (nullptr) {}

  // immobile
  fibre_t(fibre_t const&)=delete;
  fibre_t& operator=(fibre_t const&)=delete;

  ~fibre_t();
  svc_req_t *run_fibre();
};


The destructor deletes any remaining continuations in spaghetti stack

\begin{minted}{c++}
  ~fibre_t::fibre_t() {
    while(cc) {
      con_t *tmp = cc->caller;
      delete cc;
      cc = tmp;
    }
  }
```

run until either fibre issues a service request or dies

```
  svc_req_t *fibre_t::run_fibre() {
    while(cc) {
      cc=cc->resume();
      if(cc && cc->svc_req) return cc->svc_req;
    }
    return nullptr;
  }
```

Here's our new scheduler:

```
// scheduler
struct sync_sched {
  fibre_t *current; // currently running fibre, nullptr if none
  fibre_t *active;  // chain of fibres ready to run

  sync_sched() : current(nullptr), active(nullptr) {}

  // push a new active fibre onto active list
  void push(fibre_t *fresh) {
    fresh->next = active;
    active = fresh;
  }
  // pop an active fibre off the active list
  fibre_t *pop() {
    fibre_t *tmp = active;
    if(tmp)active = tmp->next;
    return tmp;
  }
  void sync_run();
  void do_read(io_request_t *req);
  void do_write(io_request_t *req);
  void do_spawn_fibre(spawn_fibre_request_t *req);
};
```

Now the scheduler has to handle the service request.

```
// scheduler subroutine runs until there is no work to do
void sync_sched::sync_run() {
  current = pop(); // get some work
  while(current) // while there's work to do
  {
    svc_req_t *svc_req = current->run_fibre();
    if(svc_req)  // fibre issued service request
      switch (svc_req->get_code())
      {
        case read_request_code_e:
          do_read(&(svc_req->io_request));
          break;
        case write_request_code_e:
          do_write(&(svc_req->io_request));
          break;
        case spawn_fibre_request_code_e:
          do_spawn_fibre(&(svc_req->spawn_fibre_request));
          break;
```

```
        }
      else // the fibre returned without issuing a request so should be dead
      {
        assert(!current->cc); // check it's adead fibre
        delete current;       // delete dead fibre
        current = pop();      // get more work
      }
    }
  }
```

The new methods now need to be written. Here's the read operation:

```
  void sync_sched::do_read(io_request_t *req) {
    fibre_t *w = req->chan->pop_writer();
    if(w) {
      *current->cc->svc_req->io_request.pdata =
        *w->cc->svc_req->io_request.pdata; // transfer data

      // null out svc requests so they're not re-issued
      w->cc->svc_req = nullptr;
      current->cc->svc_req = nullptr;

      push(w); // onto active list
      // i/o match: reader retained as current
    }
    else {
      req->chan->push_reader(current);
      current = pop(); // active list
      // i/o fail: current pushed then set to next active
    }
  }
```

The write operation is similar but differs because we always want the reader to be active:

```
  void sync_sched::do_write(io_request_t *req) {
    fibre_t *r = req->chan->pop_reader();
    if(r) {
      *r->cc->svc_req->io_request.pdata =
        *current->cc->svc_req->io_request.pdata; // transfer data

      // null out svc requests so they're not re-issued
      r->cc->svc_req = nullptr;
      current->cc->svc_req = nullptr;
```

```
      push(current); // current is writer, pushed onto active list
      current = r; // make reader current
    }
    else {
      req->chan->push_writer(current); // i/o fail: push current onto channel
      current = pop(); // reset current from active list
    }
  }
```

and finally the spawn is simple, we will make the spawned fibre current to emulate the behaviour of a subroutine call:

```
  void sync_sched::do_spawn_fibre(spawn_fibre_request_t *req) {
    current->cc->svc_req=nullptr;
    push(current);
    current = req->tospawn;
  }
```

### 4.1.1 Contracts

in all the above code the methods have pre-conditions and post-conditions and the objects have public invariants; we have not stated these. These things can be said to be the *terms* of a *contract*.

In production code, we should always state the contract terms, but we should go further: we should *prove*, or at least outline a proof, that the terms of the contract are adhered to, where we have written both parties of the contract.

For a subroutine call, we should prove the caller meets the pre-condition, for an object, that the invariants are maintained. Sometimes this is trivial: for example, the `pop_reader` function may return a `nulptr`. Did we handle both the null and non-null cases? A quick inspection shows `pop_reader` is only called in one place in `do_write`, and we have indeed checked if it has returned null, and handled both the null and non-null cases. In turn, the dereference of the returned pointer is safe, because it is inside the scope of the branch handling the non-null case.

It is less obvious that `current` cannot be null, and if it is, that the current continuation `cc` of the indicated fibre must also be non-null.

How can we be sure we have an exhaustive proof? Our code is simple enough to elaborate all possible control flow paths, and annotate each point on the path with the state of the important variables.

There is a problem with contract annotations: they make the code more verbose, which can make it harder to understand. On the other hand the specification

of the contract terms can also aid comprehension.

The canonical exemplar of this issue is the tension between dynamically typed languages like Python, in which arguments, return values, and variables do not carry type annotations, and languages like C++ where static type annotations can be used. In particular the use of the `auto` keyword in C++ for local variables is common: it improves comprehensibility by removing the need to write messy type annotations, at the expense of needing to calculate from non-local information what the type would be. For the compiler, this also removes a source of error checking.

We have not used auto in our code, but this is because our types are all fairly simple and it's useful to see exactly what they are: in general it is reasonable to use it for local variables where the type is easily deduced from lexically close code.

## 4.2 Using channels

Since we have now established all the code needed to service I/O requests we need to discover how into perform the requests and how to use the I/O as a synchronisation device. A continuation with a `resume` method that performs, directly or indirectly, a service request is said to be a *coroutine*. If no service requests can be performed executing the resume method, the continuation is a trivial coroutine or subroutine. All the routines we have presented so far are procedures, that is, none of them return a value, instead they rely on effects for their utility.

A routine can perform a service call indirectly if it invokes a coroutine, or a routine which invokes a coroutine .. and so on. Therefore being a proper coroutine depends on the transitive closure of all routines that a given routine can call.

There is a cruicial distinction between proper coroutines and procedures: mere procedures can be optimised to put their objects on the machine stack, because they cannot be suspended. In fact, an ordinary C procedure can be used, which can be even more efficient.

Now we need to design a little demo which uses channels. We will start with something apparently simple which can easily be done in C++ already: we will map a list of integers to a map of their squares. This will involve three coroutines and two channels.

The first coroutine accepts a C++ list of integers and writes them down a channel.

```
struct producer : con_t {
  ::std::list<int> *plst;
  ::std::list<int>::iterator it;
```

```
    channel_t *chan;
    union {
      void *iodata;
      int value;
    };
    io_request_t w_req;

    con_t *call(
      con_t *caller_a,
      ::std::list<int> *plst_a,
      channel_t *outchan)
    {
      caller = caller_a;
      plst = plst_a;
      pc = 0;
      w_req.chan = outchan;
      return this;
    }

    con_t *resume() override {
      switch (pc) {
        case 0:
          it = plst->begin();
          pc = 1;
          w_req.svc_code = write_request_code_e;
          w_req.pdata = &iodata;
          return this;

        case 1:
          if(it == plst->end()) {
            auto tmp = caller;
            delete this;
            return caller;
          }
          value = *it++;
          svc_req = (svc_req_t*)(void*)&w_req; // service request
          return this;
        default: assert(false);
      }
    }
  };
```

It is intersting to note that, because the pc is not reset at the end of case 2, it remains at the value 1, therefore each time the resume method is called after first entering case 1 branch, it will be called again from the case 1 entry point,

effecting a loop.

Note also we did not delete the object after it returns. We will see how this is handled and the constraints that imposes when we write the mainline.

I have used a union to hold the int to be transfered, aliased with the actual type required. The assumption here is that the int is not larger than a `void*`. This saves heap allocating the integer and passing a pointer, which avoids a memory management hassle. Again, technically, our code is ill formed because we're moving a different union component from the one we stored but it will work.

The consumer code is similar, but it constructs a list from its inputs instead.

```
struct consumer: con_t {
  ::std::list<int> *plst;
  union {
    void *iodata;
    int value;
  };
  io_request_t r_req;

  con_t *call(
    con_t *caller_a,
    ::std::list<int> *plst_a,
    channel_t *inchan_a)
  {
    caller = caller_a;
    plst = plst_a;
    r_req.chan = inchan_a;
    pc = 0;
    return this;
  }

  con_t *resume() override {
    switch (pc) {
      case 0:
        pc = 1;
        r_req.svc_code = read_request_code_e;
        r_req.pdata = &iodata;
        return this;

      case 1:
        svc_req = (svc_req_t*)(void*)&r_req; // service request
        pc = 2;
        return this;

      case 2:
```

```
          plst->push_back(value);
          pc = 1;
          return this;
        default: assert(false);
      }
    }
  };
```

You will see something very interesting here: this code is an infinite loop! This is correct. It is the usual case for coroutines! You may wonder how the loop terminates. We will soon see, but the answer briefly is that it does not terminate, it simply fails to resume.

Now we will implement a transducer which reads an integer and writes its square. There's a bit more housekeeping because it uses both an input and output channel.

```
  struct transducer: con_t {
    union {
      void *iodata;
      int value;
    };
    io_request_t r_req;
    io_request_t w_req;

    con_t *call(
      con_t *caller_a,
      channel_t *inchan_a,
      channel_t *outchan_a)
    {
      caller = caller_a;
      r_req.chan = inchan_a;
      w_req.chan = outchan_a;
      pc = 0;
      return this;
    }

    con_t *resume() override {
      switch (pc) {
        case 0:
          pc = 1;
          r_req.svc_code = read_request_code_e;
          r_req.pdata = &iodata;
          w_req.svc_code = write_request_code_e;
          w_req.pdata = &iodata;
```

```
          return this;

        case 1:
          svc_req = (svc_req_t*)(void*)&r_req; // service request
          pc = 2;
          return this;

        case 2:
          value = value * value; // square value
          svc_req = (svc_req_t*)(void*)&w_req; // service request
          pc = 1;
          return this;
        default: assert(false);
      }
    }
  };
```

Now we need to set up the system to test it.

```
  int main() {
    // create the input list
    ::std::list<int> inlst;
    for (auto i = 0; i < 20; ++i) inlst.push_back(i);

    // output list
    ::std::list<int> outlst;

    // create scheduler
    sync_sched sched;

    // create channels
    channel_t chan1;
    channel_t chan2;

    // create fibres
    fibre_t *prod = new fibre_t ((new producer)->call(nullptr, &inlst, &chan1));
    fibre_t *trans = new fibre_t ((new transducer)->call(nullptr, &chan1, &chan2));
    fibre_t *cons = new fibre_t ((new consumer)->call(nullptr, &outlst, &chan2));

    // push initial fibres onto scheduler active list
    sched.push(prod);
    sched.push(trans);
    sched.push(cons);

    // run it
```

```
    sched.sync_run();

  ::std::fflush(stdout);

    // the result is now in the outlist so print it
    ::std::cout<< "List of squares:" << ::std::endl;
    for(auto v : outlst) ::std::cout << v << ::std::endl;
  }
```

The key thing here is to note that the `sync_run` function only returns when there is no work to do. Any fibre that terminated is deleted by the scheduler. Those fibres that remain hanging on channels will be deleted by the channel destructors.

When a produce is connected to a series of transducers and terminated by a consumer, the structure is called a *pipeline*. Pipelines have some nice properties, including the fact that connection is associative. What this means is that, for example, a producer connected to a transducer is a producer, two transducers connected together are a transducer, and a transducer connected to a consumer is a consumer. This means you can defined new pipeline components by connecting arbitrary components in the sequence into new components, so the system is highly modular.

A coroutine which depends only on data it reads and writes from channels, and does not modify its environment, is said to be *pure*. Pure coroutines are referentially transparent. We can count the producer as pure, assuming the input list bound to it is not modified, and the transducer is pure, however the consumer is not.

It is useful to note that a pipeline is semantically equivalent to a monad. You can see from even this small example coroutines excel at handling streams: the producer in the example is converting a spatial list to a temporal stream.

## 4.3 Control inversion and peers

The most important thing about using coroutines is that they dispense with the master/slave relationship enforced by subroutine calling and stacks. Instead, coroutines are peers. This can only be done by fast context switching which is precisely what our system provides.

It is vital to understand why peer to peer relations are better. Consider the methods for mapping a list of integers to a list of its squares. There are two established methods for doing this with subroutines.

The first method is to use a higher order function called map that accepts the list and a per element processing function, and returns a new list. The processing

function is a callback slave of the master map function. The map has to scan the list and construct a new one, which is easy to do because it is the master and gains the advantage of structured programming, that is, the implicit coupling of parameter passing and control flow. But the slave callback is hard to write in complex applications because it keeps losing its state.

The second method is to use an iterator as a slave to scan the list, and possibly construct a new one, and write a loop to drive it that does the application dependent work. This is much better for the client programmer than writing a callback, because structured programming tracks the state. However the iterator is a callback, and is therefore hard to write for complex data structures.

What we actually want is that *both* the iteration and client computations think they're masters, so they are both easy to program, and this is precisely what our system provides: all the components are coroutines which read and write data.

When a master is converted to a slave, or vice-versa, this is called *control inversion*. The master/slave relation is also known as client/server or push/pull. It is the bane of a programmers existence to have to take a master routine and control invert it into a callback. The reason is that the structured programming advantage of implicit coupling of control flow and data stacking is lost in a callback, and has to be manually put back. In our system, all routines are masters in effect, but we have to use disciple to implement the patterns for operations like subroutine calling and returning. This is the price, along with the need to use heap allocations instead of the much faster machine stack, in order to support high speed context switching.

Although our system works and is extremely fast, it retains some of the problems of a typical stack machine which also has pointers: pointer to data in a returned continuation will dangle, because the continuation has been deleted. We will look later at some ways to solve this problem. One method is obvious: use a garbage collector. The current Felix run time system does precisely that and ensures no pointers can dangle. Garbage collection is fast and achieves better thruput than other methods, and is fully general. However naive collectors are not compatible with real time operation.

## 4.4 Thread Safety

The code we have presented so far is not thread safe. There are two things we want to do. The first is to ensure two threads running separate schedulers allow fibres to communicate with channels. The second thing we want to do is allow multiple threads to service the same active list, in effect providing a thread pool to execute queued jobs.

The first and easiest thing to do is split up the scheduler.

```
// active set
struct active_set_t {
  ::std::atomic_size_t refcnt;
  fibre_t *active;
  ::std::atomic_flag lock;
  active_set() : refcnt(1), active(nullptr), lock(ATOMIC_FLAG_INIT) {}

  active_set_t *share() { ++refcnt; return this; }
  void forget() { --refcnt; if(!atomic_load(&refcnt)) delete this; }

  // push a new active fibre onto active list
  void push(fibre_t *fresh) {
    while(lock.test_and_set(::std::memory_order_acquire); // spin
    fresh->next = active;
    active = fresh;
    lock.clear(::std::memory_order_release); // release lock
  }
  // pop an active fibre off the active list
  fibre_t *pop() {
    while(lock.test_and_set(::std::memory_order_acquire); // spin
    fibre_t *tmp = active;
    if(tmp)active = tmp->next;
    lock.clear(::std::memory_order_release); // release lock
    return tmp;
  }
};
```

We have provided an atomic flag we use as a spinlock in the push an pop operations. There is also a reference count. When a new scheduler is created, typically in a new thread, which is to share the active set, the share function is use to obtain a pointer to it, which increments the reference count. The initial creator gets a pointer to a fresh scheduler with the counter already set to 1.

When any scheduler exits or otherwise wishes to relinquish access to the active set, it must call forget. This includes the initial creator. If there are no references left to the active set, it is deleted.

Note that a corouine which somehow managed to get hold is the active list on which it is waiting, this would create a circularity. This would only cause a problem is the coroutine failed to reach the point where it forgot the active list. This could happen if the coroutine became suspended on a channel, its owning scheduling process returned, and the coroutines destructor also failed to invoke forget.

Now before we proceed we must make a decision about a new issue that arises when separate threads, running separate schedulers, meet via a successful I/O

operation. In the single threaded case, we simply pushed the writer on the active list, and executed the reader.

But now, we could execute both concurrently, or we could push both onto an active list, but which one?

The simplest answer is that when a fibre is created it is associated with a particular active set, so it will always be either put on the active set, or executed by a thread accessing that active set. To make this happen, we have to save a pointer to the active set in the fibre:

```
// fibre
struct fibre_t {
  con_t *cc;
  fibre_t *next;
  active_set_t *owner;
...
};
```

The scheduler has to be changed to this:

```
// scheduler
struct sync_sched {
  fibre_t *current; // currently running fibre, nullptr if none
  struct active_set_t *active_set;  //

  sync_sched() : current(nullptr), active(nullptr) {}
  ~sync_sched() { active_set->forget(); }

  void sync_run();
  void do_read(io_request_t *req);
  void do_write(io_request_t *req);
  void do_spawn_fibre(spawn_fibre_request_t *req);

};
```

Note there is one scheduler per thread.

Now, we need the channel I/O operators to be mostly safe. To achieve this we have little choice but to add a lock to channels.

```
// channel
struct channel_t {
  fibre_t *top;
  ::std::atomic_flag lock;
```

```
    channel_t () : top (nullptr), lock(ATOMIC_FLAG_INIT) {}
...
};
```

Now comes the hard bit. It would be tempting to lock the push and pop operations on the channel, but this will not do. Recall a reader, for example, tries to get a writer, and gets pushed if there isn't one. The problem is, a writer from another thread may get pushed in between the check and the push, and now we have broken a core invariant: a channel must be empty, contain only readers, or contain only writers. So we have to acquire a lock at the start of the read operation and release it only when it is complete.

```
void sync_sched::do_read(io_request_t *req) {
  while(lock.test_and_set(::std::memory_order_acquire); // spin
  fibre_t *w = req->chan->pop_writer();
  if(w) {
    lock.clear(::std::memory_order_release); // release lock
    *current->cc->svc_req->io_request.pdata =
      *w->cc->svc_req->io_request.pdata; // transfer data

    // null out svc requests so they're not re-issued
    w->cc->svc_req = nullptr;
    current->cc->svc_req = nullptr;

    w->owner.push(w); // onto active list
    // i/o match: reader retained as current
  }
  else {
    req->chan->push_reader(current);
    lock.clear(::std::memory_order_release); // release lock
    current = active_set->pop(); // active list
    // i/o fail: current pushed then set to next active
  }
}
```

The location of the release is a bit surprising, it comes before the data transfer. This is ok because the reader and writer are both suspended. The writer is suspended on the channel, and the reader is suspended in the function itself.

Notice also, the writer is pushed back onto its owner active list.

The pushes and pops on active lists are separately locked. It takes some thought to understand why this is enough: there is no need for the whole method to run atomically. The critical insight is the abstract semantics: the choice of active

suspension to run next is arbitrary. If another fibre is pushed before we do our operation it doesn't matter.

The critical observation is this: indeterminacy in semantic specification can enable optimisations. Contrary to popular belief indeterminacy is good.

The write operation is similar but it has an interesting twist!

```
void sync_sched::do_write(io_request_t *req) {
  while(lock.test_and_set(::std::memory_order_acquire); // spin
  fibre_t *r = req->chan->pop_reader();
  if(r) {
    lock.clear(::std::memory_order_release); // release lock
    *r->cc->svc_req->io_request.pdata =
      *current->cc->svc_req->io_request.pdata; // transfer data

    // null out svc requests so they're not re-issued
    r->cc->svc_req = nullptr;
    current->cc->svc_req = nullptr;

    if(r->owner == active_set) {
      active_set->push(current); // current is writer, pushed onto active list
      current = r; // make reader current
    }
    else {
      r->owner.push(r);
      // writer remains current if reader is foreign
    }
  }
  else {
    req->chan->push_writer(current); // i/o fail: push current onto channel
    lock.clear(::std::memory_order_release); // release lock
    current = active_set->pop(); // reset current from active list
  }
}
```

Notice that our old semantics cannot be preserved if the reader and writer come from different active sets. This is also true in the read case but the extra check was not required!

With some modifications, we can easily run our old test case as a regression test. The mainline changes to this:

```
int main() {
  // create the input list
  ::std::list<int> inlst;
  for (auto i = 0; i < 20; ++i) inlst.push_back(i);
```

```
    // output list
    ::std::list<int> outlst;

    // create scheduler
    sync_sched sched;

    // create channels
    channel_t chan1;
    channel_t chan2;

    // create fibres
    fibre_t *prod = new fibre_t ((new producer)->call(nullptr, &inlst, &chan1));
    fibre_t *trans = new fibre_t ((new transducer)->call(nullptr, &chan1, &chan2));
    fibre_t *cons = new fibre_t ((new consumer)->call(nullptr, &outlst, &chan2));

    // push initial fibres onto scheduler active list
    sched.active_set->push_new(prod);
    sched.active_set->push_new(trans);
    sched.active_set->push_new(https://www.microsoft.com/en-au/microsoft-365/visio/flowcha

    // run it
    sched.sync_run();

    // the result is now in the outlist so print it
    ::std::cout<< "List of squares:" << ::std::endl;
    for(auto v : outlst) ::std::cout << v << ::std::endl;
}
```

Note we used `push_new` instead of just `push` to set the owner of each fibre.

### 4.4.1 Process test

[Code a test of concurrent processes here]

### 4.4.2 Foreign threads test

[Code a test of foreign threads here]

## 4.5 Phase 4

### 4.5.1 Memory Management

So far we have glossed over issues of memory management.

Let's consider fibres first. We want every fibre to live in one of three places, exclusive: a fibre is running, in the active set, or hanging on a channel. This is a core invariant we need to enforce. To ensure it, client code must not be permitted to install fibres: they can create a fibre object, but they must not put it into the system. Instead, the `spawn_fibre` service call must be used, and it must be changed to accept an initial continuation rather than a fibre. The scheduler must be changed so the run method also accepts an initial continuation.

This gives the system exclusive control over fibres objects so it can ensure each one exists in precisely one of the three allowed locations. In particular, a fibre must be deleted when its continuation stack is empty.

Now, the client code creates channels and passes pointers to them around. When a channel is deleted, its destructor also deletes all fibres on the channel. This is safe because there cannot be any other references to the fibres, by the invariant we established above.

The hard bit is managing channels. A reference to a channel in a continuation disappears when the continuation is deleted, but there can be many references all over the place. We want to delete a channel when is no longer reachable, so the naive approach is to use a C++ `shared_ptr` instead of an ordinary pointer.

Unfortunately this has three serious problems. The first problem is that we end up with a circularity if a fibre A owning a reference to a channel C does an I/O operation on it. When that happens, the channel is owning the fibre also, so we have a circularity.

Now, if there are no fibres in any active set or currently running which can reach the channel, then the I/O requests of the fibres hanging on the channel can never be satisfied, so they are permanently unreachable as is the cnannel and we should delete the channel along with all those fibres.

To make this work, we need to at least discount all references a fibre makes to all channels when the fibre is put on a channel.

It is very hard to organise this! The current Felix run time does it correctly using a general garbage collector, but our aim here is to build a system capable of real time operation which excludes the use of a naive mark/sweep world stop collector.

Some wag once said that all problems in computer science can be solved by adding another level of indirection. Let us try!

We will add a new concept, a channel endpoint, which is a reference counting smart pointer to a channel. Then we will use a standard C++ `shared_ptr` to a a channel endpoint as a way to distribute access to channels. We will provide a channel constructor that returns several endpoints, and a requirement that at most one endpoint is accessible by any fibre.

Now when the top level smart pointer decrements the counter of a channel endpoint to zero, the endpoint is deleted as usual but the destructor does some-

thing else: it decrements the channel reference count. When all the endpoints are deleted, the channel is also deleted. This is all standard but now for the trick: when a fibre does I/O on a channel it does so through a channel endpoint. If the channel has reference count one, the channel and the sole endpoint we have are deleted along with all fibres on the channel, otherwise the reference count of the channel is decremented. When the fibre is reinstated by a matching I/O request the channel reference count is again incremented.

Let us suppose a fibre is hanging on a channel, and another fibre with access to the channel through a different endpoint terminates, deleting its endpoint. This in turn will delete the channel, and thus the first fibre hanging on it. The smart pointers in the continuations will then delete the first fibres endpoint, however there's a problem: the channel it refers to is already in the process of being deleted. We cannot allow the channel to be deleted twice!

Luckily, we can avoid this problem be ensuring we set the channel reference count to zero before deleting the fibres of the channel. If during this process another endpoint is deleted, it can observe the channel it refers to is already being deleted by checking the reference count, and skip deleting it.

```
// channel4.hpp

// low bit fiddling routines
inline static bool get_lowbit(void *p) {
  return (uintptr_t)p & (uintptr_t)1u;
}
inline static void *clear_lowbit(void *p) {
  return (void*)((uintptr_t)p & ~(uintptr_t)1u);
}
inline static void *set_lowbit(void *p) {
  return (void*)((uintptr_t)p | (uintptr_t)1u);
}

// channel
struct channel_t {
  ::std::atomic_size_t refcnt;
  fibre_t *top;
  ::std::atomic_flag lock;

  channel_t () : top (nullptr), lock(false) {}

  // immobile object
  channel_t(channel_endpoint const&)=delete;
  channel_t& operator= (channel_endpoint const&)=delete;

  // push a fibre as a reader: precondition it must be a reader
  // and, if the channel is non-empty it must contain only readers
```

```
    void push_reader(fibre_t *r) {
      r->next = top;
      top = r;
    }

    // push a fibre as a writer: precondition it must be a writer
    // and, if the channel is non-empty it must contain only writers
    void push_writer(fibre_t *w) {
      w->next = top;
      top = (fibre_t*)set_lowbit(w);
    }

    // pop a reader if there is one, otherwise nullptr
    fibre_t *pop_reader() {
      fibre_t *tmp = top;
      if(get_lowbit(tmp))return nullptr;
      top = top -> next;
      return tmp; // lowbit is clear, its a reader
    }

    // pop a writer if there is one, otherwise nullptr
    fibre_t *pop_writer() {
      fibre_t *tmp = top;
      if(!get_lowbit(tmp)) return nullptr;
      tmp = (fibre_t*)clear_lowbit(tmp); // lowbit is set for writer
      top = tmp -> next;
      return tmp;
    }
  };

  struct channel_endpoint_t {
    channel_t *channel;
    channel_endpoint_t(channel_t *p) : channel(p) { ++channel.refcnt; }

    // immobile object
    channel_endpoint_t(channel_endpoint_t const&)=delete;
    channel_endpoint_t& operator= (channel_endpoint_t const&)=delete;

    // create a duplicate of the current endpoint refering
    // to the same channel. Returns a chan_epref_t; a shared
    // pointer to the new endpoint. Increments the counter
    // of endpoints in the channel.
    // note, C++ must construct a single object containing
    // both the reference count and the channel endpoint.
```

```
    ::std::shared_ptr<channel_endpoint_t> dup() const {
      return make_shared<channel_endpoint_t>(channel);
    }

    ~channel_endpoint_t () {
      switch (channel->refcnt.load()) {
        case 0: break;
        case 1: delete_channel(); break;
        default: --channel->refcnt; break;
      }
    }

    void delete_channel() {
      fibre_t *top = channel->top;
      channel->top = nullptr;
      channel.refcnt = 0;
      while (top) {
        fibre_t *f = (fibre_t*)clear_lowbit(top);
        fibre_t *tmp = f->next;
        delete f;
        top = tmp;
      }
    }

  };

  // channel endpoint reference type
  // note, the refcnt is not atomic.
  // this is fine, because endpoints belong exclusively
  // to a single fibre, which cannot be executed by more
  // the one thread at once.

  using chan_epref_t = ::std::shared_ptr<channel_endpoint_t>;

  chan_epref_t make_channel() {
  }
```