Object-Oriented Choreographic Programming

SAVERIO GIALLORENZO[*], Università di Bologna, Italy and INRIA, France
FABRIZIO MONTESI, University of Southern Denmark, Denmark
MARCO PERESSOTTI, University of Southern Denmark, Denmark

We present Choral, the first language for programming choreographies (multiparty protocols) that builds on top of mainstream programming abstractions: in Choral, choreographies are objects. Given a choreography that defines interactions among some roles (Alice, Bob, etc.), an implementation for each role in the choreography is automatically generated by a compiler. These implementations are libraries in pure Java, which developers can modularly compose in their own programs to participate correctly in choreographies.

CCS Concepts: • **Computing methodologies → Concurrent programming languages**; • **Software and its engineering → Multiparadigm languages**; **Classes and objects**; *Concurrent programming structures.*

Additional Key Words and Phrases: Choreographies, Communication, Higher-kinded Types

1 INTRODUCTION

Background. Choreographies, broadly construed, are coordination plans for concurrent and distributed systems [Object Management Group 2011; W3C 2004]. Examples of choreographies include distributed authentication protocols [OpenID Foundation 2014; Sporny et al. 2011], cryptographic protocols [Diffie and Hellman 1976], and multiparty business processes [Object Management Group 2011; W3C 2004]. In software development, programmers use choreographies to agree on the interactions that communicating endpoints should enact to achieve a common goal; then, each endpoint can be programmed independently. The success of this development process hinges on achieving *choreography compliance*: when all endpoints are run together, they interact as defined by the choreographies agreed upon (Figure 1).

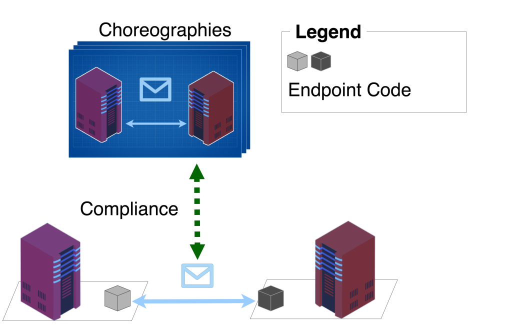

Fig. 1. Choreography compliance: endpoints should communicate as intended by the choreographies that they engage in.

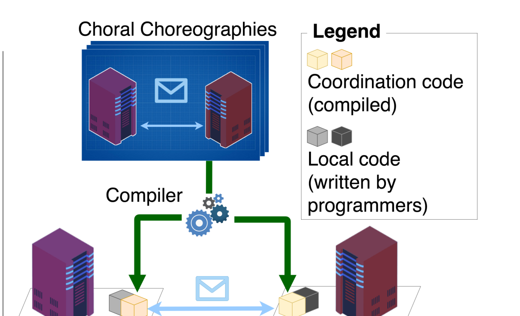

Fig. 2. Our proposal: compliant-by-construction libraries are automatically compiled from choreographies in Choral.

[*]Work mainly performed while the author was employed at the University of Southern Denmark.

Authors' addresses: Saverio Giallorenzo, Department of Computer Science and Engineering, Università di Bologna, Mura Anteo Zamboni 7, Bologna, 40126, Italy , INRIA, Sophia-Antipolis, France, saverio.giallorenzo@gmail.com; Fabrizio Montesi, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, Odense, 5230, Denmark, fmontesi@imada.sdu.dk; Marco Peressotti, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, Odense, 5230, Denmark, peressotti@imada.sdu.dk.

Achieving choreography compliance is hard, because of some usual suspects of concurrent and distributed programming: predicting how multiple programs will interact at runtime is challenging [O'Hearn 2018], and mainstream programming languages do not adequately support programmers in reasoning about coordination in their code [Leesatapornwongsa et al. 2016; Lu et al. 2008]. Additionally, choreographies are complex. At a minimum, choreographies define the expected communication flows among their roles (abstractions of endpoints, like "Alice", "Bob", "Buyer", etc.) [Intl. Telecommunication Union 1996]. However, often choreographies include also computational details of arbitrary complexity, for example: pre- or post-processing of data (encryption, validation, anonymisation, etc.), state information, and decision procedures to choose among alternative behaviours. These computational details are essential parts of many protocols, ranging from security protocols to business processes.

In response to the challenge of choreography compliance, researchers investigated methods to relate choreographies to endpoint programs—many are reviewed in [Ancona et al. 2016; Hüttel et al. 2016]. Initially, these methods focused on simple choreography languages without computation that were inspired by, e.g., communicating automata, process calculi, and session types [Basu et al. 2012; Bravetti and Zavattaro 2007; Honda et al. 2016; Qiu et al. 2007]. Some of the ideas developed for choreographies were later adopted in the paradigm of *choreographic programming* [Carbone and Montesi 2013; Montesi 2013], where choreographies are written in a Turing-complete programming language that allows for defining arbitrary computation at endpoints. Thanks to the capability of combining computation with coordination, choreographic programming languages can capture realistic protocols that include data manipulation and decision procedures, including encryption strategies (as in security protocols), retry strategies (as in transport protocols), and marshalling (as in application protocols). In choreographic programming, compliance is obtained by construction: given a choreography, a compiler automatically translates it to a set of compliant endpoint implementations. Choreographic programming has been shown to have promising potential in multiple contexts, including information flow [Lluch-Lafuente et al. 2015], distributed algorithms [Cruz-Filipe and Montesi 2016], cyber-physical systems [López and Heussen 2017; López et al. 2016], and self-adaptive systems [Dalla Preda et al. 2017].

The problem. Current approaches to choreographic programming come at a significant cost to modularity, and it remains unclear how the benefits of this paradigm can be applied to mainstream software development.

Typically, the endpoint code that implements a choreography comes in libraries that developers can use in their applications through an API [Atzori et al. 2010; Murty 2008; Wilder 2012]. For example, a library that implements a choreography for user authentication might provide a method `authenticate` that a web service can invoke to run (its part of) the protocol in collaboration with a connected client. The implementation of `authenticate` might involve a series of communications between the client and the web service. Potentially, a third-party identity provider might be involved as well, and messages might include passwords that should be hashed. Usually, and ideally, all these details are hidden from the API (the signature of the method) exposed to the developer of the web service. This allows the developer to minimise coupling between their implementation of the functionalities offered by the web service and the choreography used for authentication, bringing the expected benefit: the library implementing the choreography and the rest of the code of the web service can be updated independently and recombined, as long as the API provided by the former remains compatible with that expected by the latter. This benefit is key to large-scale software development, and it is an initial assumption for modern development practices like microservices and DevOps [Dragoni et al. 2017].

Unfortunately, previous frameworks for choreographic programming do not support modular software development [Dalla Preda et al. 2017; Montesi 2013]. The code that these frameworks generate for each endpoint is a "black box" program without an API: it can be executed, but it is not designed to be composed by programmers within larger codebases. Thus, these frameworks fall short of providing the aforementioned benefit of modularity. Furthermore, the common scenario of programming an endpoint that participates in multiple choreographies is not supported, and neither is programming an endpoint where participating in a choreography is only part of what the endpoint does.

A major factor that makes modularity challenging is that current choreography languages are based on behavioural models (process calculi, communicating automata, etc.), which makes translating choreographies into libraries based on mainstream abstractions (data, functions, objects, etc.) nontrivial. For the simpler setting of choreography languages without computation, we know that choreographies can be translated to libraries that offer a "fluid" API. For instance, Scalas et al. [2017] produce object-oriented libraries whose APIs enforce the invocation of `send` and `receive` methods in the right order: if an endpoint should send, receive, and then send, the developer will be forced to write something like `o.send(..).receive(..).send(..)`. However, this approach leaks the structure of the communication flow implemented by the library; thus, future versions of the library that adopt a different structure (e.g., an unnecessary communication might be removed, or some communications might be bunched together for efficiency) would require rewriting the application that uses the library. Furthermore, this approach does not let the choreography programmer decide on how the generated API will look like; thus, we cannot use this method to generate drop-in replacements for existing libraries.

In summary, we still have to discover how the principles of choreographic programming can be applied to mainstream programming. The aim of this article is to fill this gap.

This article. We present Choral, a new choreographic programming framework that supports modularity and is based on mainstream programming concepts (from object-oriented programming). To demonstrate applicability, Choral is compatible with Java, but our ideas apply to most statically-typed object-oriented languages.

The fulcrum of Choral is a new interpretation of choreographies that builds naturally on top of existing language abstractions: in Choral, *objects are choreographies*. The starting point for this interpretation is a generalisation of the key idea found in Lambda 5 [Murphy VII et al. 2004], the model that inspired the research line on multitier programming [Murphy VII et al. 2007; Neubauer and Thiemann 2005; Serrano et al. 2006; Weisenburger et al. 2020]. In Lambda 5, each data type is located at a (single) place, which enables reasoning on spatially-distributed computation. Choral generalises these types from single to *multiple* locations, which allows us to express that an object is implemented choreographically: Choral objects have types of the form `T@(A1, ..., An)`, where `T` is the usual interface of the object, and `A1`, …, `An` are the roles that collaboratively implement the object.

As an example, consider the case of a multiparty choreography for distributed authentication, where a service authenticates a client via a third-party identity provider. We can define such a choreography as a Choral object of type `DistAuth@(Client, Service, IP)` (`IP` is short for identity provider). The object can implement methods that involve multiple roles. For example, it can offer a method `authenticate` with the following signature.

```
Optional@Service<AuthToken> authenticate(Credentials@Client credentials)
```
Choral Code

Invoking method `authenticate` with some `Credentials` located at `Client` returns an authorisation token at `Service` (`Optional`, since authentication might fail), denoting movement of data.

We leverage our choreographies-as-objects interpretation to develop a methodology for choreography compliance that supports modularity and is compatible with mainstream programming. We depict this methodology in Figure 2. Given the code of a Choral object with some roles, a compiler produces a compliant-by-construction software library in pure Java for each role ("coordination code" in the figure): each library contains the local implementation of what its role should do to execute the choreography correctly. These libraries offer Java APIs derived from the source choreographies, which reveal only the details pertaining the implemented role. When a software developer programs an endpoint that should engage in a choreography, they can just take the library compiled for the role that they want to play and use it through its Java API. Through such APIs, developers can modularly compose multiple libraries with their own code ("local code" in the figure), thus gaining the ability to participate in multiple choreographies.

The Java code compiled from method `authenticate` for role `Service` has the signature below.

```
Optional<AuthToken> authenticate()
```
Generated Code

Parameter `credentials` from the choreographic method has disappeared, since its type does not include `Service`; conversely, the return type `Optional<AuthToken>` remains.

Contributions. We outline our main contributions.

Language. We present Choral, the first choreographic programming language based on mainstream abstractions and interoperable with a mainstream language (Java). The key novelty of the Choral language is that data types are higher-kinded types parameterised on roles. We leverage this feature to bring the key aspects of choreographies to object-oriented programming (spatial distribution, interaction, and knowledge of choice). Choral is also the first truly higher-order choreographic language, where choreographies passed as arguments can carry state and invocations of higher-order choreographies require no centralised coordination [Dalla Preda et al. 2017; Demangeon and Honda 2012].

Integrating object-oriented principles with choreographies brings key benefits also in the other direction, in the sense that we gain a much richer choreographic language than the state of the art. By using subtyping, we can define abstract APIs for choreographies that can be implemented in different ways and communication behaviours, bringing the usual substitution principle for objects to choreographies. Method overloading allows us to specialise computation based on the locations of arguments—which is a new dimension for overloading. Semantic parametricity enables the writing of reusable choreographies that treat uniformly parameters that implement a shared API. These features allow us to generalise choreographies from assuming a fixed communication primitive to user-definable communication methods. This frees Choral from commitments to any communication technology or middleware. Furthermore, we can define in Choral the first hierarchy of "channel types" for choreographies, which can be used to represent at the level of types the topological assumptions of a choreography. In general, users are free to define other "channel types" to support more topologies.

We implement a type checker that, in addition to the expected checks for an object-oriented language, detects also coding errors related to roles. For example, our checker can rule out computation at a role that erroneously accesses data at another role without proper communications. This makes distribution errors manifest to the programmer. Our typing also supports the reuse of all existing Java classes and interfaces in Choral, because every such type can be lifted to a Choral type located at a single role.

In Choral, choreographies are concrete software artifacts. This poses questions such as what code should go in these artifacts and what code should remain local, and how these two parts of software should interact through APIs. We elicit these questions and report on our experience in addressing them throughout the article, where they become relevant.

Compiler. We implement a compiler that translates Choral source into Java libraries that comply with the source choreography: the code generated for each role performs the actions prescribed by the choreography. With our compiler, the programmer of the choreography is in control of the generated APIs: the APIs for each role follow the same structure found in the choreography, but where all data types and parameters in method signatures that do not pertain the role are omitted, as we exemplified previously for the `authenticate` method.

Testing. We present the first testing tool for choreographic programming. Since choreography implementations are spread over multiple components (one for each role), testing choreographies can be difficult because it calls for integration testing. Our testing tool leverages Choral to write "choreographic tests" that look like simple unit tests at the choreographic level, but are then compiled to integration tests that integrate the respective implementations of all roles.

Evaluation. We explore the expressivity of Choral with use cases of different kinds, covering security protocols, cyberphysical systems connected to the cloud, and parallel algorithms. For these use cases, we discuss relevant code excerpts and development techniques induced by Choral.

We then move from our own examples to a real-world comparison.

First, we show how Choral can be used to transition existing Java programs to choreographies. We reimplement in Choral the Java implementation of Karatsuba's algorithm for fast multiplication, which yields a parallel implementation. Furthermore, we reimplement a clone of Twitter developed by the Spring team. Implementing this system in Choral requires identifying roles, which makes the original monolithic application much more modular.

Second, we compare Choral to a popular framework for concurrent programming in Java based on actors (Akka). In particular, we identify the key differences in the development processes and resulting architectures induced by the two technologies. We find that Choral provides concrete benefits in avoiding subtle concurrency bugs.

Third, we carry out a quantitative evaluation of how Choral impacts software development and runtime performance. Thanks to its choreographic approach, Choral consistently leads to smaller codebases. Moreover, the impact of our compiler on the speed of the edit-compile cycle is negligible (milliseconds). Finally, we show that the runtime performance of the code generated by Choral is not worse (and often better) than that of alternative implementations in Java and Akka.

Outline. We overview Choral with simple examples in Section 2, and give more realistic use cases in Section 3. The syntax and implementation of Choral are discussed in Section 4, and testing in Section 5. We evaluate Choral in depth in Section 6. Related and future work are discussed in Section 7. We draw conclusions in Section 8.

2 CHORAL IN PRACTICE

We start with an overview of the key features of Choral. First, spatial distribution: the expression of computation that takes place at different roles (Section 2.1). Second, interaction: the coding of data exchanges between roles (Section 2.2). Third, knowledge of choice: the coordination of roles to choose between alternative behaviours (Section 2.3).

The Choral language is quite big. Its usefulness depends on the capability to produce software libraries whose APIs look like "idiomatic" Java APIs, so we chose to incorporate a substantial set of features, which would commonly be considered necessary to use and produce Java APIs: Choral supports classes, interfaces, generics, inheritance, and method overloading. APIs generated by Choral support lambda expressions, in the sense that Java programmers can pass lambda expressions as arguments to our APIs. (Just as in Java, Choral sees these arguments as objects.) Supporting the Java syntax for lambda expressions inside of Choral programs is not necessary for our objective, since they can be encoded as objects, so we leave it to future work on ergonomics.

In the rest of this section, we explain the key aspects of Choral by assuming that the reader is familiar with Java. The reader can assume that language constructs that have the same syntax as Java behave as expected (modulo our additions, which we explain in the text).

2.1 Roles and data types

Hello roles. All values in Choral are distributed over one or more roles, using the `@`-notation seen in Section 1. The degenerate case of values involving one role allows Choral to reuse existing Java classes and interfaces, lifted mechanically to Choral types and made available to Choral code. For example, the literal `"Hello from A"@A` is a string value `"Hello from A"` *located at role* `A`. Code involving different roles can be freely mixed in Choral, as in the following snippet.

```
1 class HelloRoles@(A,B) {
2  public static void sayHello() {
3   String@A a = "Hello from A"@A;
4   String@B b = "Hello from B"@B;
5   System@A.out.println(a);
6   System@B.out.println(b);
7 }}
```
Choral Code

The code above defines a class, `HelloRoles`, parameterised over two roles, `A` and `B`. Line 3 assigns the string `"Hello from A"` located at `A` (`"Hello from A"@A`) to variable `a` of type "String at `A`" (`String@A`), and then the same but for a string located at `B`. Then, Line 4 prints variable `a` by using the `System` object at `A` (`System@A`), and then prints variable `b` at role `B`.

Roles are part of data types in Choral, adding a new dimension to typing. For example, the statement `String@A a = "Hello from B"@B` would be ill-typed, because the expression on the right returns data at a different role from that expected by the left-hand side.

Formally, in Choral, `String` is a type of a higher kind (or type constructor): it takes a role in order to return a type of the same kind of Java types [Moors et al. 2008]. The code `String@A` represents the instantiation of `String` at role `A`. Any Java type is automatically liftable to a Choral type with a single role parameter by following the same reasoning, enabling interoperability. Type constructors in Choral are not limited to a single role in general. We are going to see examples with multiple roles and more complex type parameters later in this section.

From Choral to Java. Given class `HelloRoles`, the Choral compiler generates for each role parameter a Java class with the behaviour for that role, in compliance with the source class. Here, the Java class for role `A` is `HelloRoles_A` and the class for `B` is `HelloRoles_B`.

```
class HelloRoles_A {
 public static void sayHello() {
  String a = "Hello from A";
  System.out.println( a );
}}
```
Generated Code

```
class HelloRoles_B {
 public static void sayHello() {
  String b = "Hello from B";
  System.out.println( b );
}}
```
Generated Code

Each generated class contains only the instructions that pertain that role. If Java developers want to implement the behaviour of method `sayHello` for a specific role of the `HelloRoles` choreography, say `A`, they just need to invoke the generated `sayHello` method in the respective generated class (`HelloRoles_A`). If all Java programs interested in participating to `HelloRoles` do that, then their resulting global behaviour complies by construction with the source choreography.

Notice that the code compiled for `A` and `B` will not interact and can therefore proceed fully concurrently, because the choreography does not prescribe so. In general, choreographic programming languages are expected to generate code that interacts only to enact the communications that the programmer specified in the choreography [Montesi 2022]. Choral follows this adequacy principle. We discuss how to program interactions in Section 2.2.

Distributed State. Fields of Choral classes carry state and can be distributed over different roles. For example, a class `BiPair` can define a "distributed pair" storing two values at different roles.

```
1 class BiPair@(A,B)<L@C,R@D> {
2  private L@A left;
3  private R@B right;
4  public BiPair(L@A left, R@B right) { this.left = left; this.right = right; }
5  public L@A left() { return this.left; }
6  public R@B right() { return this.right; }
7 }
```
Choral Code

Class `BiPair` is distributed between roles `A` and `B` and has two fields, `left` and `right`. In general, every class or interface in Choral is always parameterised on at least one role, and is hence a type constructor. The class is also parameterised on two data types, `L` and `R`, which exemplifies our support for generics [Naftalin and Wadler 2007]. At line 1, `L@C` specifies that `L` is expected to be a data type parameterised over a single role, abstracted by `C`; similarly for `R@D`. Naming the role parameters of `L` and `R` does not add any information in this particular example (we only need to know that they have one parameter). However, naming role parameters in generics is useful for expressing type bounds in extends clauses as discussed later in this section. Choral interprets role parameter binders as in Java generics: the first appearance of a parameter is a binder, while subsequent appearances of the same parameter are bound. Observe that the scope of role parameters `C` and `D` is limited to the declaration of the type parameters `L` and `R`, respectively—we use distinct names exclusively for readability. At lines 2 and 3 we have the two fields, `left` and `right`, respectively located at `A` and `B` as stated by the types `L@A` and `R@B`, the constructor is at line 4, while accessors to the two fields are at lines 5–6.

Data structures like `BiPair` are useful when defining choreographies where the data at some role needs to correlate with data at another role, as with distributed authentication tokens. We apply them to a use case in Section 3.1.

2.2 Interaction

Choral programs become interesting when they contain interaction between roles—otherwise, they are simple interleavings of local independent behaviours by different roles, as in `HelloRoles`.

Choreography models typically come with some fixed primitives for interaction, e.g., sending a value from a role to another over a channel [Carbone et al. 2012]. Thanks to our data types parameterised over roles, Choral is more expressive: we can *program* these basic building blocks and then construct more complex interactions compositionally. This allows us to be specific about the requirements of choreographies regarding communications, leading to more reusable code. For instance, if a choreography needs only a directed channel, then our type system can see by subtyping that a bidirectional channel is also fine.

Directed data channels. We start our exploration of interaction in Choral from simple directed channels for transporting data. In Choral, such a channel is just an object (if you prefer, call it a choreography) that takes data from one place to another. We can specify this as an interface.

```
interface DiDataChannel@(A,B)<T@C> { <S@D extends T@D> S@B com(S@A m); }
```
Choral Code

A `DiDataChannel` is the interface of a directed channel between two roles, abstracted by `A` and `B`, for transmitting data of a given type, abstracted by the type parameter `T`, from `A` to `B` (hence the number of role parameters in `T`). Data transmission is performed by invoking the only method of the interface: `com` which takes any value of a subtype of `T` located at `A`, `S@A`, and returns a value of type `S@B`. The type parameter `S` of method `com` has `T` as upper bound (we can read the expression `S@D extends T@D` as "for any role `D`, `S@D extends T@D`") and allows us to transmit data of types more specific than `T` without losing type information (as it would be if the signature of `com` was simply `T@B com(T@A m)`).

Parameterising data channels over the type of transferrable data (`T`) is important in practice for channel implementors, because they often need to deal with data marshalling. Choral comes with a standard library that offers implementations of our channel APIs for a few common types of channels, e.g., TCP/IP sockets supporting JSON objects and shared memory channels. Users can provide their own implementations.

Using a `DiDataChannel`, we can write a simple method that sends a string notification from a `Client` to a `Server` and logs the reception by printing on screen.

```
notify(DiDataChannel@(Client,Server)<String> ch, String@Client msg) {
 String@Server m = ch.<String>com(msg);
 System@Server.out.println(m);
}
```
Choral Code

Note that `String` is a valid instantiation of `T` of `DiDataChannel` because we lift all Java types as Choral types parameterised over a single role.

Alien data types. Compiling `DiDataChannel` to Java poses an important question: what should be the return type of method `com` in the code produced for role `A`? Since the return type does not mention `A` (we say that it is *alien* to `A`), a naïve answer to this question could be **`void`**, as follows.

```
interface DiDataChannel_A<T> { <S extends T> void com(S m); }
```
Tentative Code

It turns out that this solution does not work well with expressions that compose multiple method calls, including chaining like `m1(e1,e2).m2(e3)` and nesting like `m1(m2(e))`. As a concrete example, consider a simple round-trip communication from `A` to `B` and back.

```
static <T@C> T@A roundTrip(DiDataChannel@(A,B)<T> chAB, DiDataChannel@(B,A)<T> chBA, T@A msg)
{ return chBA.<T>com(chAB.<T>com(msg)); }
```
Choral Code

Method `roundTrip` takes two channels, `chAB` and `chBA`, which are directed channels respectively from `A` to `B` and from `B` to `A`. The method sends the input `msg` from `A` to `B` and back by nested `com`s and returns the result at `A`.

A structure-preserving compilation of method `roundTrip` for role `A` would be as follows.

```
static <T> T roundTrip(DiDataChannel_A<T> chAB, DiDataChannel_B<T> chBA, T msg)
{ return chBA.<T>com(chAB.<T>com(msg)); }
```
Generated Code

Observe how the inner method call, `chAB.com<T>(msg)`, should return something, such that it can trigger the execution of the outer method call to receive the response. Therefore, the `com` method of `DiDataChannel_A` cannot have **`void`** as return type.

Programming language experts have probably guessed by now that the solution is to use unit values instead of **`void`**. Indeed, Choral defines a singleton type `Unit`, a final class that our compiler uses instead of **`void`** to obtain Java code whose structure resembles its Choral source code.

We now show the Java code produced by our compiler from `DiDataChannel` for both `A` and `B`.

```
interface DiDataChannel_A<T>{
 <S extends T> Unit com(S m);
}
```
Generated Code

```
interface DiDataChannel_B<T>{
 <S extends T> S com(Unit m);
}
```
Generated Code

Given these interfaces, the compilation of `roundTrip` for role `A` is well-typed and correct Java code. An alternative to using `Unit` would have been to give up on preserving structure in the compiled code; we chose in favour of our solution because preserving structure makes it easier to read and debug the compiled code (especially when comparing it to the source choreography), and also makes our compiler simpler.

The users of our compiled libraries are not forced to passing `Unit` arguments to methods, as for method `com` of `DiDataChannel_B`: for methods like these, our compiler provides corresponding "courtesy methods" that take no parameters and inject `Unit`s automatically.

Bidirectional channels. An immediate generalisation of directed data channels brings us to bidirectional data channels, specified by `BiDataChannel`.

```
interface BiDataChannel@(A,B)<T@C,R@D> extends DiDataChannel@(A,B)<T>,DiDataChannel@(B,A)<R>{
  <S@C extends T@C> S@B com(S@A m); // inherited from DiDataChannel@(A,B)<T>
  <S@C extends R@C> S@A com(S@B m); // inherited from DiDataChannel@(B,A)<R>
}
```
Choral Code

A `BiDataChannel` is parameterised over two types: `T` is the type of data that can be transferred from `A` to `B` and, vice versa, `R` is the type of data that can be transferred in the opposite direction. This is obtained by multiple type inheritance: `BiDataChannel` extends `DiDataChannel` in one and the other direction, which allows for using modularly a bidirectional data channel in code that has the weaker requirement of a directed data channel in one of the two supported directions. Distinguishing the two parameters `T` and `R` is useful for protocols that have different types for requests and responses, like HTTP. Extending `DiDataChallen` twice does not result in any clashes since `A` and `B` play different roles in each supertype. This "twin" inheritance results in the overload of method `com` one for each communication direction supported by the channel. This overload does not result in any clash in the generated code as illustrated by the code generated for `A` (code generated for `B` is symmetric).

```
interface BiDataChannel_A<T,R> extends DiDataChannel_A<T>, DiDataChannel_B<R> {
  <S extends T> Unit com(S m); // inherited from DiDataChannel_A<T>
  <S extends R> S com(Unit m); // inherited from DiDataChannel_B<R>
}
```
Generated Code

We discuss more types of channels (including symmetric channels) in Section 2.4 and provide more details on inheritance and overloading in Section 4.

Forward chaining. We use bidirectional channels to define a choreography for remote procedure calls, called `RemoteFunction`, which leverages the standard Java interface `Function<T,R>`.

```
1 class RemoteFunction@(Client,Server)<T@A,R@B> {
2  private BiDataChannel@(Client,Server)<T,R> ch;
3  private Function@Server<T,R> f;
4  public RemoteFunction(BiDataChannel@(Client,Server)<T,R> ch, Function@Server<T,R> f){
5   this.ch = ch; this.f = f;
6  }
7  public R@Client call(T@Client t) {
8   return ch.<R>com(f.apply(ch.<T>com(t)));
9 }}
```
Choral Code

In the experience that we gained by programming larger Choral programs (as those in Section 3), compositions of method invocations including data transfers as in line 4 of `RemoteFunction` are rather typical. In these chains, we read data transfers from right to left (innermost to outermost invocation), but most choreography models in the literature use a left-to-right notation (as in "Alice sends 5 to Bob"). To make Choral closer to that familiar choreographic notation, we borrow the forward chaining operator `>>` from F# [Petricek and Skeet 2009], so that `exp >> obj::method` is syntactic sugar for `obj.method(exp)`. For example, we can rewrite method `call` of `RemoteFunction` as follows, which is arguably more readable and recovers a more familiar choreographic notation.

```
public R@Client call(T@Client t){return t >> ch::<T>com >> f::apply >> ch::<R>com;}
```
Choral Code

Using Choral libraries. As mentioned for Channels, when we compile the `RemoteFunction` class above, we obtain two Java classes: a `RemoteFunction_Client`, which sends some data to the `Server` for processing and waits for its response, and a `RemoteFunction_Server`, which, upon reception, feeds the data into a `Function` and sends back to the `Client` its result.

The `RemoteFunction_Server` is an interesting example of how users interact with Choral libraries. The code generated from Choral is (snippet):

```
class RemoteFunction_Server<T,R> {
 private BiDataChannel_B <T,R> ch;
 private Function <T,R> f;
 public RemoteFunction_Server(BiDataChannel_B<T,R> ch, Function<T,R> f) { /*...*/ }
 public Unit call() { /*...*/ }
}
```
Generated Code

A user of the `RemoteFunction_Server` can interact in the choreography by providing the definition of the `Function` at the creation of the object. In general, this is how we expect users to integrate Choral-generated code with their "local code", i.e., code parametric to the choreography that users can implement locally, without any coordination with the other participants (save the APIs induced by Choral-generated code). For example, the snippet below is from a Java class that uses `RemoteFunction_Server` to provide a remote procedure for checking if an integer is even.

```
1 BiDataChannel_B<Integer,Boolean> channel = /*...*/;
2 new RemoteFunction_Server<Integer, Boolean>(channel, i -> i %
```
Local Code

Here, at line 2 (second argument of the constructor), we provide the definition of the `Function` using Java Lambdas functional syntax.

2.3 Knowledge of choice

Knowledge of choice is a hallmark challenge of choreographies: when a choreography chooses between two alternative behaviours, roles should coordinate to ensure that they agree on which behaviour should be implemented [Castagna et al. 2011].

We exemplify the challenge with the following code, which implements the consumption of a stream of items from a producer `A` to a consumer `B`.

```
// wrong implementation
consumeItems(DiDataChannel@(A,B)<Item@C> ch, Iterator@A<Item> it, Consumer@B<Item> consumer){
 if (it.hasNext()){
  it.next() >> ch::<Item>com >> consumer::accept;
  consumeItems(ch, it, consumer);
}}
```
Choral Code

Method `consumeItems` takes a channel from `A` to `B`, an iterator over a collection of items at `A`, and a consumer function for items at `B`. Role `B` works reactively, where its consumer function is invoked whenever the stream of `A` produces an element: if the iterator can provide an item (line 3), it is transmitted from `A` to `B`, consumed at `B`, and the method recurs to consume the other items (line 4).

The reader familiar with choreographies should recognise that this method implementation is *wrong*, due to (missing) knowledge of choice: the information on whether the if-branch should be entered or not is known only by `A` (since it evaluates the condition), so `B` does not know whether it should run lines 4–5 (receive, consume, and recur), or do nothing and terminate.

In choreographic programming, knowledge of choice is typically addressed by equipping the choreography language with a "selection" primitive to communicate constants drawn from a dedicated set of "labels" [Carbone and Montesi 2013; López et al. 2016]. This makes it possible for the compiler to build code that can react to choices made by other roles, inspired by a theoretical operator known as merging [Carbone et al. 2012]. In Choral, we adapt this practice to objects. Notably, Choral is expressive enough that we do not need to add a dedicated primitive, nor a dedicated set of labels.

We define a method-level annotation `@SelectionMethod`, which developers can apply only to methods that can transmit instances of enumerated types between roles (the compiler checks for this condition). For example, we can specify a directed channel for sending such enumerated values with the following `DiSelectChannel` interface.

```
interface DiSelectChannel@(A,B) {
 @SelectionMethod
 <T@C extends Enum@C<T>> T@B select(T@A m);
}
```
Choral Code

Our compiler assumes that implementations of methods annotated with `@SelectionMethod` return at the receiver the same value given at the sender. (This is part of the contract for channels, and it is a standard assumption in implementations of choreographies.)

Typically, channels used in choreographies are assumed to support both data communications and selections. We can specify this with `DiChannel` (directed channel), a subtype of both `DiDataChannel` and `DiSelectChannel` (we include inherited methods for convenience).

```
interface DiChannel@(A,B)<T@C> extends DiDataChannel@(A,B)<T>, DiSelectChannel@(A,B) {
  <S@C extends T@C> S@B com(S@A m);          // inherited from DiDataChannel@(A,B)
  <S@C extends Enum@C<T>> T@B select(T@A m); // inherited from DiSelectChannel@(A,B)
}
```
Choral Code

Using `DiChannels`, we can update `consumeItems` to respect knowledge of choice.

```
enum Choice@A { GO, STOP }
```
Choral Code

```
consumeItems(DiChannel@(A,B)<Item@C> ch, Iterator@A<Item> it, Consumer@B<Item> consumer) {
 if (it.hasNext()) {
  ch.<Choice>select(Choice@A.GO);
  it.next() >> ch::<Item>com >> consumer::accept;
  consumeItems(ch, it, consumer);
 } else {
  ch.<Choice>select(Choice@A.STOP);
}}
```
Choral Code

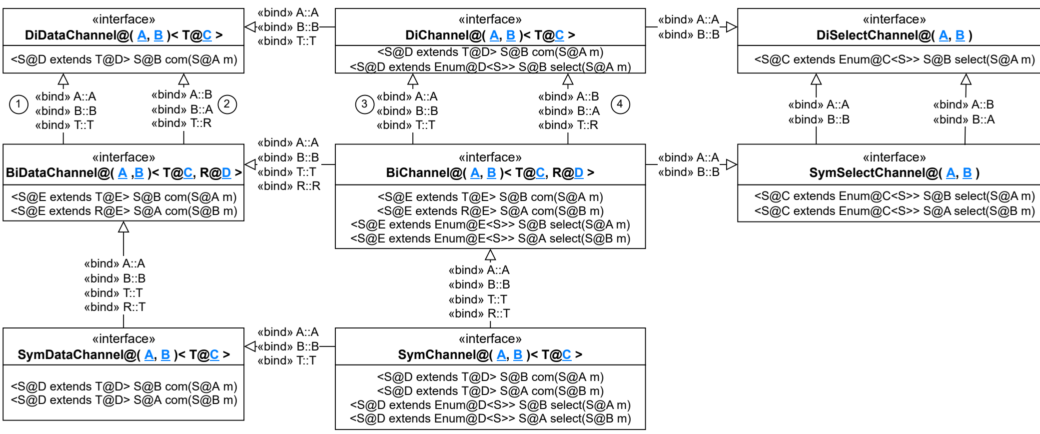

Fig. 3. UML class diagram of the hierarchy of the `*Channel` interfaces.

Differently from the previous broken implementation of `consumeItems`, now role `A` sends a selection of either `GO` or `STOP` to `B`. Role `B` can now inspect the received enumerated value to infer whether it should execute the code for the if- or the else-branch of the conditional. This information is exploited by our static analyser to check that `consumeItems` respects knowledge of choice, and also by our compiler to generate code for `B` that reacts correctly to the choice performed by `A`. (A more extensive example containing also the code compiled for the receiver is given in Section 3.1.)

Our compiler supports three features to make knowledge of choice flexible. Firstly, our knowledge of choice check works with arbitrarily-nested conditionals. Secondly, knowledge of choice can be propagated transitively. Say that a role `A` makes a choice that determines that two other roles `B` and `C` should behave differently, and `A` informs `B` of the choice through a selection. Now either `A` or `B` can inform `C` with a selection, because our compiler sees that `B` now possesses knowledge of choice. Thirdly, knowledge of choice is required only when necessary: if `A` makes a choice and another role, say `B`, does not need to know because it performs the same actions (e.g., receiving an integer from `A`) in both branches, then no selection is necessary. We explain the technicalities behind this in Section 4.

2.4 The family of Choral channels

Choral types give us a new way to specify requirements on channels that prior work implicity assumed, leading to the definition of a family of channel interfaces diagrammed in Figure 3.

From the left-most column in Figure 3, at the top, we find `DiDataChannel`, representing a directed channel parameterised over `T` (the type of the data that can be sent). We obtain `BiDataChannel`, a bidirectional data channel, by extending `DiDataChannel` once for each direction: ① it binds the role parameters of one extension in the same order given for the role parameters of `BiDataChannel`, giving us a direction from `A` to `B` and ② it binds the role parameters of the other extension in the opposite way, giving us a direction from `B` to `A`. The result is that `BiDataChannel` defines two `com` methods: one transmitting from `A` to `B`, the other from `B` to `A`. The last lines in ① and ② in Figure 3 complete the picture: the first generic data type `T` binds data from `A` to `B`, second generic data type `R` binds data from `B` to `A`. The `SymDataChannel` in Figure 3, by extending the `BiDataChannel` interface and binding the two generic data types `T` and `R` with its only generic data type `T`, defines a bidirectional data channel that transmits one type of data, regardless its direction.

The right-most vertical hierarchy in Figure 3 represents channels supporting selections and it follows a structure similar to that of data channels. A `DiSelectChannel` is a directed selection channel and a `SymSelectChannel` is the bidirectional version—there is no `BiSelectChannel` since both directions exchange the same enumerated types.

The vertical hierarchy in the middle column of Figure 3 is the combination of the left-most and right-most columns. Interface `DiChannel` is a directed channel that supports both generic data communications and selections. `BiChannel` is its bidirectional extension (③ and ④ in Figure 3), and `SymChannel` is the symmetric extension of `BiChannel`. The snippet below contains the definition of the interface `BiChannel`.

```
interface BiChannel@(A,B)<T@C, RD> extends
    DiChannel@(A,B)<T>,         // A BiChannel is a pair of directed channels
    DiChannel@(B,A)<R>,         // in opposite directions
    BiDataChannel@(A,B)<T,R>,   // that supports data
    BiSelectChannel@(A,B)       // and selections
  { } // we do not define any new methods, since they are all inherited
```

This definition means that for any pair of distinct roles `C`, `D` and for any pair of types `S`, `P` (with one role parameter), `BiChannel@(C,D)<S,P>` is a subtype of `DiChannel@(C,D)<S>`, `DiChannel@(D,C)<P>`, `BiDataChannel@(C,D)<S,P>`, and `BiSelectChannel@(C,D)`.

2.5 Handling exceptions

Typically, choreographic languages assume that reliable communications [Carbone et al. 2008; Carbone and Montesi 2013; Dalla Preda et al. 2017]. The only exception is the language theory in [Montesi and Peressotti 2017], which shows that one can relax this assumption, by allowing the choreographic language to handle local exceptions. In Choral, we follow the same strategy, which we briefly illustrate here.

Choral can invoke Java code, which might raise an exception. Plain Java code is always located at one role, and therefore the same holds for exceptions (exceptions are "local"). We exemplify how we treat exceptions with the following choreography, where a role `B` uses the Java standard library to save on disk some text communicated from another role `A`.

```
1  public Result@A<String, String> save(
2   SymDataChannel@(A,B)<String> chAB, String@A text, Path@B path
3  ) {
4   String@B textB = text >> ch::<String>com;
5   Result@B<String, String> result;
6   try {
7    Files@B.writeString(path, text);
8    result = Result@B.ok("Saved"@B);
9   } catch(IOException@B ex) {
10   result = Result@B.err(ex.getMessage());
11  }
12  return result >> ch::<String>com;
13 }
```
Choral Code

Above, we start by communicating the text to be saved from `A` to `B` (line 4). We then declare at `B` a result variable (line 5), which will store either a success or error message which `B` later communicates to `A` (line 12). In line 7, `B` attempts at saving the received text.

This choreography might incur execution errors related to communication or file writing. Exceptions encapsulate these errors. For example, the invocation of method `writeString` might throw an `IOException` located at `B`, which we handle with the **`try-catch`** block at lines 6–11.

Also method `com` can throw exceptions, depending on the implementation of channel `ch`. Channels for remote communication (e.g., based on TCP/IP sockets) in the Choral library use the following strategy: the sender attempts at sending a message until its network stack accepts the task (using exponential backoff and bound to the number of attempts, to guarantee termination); likewise, the receiver attempts at receiving until a timeout expires. If the sender ultimately fails at relaying the message to its local network stack, the channel throws a `SendException` at the sender (`A` in our example). If the receiver timeouts before receiving a message, the channel raises a `TimeoutException` at the receiver (`B` in our example).

As a design choice, we left the exceptions of method `com` unchecked. The idea is that the implementation of channels should do their best to deliver messages, and when this is not possible the local code that uses the code generated from a choreography should deem the execution of the choreography unsuccessful. The local code is free to catch these exceptions and attempt recovery, for example by executing the choreography again (as in actor frameworks [Wyatt 2013]).

However, our implementation of `com` in the Choral standard library is just a default, Choral does not hardcode any communication semantics. The user is free to implement alternative communication methods that expose an API which the caller choreography might use to handle network errors, e.g., lost messages. For instance, we might account for lossy communications between `A` and `B` within the above choreography as follows.

```
public Result@A<String, String> save(
 LossySymDataChannel@(A,B)<String> chAB, String@A text, Path@B path
) {
 Optional@B<String> textB = text >> ch::<String>lossyCom;
 Result@B<String, String> result;
 if(!textB.isEmpty()) {
  try {
   Files@B.writeString(path, textB.get());
   result = Result@B.ok("Saved"@B);
  } catch(IOException@B ex) {
   result = Result@B.err(ex.getMessage());
  }
 } else {
  result = Result@B.err("Network error");
 }
 return result >> ch::<String>com;
}
```

Choral Code

Channel `chAB` is now a `LossySymDataChannel`, which in addition to method `com` offers also method `lossyCom`. The latter does not throw exceptions in case of communication failures, but rather returns an `Optional` value that contains the received value in case of success or is empty otherwise.

Choosing which errors should be dealt with by a choreography and which errors should be raised as unrecoverable exceptions to the local code is a design trade-off that derives from the usual tension between robustness versus simplicity. This trade-off is typical of coordination protocols and exists independently from Choral, which is why we decided to leave the programmer free to navigate this spectrum. Gathering from our own experience with Choral: protocols whose design assumes a reliable network layer should not deal with communication errors within the choreography (e.g., the Diffie-Hellman protocol for key exchange, which we implemented for our evaluation in Section 6); instead, choreographies implementing protocols designed to deal with network errors should specify the handling of those errors (e.g., implementations of objects dedicated to data transfer, like our channels). Choral is quite flexible regarding these aspects, since a channel API can offer both methods that: raise exceptions, like method `com`, meaning that the communication is essential; or wrap failures in data types (no exceptions are raised), like method `lossyCom`, meaning that the communication is not essential and that the choreography can handle internally the failure. The programmer can use the different methods within the same choreography to pinpoint which communications are deemed essential and which are not.

2.6 What goes in a choreography?

We have just looked at the design issue of whether errors should be dealt with in the choreography or in the local code that uses the (communication code compiled from the) choreography. This is an instance of the more general issue of protocol design: what should be part of a choreography? This issue exists even when designing choreographies informally (without Choral) because one needs to choose what details are fixed in the protocol and what is instead left to the discretion of the local code. There is no one-size-fits-all solution since these choices are influenced by the concrete use case that the choreography deals with.

Consider, for example, the widely adopted Diffie-Hellman protocol for cryptographic key exchange [Diffie and Hellman 1976]. Integral parts of the protocol specify both computation and communication. In the protocol, two parties, e.g., `Alice` and `Bob`, use two pairs of keys (a private and a public one) to generate a shared secret, which they can later use for symmetric encryption. Formally, let p be a prime number and g be a primitive root modulo p, sA be a secret key held by `Alice`, and sB be a secret key held by `Bob`. First, `Alice` computes her public key $pA = (g^{sA} \bmod p)$ and, likewise, `Bob` computes his public key $pB = (g^{sB} \bmod p)$. Then, `Alice` and `Bob` exchange their public keys, which they can use to generate their shared secret s:

$$\overbrace{\underbrace{(g^{sB} \bmod p)}_{\texttt{Bob}\text{'s } pB}{}^{sA} \bmod p}^{\texttt{Alice}\text{'s side}} \quad = \quad \underbrace{s}_{\text{shared secret}} \quad = \quad \overbrace{\underbrace{(g^{sA} \bmod p)}_{\texttt{Alice}\text{'s } pA}{}^{sB} \bmod p}^{\texttt{Bob}\text{'s side}} .$$

Since the computations performed by `Alice` and `Bob` are essential to the protocol, any faithful Choral implementation shall include those details too: doing otherwise would mean implementing a different protocol. The following is a snippet of the implementation that we have written for our evaluation in Section 6 (with variables renamed to match our description above).

```
BigInteger@Bob pA = g.modPow(sA, p) >> channel::<BigInteger>com;
BigInteger@Bob s = pA.modPow(sB, p);
```

Choral Code

Differently from these computational details, the Diffie-Hellman protocol does not fix the implementation of the channel used to communicate data. It is therefore reasonable that a Choral implementation of the protocol is parameterised on this implementation. In our case, this is reflected by the fact that `channel` is a parameter of the method that contains our code above.

An example of a choreography where the definition of computation is completely abstracted away is the `consumeItems` method in Section 2.3. The choreography fixes the coordination between the participants, but not how they produce or consume the data to be exchanged. The latter is to be defined by either local code or another choreography that invokes `consumeItems`.

In general, how much computation should be defined in a choreography forms a spectrum. A “good” choreographic programming language should thus give freedom to define or abstract away computation at will. Choral provides this capability through the standard facilities of object-oriented programming (parameters, inheritance, etc.).

3 USE CASES

We dedicate this section to illustrate how the features of Choral contribute in writing realistic choreographies. We start with a protocol for distributed authentication (Section 3.1), which we then reuse modularly in another use case from the healthcare sector that mixes cloud computing, edge computing, and Internet of Things (IoT) (Section 3.2). Finally, we show a use case on parallel computing, by showing a distributed implementation of merge sort (Section 3.3).

3.1 Distributed Authentication

We write a choreography for distributed authentication, inspired by the single sign-on authentication scheme: an `IP` ("Identity Provider", also known as central authentication service) authenticates a `Client` that accesses a third-party `Service`.

We start by introducing an auxiliary class, `AuthResult`, that we will use to store the result of authentication. The idea is that, after performing the authentication protocol, both the `Client` and the `Server` should have an authentication token if the authentication succeeded, or an "empty" value if it failed. We model this by extending the `BiPair` class presented in Section 2.

```
public class AuthResult@(A,B)
  extends BiPair@(A,B)< Optional@A<AuthToken>, Optional@B<AuthToken> > {
 public AuthResult(AuthToken@A t1, AuthToken@B t2) {
  super(Optional@A.<AuthToken>of(t1), Optional@B.<AuthToken>of(t2));
 }
 public AuthResult() {
  super(Optional@A.<AuthToken>empty(), Optional@B.<AuthToken>empty());
}}
```
Choral Code

The constructors of `AuthResult` guarantee that either both roles (`A` and `B`) have an optional containing a value or both optionals are empty (`Optional` is the standard Java type). Since `AuthResult` extends `BiPair`, these values are locally available by invoking the `left` and `right` methods.

We now present the choreography for distributed authentication, as the `DistAuth` class below.

```
enum AuthBranch { OK, KO }

public class DistAuth@(Client,Service,IP){
 private TLSChannel@(Client,IP)<Object> ch_Client_IP;
 private TLSChannel@(Service,IP)<Object> ch_Service_IP;

 public DistAuth(
  TLSChannel@(Client,IP)<Object> ch_Client_IP,
  TLSChannel@(Service,IP)<Object> ch_Service_IP ) {
  this.ch_Client_IP = ch_Client_IP;
  this.ch_Service_IP = ch_Service_IP;
 }

 private String@Client calcHash(String@Client salt, String@Client pwd) { /*...*/ }

 public AuthResult@(Client,Service) authenticate(Credentials@Client credentials) {
  String@Client salt = credentials.username
   >> ch_Client_IP::<String>com >> ClientRegistry@IP::getSalt >> ch_Client_IP::<Stri
```
Choral Code

```
19   Boolean@IP valid = calcHash(salt, credentials.password)
20    >> ch_Client_IP::<String>com >> ClientRegistry@IP::check;
21   if (valid) {
22    ch_Client_IP.<AuthBranch>select(AuthBranch@IP.OK);
23    ch_Service_IP.<AuthBranch>select(AuthBranch@IP.OK);
24    AuthToken@IP t = AuthToken@IP.create();
25    return new AuthResult@(Client,Service)(
26     ch_Client_IP.<AuthToken>com(t), ch_Service_IP.<AuthToken>com(t)
27    );
28   } else {
29    ch_Client_IP.<AuthBranch>select(AuthBranch@IP.KO);
30    ch_Service_IP.<AuthBranch>select(AuthBranch@IP.KO);
31    return new AuthResult@(Client,Service)();
32   }
33 }}
```

Choral Code

Class `DistAuth` is a *multiparty* protocol parameterised over three roles: `Client`, `Service`, and `IP` (for Identity Provider). It composes two channels as fields (lines 4–5), which respectively connect `Client` to `IP` and `Service` to `IP`—hence, interaction between `Client` and `Service` can only happen if coordinated by `IP`. The channels are of type `TLSChannel`, a class for secure channels from the Choral standard library that uses TLS for security and the Kryo library [Grotzke 2020] for marshalling and unmarshalling objects. Class `TLSChannel` implements interface `SymChannel`, from Section 2, so it can be used in both directions. The private method `calcHash` (omitted) implements the local code that `Client` uses to hash its password.

Method `authenticate` (lines 16–33) is the key piece of `DistAuth`, which implements the authentication protocol. It consists of three phases. In the first phase, lines 17–18, the `Client` communicates its `username` to `IP`, which `IP` uses to retrieve the corresponding salt in its local database `ClientRegistry`; the salt is then sent back to `Client`. The second phase (lines 19–20) deals with the resolution of the authentication challenge. `Client` computes its hash with the received salt and its locally-stored password and sends this to `IP`. `IP` then checks whether the received hash is valid, storing this information in its local variable `valid`. The result of the check is a `Boolean` stored in the `valid` variable located at `IP`. The first two phases codify some best practices for distributed authentication and password storage [Grassi et al. 2017]: the identity provider `IP` never sees the password of the client, but only its attempts at solving the challenge (the `salt`), which `Client` can produce with private information (here, its password). In the third phase (lines 21–32), `IP` decides whether the authentication was successful or not by checking `valid`. In both cases `IP` informs the `Client` and the `Service` of its decision, using selections to distinguish between success (represented by `OK`) or failure (represented by `KO`). In case of success, `IP` creates a new authentication token (line 24) and communicates the token to both `Client` and `Service` (inner calls to `com` at line 26). The protocol can now terminate and return a distributed pair (an `AuthResult`) that stores the same token at both `Client` and `Service`, which they can use later for further interactions (line 25). In case of failure, the method returns an authentication result with empty `Optionals` (line 31).

New to choreographic programming, `DistAuth` is a higher-order choreography: the channels that it composes are choreographies for secure communication that carry state—the result of the TLS handshake, which method `com` of `TLSChannel` uses internally. Taking this even further, we could overload method `authenticate` with a continuation-passing style alternative that, instead of or returning a result, takes as parameters choreographic continuations (objects that involve `Client` and `Service`) to be called respectively in case of success (line 25) or failure (line 31).

Compilation. We now discuss key parts of the compilation of DistAuth for role Client, i.e., the Java library that clients can use to authenticate to an identity provider and access a service.

```
1 public class DistAuth_Client {
2  private TLSChannel_A<Object> ch_Client_IP;
3 
4  public DistAuth_Client(TLSChannel_A <Object> ch_Client_IP) {
5   this.ch_Client_IP = ch_Client_IP;
6  }
7 
8  private String calcHash( String salt, String pwd ) { /*...*/ }
9 
10  public AuthResult_A authenticate(Credentials credentials) {
11   String salt = ch_Client_IP.<String>com(ch_Client_IP.<String>com(credentials.username));
12   ch_Client_IP.<String>com(calcHash(salt, credentials.password));
13   switch(ch_Client_IP.<AuthBranch>select(Unit.id)) {
14    case OK -> { return new AuthResult_A( ch_Client_IP.<AuthToken>com(Unit.id), Unit.id); }
15    case KO -> { return new AuthResult_A(); }
16    default -> { throw new RuntimeException( /*...*/ ); }
17 }}}
```

Generated Code

The field, constructor, and method at lines 2–8 are straightforward projections of the source class for role Client—fields and parameters pertaining only other roles disappeared. The interesting code is at lines 10–17, which defines the local behaviour of Client in the authentication protocol. Note that forward chainings (>>) become plain nested calls in Java (lines 11 and 12). In line 11, the client sends its username to the identity provider and receives back the salt. Recall from Section 2 that the innermost invocation of method com returns a Unit, since the client acts as sender here. Once the username is sent, the innermost com returns and we run the outermost invocation of com, which received the salt through the channel with the identity provider. In line 12, the Client sends the computed hash to the identity provider.

Line 13 exemplifies how our compiler implements knowledge of choice for roles that need to react to decisions made by other roles. The client receives an enumerated value of type AuthBranch, which can be either OK or KO, through the channel with the identity provider. Then, a switch statement matches the received value to decide whether (case OK) we shall receive an authentication token from the identity provider and store it as an AuthResult_A or (case KO) authentication failed.

3.2 A use case from healthcare: handling streams of sensitive vitals data

In this use case, we exemplify how developers can compose locally the libraries generated by independent choreographies, using a healthcare use case inspired by previous works on edge computing and pseudonimisation [Giallorenzo et al. 2019; Swaroop et al. 2019].

Suppose that a "healthcare service" in a hospital needs to gather sensitive data about vital signs (we call them vitals) from some IoT devices (e.g., smartwatches, heart monitors), and then upload them to the cloud for storage. This is a typical scenario that requires integration of libraries for participating in choreographies at the local level. We shall carry out the following two steps.

(1) Define a new choreography class, called VitalsStreaming, that prescribes how data should be streamed from an IoT Device monitoring the vitals of a patient to a data Gatherer; this choreography will enforce that the Gatherer processes only data that is (a) correctly cryptographically signed by the device and (b) pseudonymised.

(2) Implement the healthcare service a local Java class, called HealthCareService, that combines the Java library compiled from VitalsStreaming to gather data from the IoT devices

with the Java library compiled from our previous `DistAuth` example to authenticate with the cloud storage service through a third-party service (this could be, e.g., a national authentication system) and upload the data.

Vitals choreography. `VitalsStreaming` implements the choreography for streaming vitals.

```
1  public enum StreamState@R { ON, OFF }
2
3  public class VitalsStreaming@(Device,Gatherer) {
4   private SymChannel@(Device,Gatherer)<Object> ch;
5   private Sensor@Device sensor;
6
7   public VitalsStreaming(SymChannel@(Device,Gatherer)<Object> ch, Sensor@Device sensor) {
8    this.ch = ch;
9    this.sensor = sensor;
10  }
11
12  private static Vitals@Gatherer pseudonymise(Vitals@Gatherer vitals) { /*...*/ }
13  private static Boolean@Gatherer checkSignature(Signature@Gatherer signature) { /*...*/ }
14
15  public void gather(Consumer@Gatherer<Vitals> consumer) {
16   if (sensor.isOn()) {
17    ch.<StreamState>select(StreamState@Device.ON);
18    VitalsMsg@Gatherer msgOpt = sensor.next() >> ch::<VitalsMsg>com;
19    if (checkSignature(msg.signature())) {
20     msg.content() >> this::pseudonymise >> consumer::accept;
21    }
22    gather(consumer);
23   } else {
24    ch.<StreamState>select(StreamState@Device.OFF);
25 }}}
```

Choral Code

In lines 3–5, class `VitalsStreaming` composes a channel between the `Device` and the `Gatherer` and a `Sensor` object located at the `Device` (for obtaining the local vital readings). Line 12 defines a method that pseudonymises personal data in `Vitals` at the `Gatherer`. Likewise, line 13 contains a method that the `Gatherer` uses to check that a message signature is valid. (We omit the bodies of these two static methods, which are standard local methods.) The interesting part of this class is method `gather` (lines 15–25). The `Device` checks whether its sensor is on (line 16) and informs the `Gatherer` of the result with appropriate selections for knowledge of choice (lines 17 and 24). If the sensor is on, then `Device` sends its next available reading to `Gatherer` (line 18). `Gatherer` now checks that the message is signed correctly (line 19); if so, it pseudonymises the content of the message and then hands it off to a local consumer function. Notice that `Gatherer` does not need to inform `Device` of its local choice since it does not affect the code that `Device` needs to run. We then recursively invoke `gather` to process the next reading.

Local code of the healthcare service. The local implementation of the healthcare service acts as `Gatherer` in the `VitalsStreaming` choreography (to gather the data) and as the `Client` in the `DistAuth` choreography (to authenticate with the cloud storage). So we compose the compiled Java classes `VitalsStreaming_Gatherer` and `DistAuth_Client`, respectively.

```
public class HealthCareService {
 public static void main() {
  TLSChannel_A toIP = HealthIdentityProvider.connect();
```

Local Code

```
4     MQTTClient toStorage = HealthDataStorage.connect();
5     AuthResult_A authResult = new DistAuth_Client(toIP).authenticate(getCredentials());
6     authResult.left().ifPresent(token -> {
7      DeviceRegistry
8       .parallelStream()
9       .forEach(device ->
10       Supervision.restart(() ->
11        new VitalsStreaming_Gatherer(device.connect())
12         .gather(data -> toStorage.com(new StorageMesg(token, data)))
13       )
14      );
15    });
16   }
17   private static Credentials getCredentials() { /* ... */ }
18  }
```

Local Code

The main method above idiomatically combines Java standard libraries with those generated by our compiler. In lines 3 and 4, we use auxiliary methods to connect to the identity provider (which implements IP in DistAuth) and the data storage service (which implements Service in DistAuth)—these services are provided by third parties, e.g., the national health system and some cloud provider. We choose a TLS channel to enact authentication. Instead, for communications with devices and storage, we use the MQTT protocol, which is typical for IoT applications [Hunkeler et al. 2008]—MQTTClient implements (the projection of) interface SymChannel, such that connectivity issues are dealt with. In line 5, we run distributed authentication as the Client. In line 6, we check if we successfully received an authentication token by inspecting the optional result. If so, we obtain a parallel stream of Device objects from a local registry (lines 7–8). Each device (line 9) is handled by a worker that uses a restart supervision strategy (line 10), i.e., if the projection of the choreography VitalsStreaming encounters an unrecoverable error, it is restarted. In line 11, we create a new instance of VitalsStreaming_Gatherer (the code compiled for Gatherer from VitalsStreaming), which is passed an MQTT channel for communicating with the device (obtained by device.connect()). Finally, in line 12, we call the gather method to engage in the VitalsStreaming choreography with each device, passing a consumer function that sends the received data to the cloud storage service (including the authentication token).

Notice that we do not need to worry about pseudonymisation or signature checking in the local code since all these details are dealt with by the code compiled from VitalsStreaming.

3.3 Mergesort

The last use case that we present is a three-way concurrent implementation of merge sort [Knuth 1998], which exemplifies the design of parallel algorithms in Choral. Our implementation leverages role parameterisation such that participants collaboratively switch the roles that they play at runtime.

We represent the three concurrent parties as the roles A, B, and C. The idea is to follow the steps of standard merge sort, with A acting as "master" and the other as workers. Specifically, A divides the unsorted list into two sublists and then communicates them to B and C, respectively. We then recursively invoke merge sort on each sublist, but with *switched roles*: in one call, B becomes the master that uses A and C as workers; in the other call, C is the master using A and B as workers. B and C then return their sorted sublists to A, which can merge them as usual.

The sequence diagram in Figure 4 represents the pattern of how data is exchanged in our choreography by three endpoint nodes for an input list $[15, 3, 14]$ (we omit selections).

We use numbered subscripts to denote the round that each interaction belongs to. Node$_1$ starts by playing role A and holds the initial list, while the other two nodes initially play the worker roles. In the first round, Node$_1$ asks Node$_2$ and Node$_3$ to sort the sublists obtained from the initial list. This starts a recursive call (second round) where Node$_2$ is the master and the others are workers that help it to sort its sublist. Node$_2$ now splits its sublist into smaller lists and asks the other two nodes to sort them (sort$_2$). When this round is completed, each node contains a sorted sublist, and we can get up the recursion stack to the nodes playing their original roles, where now A collects the results from the others (B and C coordinate to decide who communicates first). The logic that we have just described is implemented by the following `Mergesort` class.

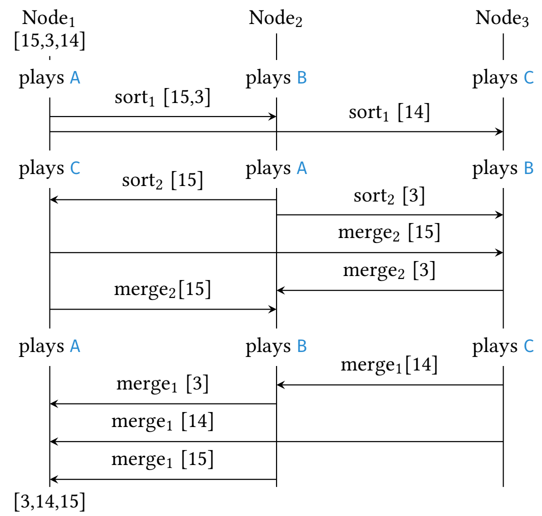

Fig. 4. Sequence diagram of data exchanges in the three-way distributed merge sort (selections omitted).

```
enum Choice@R { L, R }
public class Mergesort@(A,B,C){
 SymChannel@(A,B)<Object> ch_AB;
 SymChannel@(B,C)<Object> ch_BC;
 SymChannel@(C,A)<Object> ch_CA;

 public Mergesort(
  SymChannel@(A,B)<Object> ch_AB,
  SymChannel@(B,C)<Object> ch_BC,
  SymChannel@(C,A)<Object> ch_CA ) {
  this.ch_AB = ch_AB; this.ch_BC = ch_BC; this.ch_CA = ch_CA;
 }

 public List@A<Integer> sort(List@A<Integer> a){
  if (a.size() > 1@A) {
   ch_AB.<Choice>select(Choice@A.L);
   ch_CA.<Choice>select(Choice@A.L);
   Mergesort@(B,C,A) mb = new Mergesort@(B,C,A)(ch_BC, ch_CA, ch_AB);
   Mergesort@(C,A,B) mc = new Mergesort@(C,A,B)(ch_CA, ch_AB, ch_BC);
   Double@A pivot = a.size() / 2@A >> Math@A::floor>> Double@A::valueOf;
   List@B<Integer> lhs = a.subList(0@A,pivot.intValue())
    >> ch_AB::<List<Integer>>com >> mb::sort;
   List@C<Integer> rhs = a.subList(pivot.intValue(), a.size())
    >> ch_CA::< List<Integer> >com >> mc::sort;
   return merge(lhs, rhs);
  } else {
   ch_AB.<Choice>select(Choice@A.R);
   ch_CA.<Choice>select(Choice@A.R);
   return a;
```
Choral Code

```
30      }
31    }
32
33    private List@A<Integer> merge(List@B<Integer> lhs, List@C<Integer> rhs) {
34     if( lhs.size() > 0@B ) {
35      ch_AB.< Choice >select( Choice@B.L );
36      ch_BC.< Choice >select( Choice@B.L );
37      if( rhs.size() > 0@C ){
38       ch_CA.< Choice >select( Choice@C.L );
39       ch_BC.< Choice >select( Choice@C.L );
40       ArrayList@A< Integer > result = new ArrayList@A< Integer >();
41       if( lhs.get( 0@B ) <= ch_BC.< Integer >com( rhs.get( 0@C ) ) ){
42        ch_AB.< Choice >select( Choice@B.L );
43        ch_BC.< Choice >select( Choice@B.L );
44        lhs.get( 0@B ) >> ch_AB::< Integer >com >> result::add;
45        merge( lhs.subList( 1@B, lhs.size() ), rhs ) >> result::addAll;
46        return result;
47       } else {
48        ch_AB.< Choice >select( Choice@B.R );
49        ch_BC.< Choice >select( Choice@B.R );
50        rhs.get( 0@C ) >> ch_CA::< Integer >com >> result::add;
51        merge( lhs, rhs.subList( 1@C, rhs.size() ) ) >> result::addAll;
52        return result;
53       }
54      } else {
55       ch_CA.< Choice >select( Choice@C.R );
56       ch_BC.< Choice >select( Choice@C.R );
57       return lhs >> ch_AB::< List< Integer > >com;
58      }
59     } else {
60      ch_AB.< Choice >select( Choice@B.R );
61      ch_BC.< Choice >select( Choice@B.R );
62      return rhs >> ch_CA::< List< Integer > >com;
63   }}}
```
Choral Code

The sorting algorithm is implemented by the `sort` method, which uses the private `merge` method to recursively handle the point-wise merging of ordered lists. For lists of size greater than 1, the algorithm creates two new `Mergesort` objects by instantiating roles such that they get switched as we discussed (lines 18–19), splits the list at the master, communicates the resulting sublists to the workers (lines 21–22 and 23–24), recursively invokes merge sort with the switched roles (still lines 21–22 and 23–24), and finally merges the results (line 25).

The remaining code resembles (the choreography of) typical parallel merge sort implementations. A key benefit of Choral for parallel programming is that the compiled code is deadlock-free by construction, as usual for choreographic programming [Carbone and Montesi 2013].

4 IMPLEMENTATION

We discuss the main elements of the implementation of Choral. First, we show its syntax and comment on the main differences with Java's. Then, we present the Choral type checker, including examples of the main errors related to roles that it detects and related error messages. Finally, we describe the key components of the Choral compiler.

Literals	*lit*	::=	`null@(A)` | `true@A` | `false@A` | `"a"@A` | ... | `1@A` | ...
Program	*P*	::=	*P Interface* | *P Class* | *P Enum* | *P EOF*
Enum	*Enum*	::=	*AN MD* **enum** *id*@A{*id*}
Interface	*Interface*	::=	*AN MD* **interface** *id*@(A)⟨*FTP*⟩ **extends** *TE*, *TE*{*MDef*;}
Annotation	*AN*	::=	@*id*(*id* = *lit*)
Modifiers	*MD*	::=	**public** | **protected** | **private** | **abstract** | **final** | **static**
Formal Type Param.	*FTP*	::=	*id*@(A) **extends** *TE* & *TE*
Type Expr.	*TE*	::=	*id*⟨*TE*⟩ | *id*@(A)⟨*TE*⟩ | **void**
Method Def.	*MDef*	::=	*AN MD* ⟨*FTP*⟩ *TE id* (*TE id*)
Class	*Class*	::=	*AN MD* **class** *id*@(A)⟨*FTP*⟩ **extends** *TE* **implements** *TE*, *TE* {*CField CConst MDef*; *MDef*{*Stm*}}
Class Field.	*CField*	::=	*AN MD TE id*;
Class Con.	*CConst*	::=	*AN MD* ⟨*FTP*⟩ *id*(*TE id*){*Stm*}
Statement	*Stm*	::=	*nil* | **return** *Exp*; | *Exp*; *Stm* | *TE id* = *Exp*; *Stm*
		|	*Exp AsgOp Exp*; *Stm* | **if**(*Exp*){*Stm*}**else**{*Stm*} *Stm*
		|	{*Stm*} *Stm* | **try**{*Stm*}**catch**(*TE id*){*Stm*} *Stm*
Expression	*Exp*	::=	*lit* | *FAcc* | *Exp BinOp Exp* | *Exp*.*Exp* | ⟨*TE*⟩*id*(*Exp*)
		|	**new** ⟨*TE*⟩*id*@(A)⟨*TE*⟩(*Exp*) | *id*@(A).⟨*TE*⟩*id*(*Exp*) | *Exp* >> *EChain*
Field Acc.	*FAcc*	::=	*id* | *id*@(A).*id*
Exp. Chain	*EChain*	::=	*FAcc*.*id*::*id* | *id*@(A)⟨*TE*⟩::**new**
Assign Op.	*AsgOp*	∈	{=, +=, -=, *=, /=, &!=, |!=, %!=}
Binary Op.	*BinOp*	∈	{||, &&, |, &, ==, !=, <, >, <=, >=, +, -, *, /, %}

Fig. 5. Syntax of the Choral language.

4.1 Language

Figure 5 displays the grammar of Choral; dashed underlines denote optional terms and solid overlines denote sequences of terms of the same sort. We omit syntax for packages and imports, which is as in Java. Reserved identifies like **super** and **this** are considered as identifiers in the grammar but are given treated by out compiler like their Java counterpart. The key syntactic novelties are underlined; they consist of *i*) syntax for declaring and instantiating role parameters and *ii*) the forward chaining operator `>>` (cf. Section 2).

Role parameters have a separate namespace, and always appear in expressions like `@(A1,...,An)` that follow the name of a class, interface, enum, or type parameter e.g., `DiChannel@(A,B)`. Also, they are introduced only by the declaration of a type (e.g., `class Foo@(A,B)`) or a type parameter (e.g., `<T@(A,B) extends Foo@(A,B) & Bar@(B,A)>`) and their scope is limited to the defining type, similar to type parameters in Java. The snippet below contains an example of shadowing of role parameters; for each use of role `A`, we show its binding site with an arrow.

```
interface Foo@(A,B) extends Bar@(A,B) { <T@(A,B) extends Foo@(A,B) & Bar@(B,A)> T@(A,B) m();}
```

The Choral type checker covers all common Java type errors (illegal type conversions, access to type members, etc.), as exemplified below. Indeed, when checking a Choral program with exactly one role parameter, the Choral type-checker acts exactly like its Java counterpart.

Compiler error

```
Integer@A x = "foo"@A;
------------^
Incompatible types: expecting 'Integer@A' found 'String@A'.

return x.length();
---------^
Cannot resolve method 'length' in 'Integer@(A)'
```

Roles. The novelties compared to Java compilers emerge when two or more roles are involved since in this settings programmers can make new errors that are pertinent to Choral and specifically due to the misuse of role parameterised types. In many of the examples discussed so far in the paper, role parameters can be thought of as Java generics. Although this is a good working approximation, some care is necessary in handling type instantiation due to some substantial differences between role and type parameters.

One type of such errors is that data types are instantiated using incompatible roles. Instances of the same type with different roles parameters represent values located at different roles and this restricts their usage.

Compiler error

```
String@A x = "foo"@B; // error, same local type but at different roles
-----------^
Incompatible types: expecting 'String@A' found 'String@B'.
```

The order in which roles aper carries meaning since role parameters are positional—like type parameters in Java generics.

Compiler error

```
void m(SymChannel@(A,B)<T> x) {
  SymChannel@(A,B)<T> a = x; // matching roles
  SymChannel@(B,A)<T> b = x; // error, same roles but wrong positions
------------------------^
Incompatible types: expecting 'SymChannel@(B,A)<T>' found 'SymChannel@(A,B)<T>'.
```

Differently from Java generics, role parameter cannot appear multiple times in the same type since this corresponds to requiring that they play multiple roles in the same choreography. In the snippet below, `A` must play both the sender and receiver for the directed channel `c`.

Compiler error

```
DiChannel@(A,A)<String> c;
-------------^
Illegal type instantiation: role 'A' must play exactly one role in 'DiChannel'.
```

Forbidding role aliasing is an established restriction in choreographic programming since aliasing introduces self-communication, which would potentially break deadlock-freedom and the capability to produce separate code for each role (unless roles are provably not-aliased).

Subtyping. Choral types form a hierarchy defined following the same principles used by Java. This hierarchy is used to check if values are compatible type as expected.

```
void m(BiChannel@(A,B)<T,T> x) {                                              Compiler error
  DiChannel@(A,B)<T>  a=x; // BiChannel@(A,B)<T,T> extends DiChannel@(A,B)<T>
  DiChannel@(B,A)<T>  b=x; // BiChannel@(A,B)<T,T> extends DiChannel@(B,A)<T>
  SymChannel@(A,B)<T> c=x; // error, BiChannel@(A,B)<T,T> does not extend SymChannel@(A,B)<T>
-----------------------^
Incompatible types: expecting 'SymChannel@(A,B)<T>' found 'BiChannel@(A,B)<T>'.
```

Classes and interfaces define their supertypes by extending and implementing other classes and their interfaces with the same set of roles. This restriction provides a substitution principle that elicits all roles involved in a choreography.

```
interface AuditedDiChannel@(A,B,Auditor)<T@C> extends DiChannel@(A,B)<T> {/*...*/}   Compiler error
------------------------------------------------------^
 Illegal inheritance: 'AuditedDiChannel@(A,B,Auditor)' and 'DiChannel@(A,B)<T>' must have the
      same roles.

interface ReplicatedList@(A,Replica)<T@B> extends List@A<T> {/*...*/}
--------------------------------------------------^
 Illegal inheritance: 'ReplicatedList(A,Replica)' and 'List@A<T>' must have the same roles.
```

In some cases, "hidden roles" in choreographies might be useful, e.g., to add external auditing or data replication as an extension of an existing choreography. Alas, this introduces security concerns (channels may have hidden bystanders) or complex communication semantics (what is the meaning of sending a `ReplicatedList@(A,B)` over a channel expecting a `List@A`?). These are general open problems for choreographies, left to future work.

Cyclic inheritance is not allowed and the type checker does not discriminate over role parameters. As an example, consider the `SymChannel` interface; given its symmetric nature, one might be tempted to force this equality by having `SymChannel@(A,B)` subtype `SymChannel@(B,A)`.

```
interface SymChannel@(A,B)<T@C> extends SymChannel@(B,A)<T> { /* ... */ }   Compiler error
----------------------------------------^
Cyclic inheritance: 'SymChannel' cannot extend 'SymChannel'.
```

However, allowing declarations like the one above in Choral would result in cyclic inheritance errors in Java, as exemplified by the (manual) projection below.

```
interface SymChannel_A<T> extends SymChannel_B<T> { /* ... */ } // Projection for A
interface SymChannel_B<T> extends SymChannel_A<T> { /* ... */ } // Projection for B
```

To have channels that are instances of both `SymChannel@(A,B)` and `SymChannel@(B,A)` one needs to define a subtype of both as in the snippet below.

```
interface PeerChannel@(A,B)<T@C> extends SymChannel@(A,B)<T>, SymChannel@(B,A)<T> {
  <S@C extends T@C> S@B com(S@A m);           // inherited
  <S@C extends T@C> S@A com(S@B m);           // inherited
  <S@C extends Enum@C<S>> S@B select(T@A m); // inherited
  <S@C extends Enum@C<S>> S@A select(T@B m); // inherited

  PeerChannel@(B,A)<T> flip();               // roles A and B are interchangeable
}                                                                            Choral Code
```

By returning an instance of the same interface but with the roles flipped, the method `flip()` introduced by the interface `PeerChannel@(A,B)`, prescribes that the roles `A` and `B` are interchangeable peers.

Finally, primitive types (`int@A`, `bool@A`, etc.) follow the same rules of Java for subtyping, conversions, autoboxing, and autounboxing (when roles match, otherwise the compiler will return a role mismatch error).

Overloading. The Choral type checker refines overload equivalence: it can discriminate overloaded methods by considering roles (e.g., `m(Char@B x)` and `m(Char@A x)` below). It also predicts potential clashes in the compiled Java code. Consider the following snippet and error message.

Compiler error

```
class Foo@(A,B) {
  void m(Char@B x) { /* ... */ }  //        void m() at A and void m(Char x) at B
  void m(Char@A x) { /* ... */ }  //        void m(Char x) at A and void m() at B
  void m(Long@A x) { /* ... */ }  // error, void m(Long x) at A and void m() at B
-------^
Illegal overload: 'm(Long@A x)' and 'm(Char@A x)' have the same signature for role 'B'.
```

The last two signatures are distinguishable in Choral, since each method has different parameter types. However, this information is only available to role `A`, while the projection of both signatures at role `B` coincide. This is an instance of knowledge of choice but, differently from conditionals, it cannot be addressed locally (within the class/interface) because extending classes may introduce new branches and new points of choice by overriding and overloading, as in the example below.

Choral Code

```
class Bar@(A,B) extends Foo@(A,B) { void m(Integer@A x) { /* ... */ } }
```

Exceptions. Like every other type lifted from Java, exceptions are located at one role. This design choice allows us to preserve the expected type hierarchy in the generated code and have `java.lang.Exception` as the supertype of all exceptions. The Choral compiler then enforces that a **`try-catch`** block is located at exactly one role.

Compiler error

```
String@A fetch(DiChannel@(A,B)<T> ch, String@B file) {
  try { return RemoteReader@(A,B).read(ch,file); } catch (IOException@B e) { return null@A; }
--^
Non-local try-catch: try-catch must be at a single role, found 'A' and 'B'.
```

Allowing multiple roles in the mechanics of exceptions introduces a knowledge-of-choice situation where all roles need to obtain information about which handler to execute, if any, and when. The theory of choreographies has explored the problem of exceptions and their chorographic handling but all models proposed so far assume reliable communications among all roles and either rely on specialised orchestration primitives in the target language (i.e., some form of middleware) [Carbone 2009; Carbone et al. 2008; Fowler et al. 2019] or synthesise new communications for recovery [Neykova and Yoshida 2017].

4.2 Compiler

Our compiler consists of a pipeline of three stages: parsing, type checking, and projection. Parsing is unsurprising, so we do not describe it here. Type checking operates as we have just discussed. Projection is the component that, given well-typed Choral code, produces a choreography-compliant Java library for each role.

We discuss the most important parts of projection, reporting its full formalisation in Appendix A. The projection of a Choral **`class`**, **`interface`**, or **`enum`** generates a corresponding Java term for each role parameter. If there are two or more roles, each Java artefact name is suffixed with the role that it implements, e.g., the Java class compiled from `class Foo(A,B)` for role A is called `Foo_A`. If the Choral class has exactly one role, then we use the same name, e.g., `class Integer@A` becomes `class Integer`. (This erases friction for the integration of Java types within Choral.)

Formally, projection is a (partial) function that, given a Choral term *Term* and the role A that we wish to generate the Java implementation of, returns a Java term, written $(\!|Term|\!)^{\mathsf{A}}$. The projection $(\!|TE|\!)^{\mathsf{A}}$ of a type expression *TE* at a role A is recursively defined below—we use the auxiliary function roleName(id, i) to retrieve the name of the i-th role parameter from the definition of *id*.

$$(\!|id\texttt{@}(\overline{\mathsf{B}})\texttt{<}\overline{TE}\texttt{>}|\!)^{\mathsf{A}} = \begin{cases} id\texttt{<}\overline{(\!|TE|\!)^{\mathsf{A}}}\texttt{>} & \text{if } \overline{\mathsf{B}} = \mathsf{A} \\ id\_A'\texttt{<}\overline{(\!|TE|\!)^{\mathsf{A}}}\texttt{>} & \text{if } \mathsf{A} \text{ is the } i\text{-th element of } \overline{\mathsf{B}} \text{ and roleName}(id, i) = A' \\ \texttt{Unit} & \text{otherwise} \end{cases}$$

The projection $(\!|Exp|\!)^{\mathsf{A}}$ of an expression *Exp* at role A is defined following a similar intuition: it is a recursive stripping of role information as long as A occurs in the type of *Exp* or any of its subterms (written $\mathsf{A} \in \text{rolesOf}(Exp)$), otherwise it is the only instance of the singleton `Unit` (stored in its static field `id`), as illustrated by the cases of static field access and constructor invocation below.

$$(\!|id\texttt{@}(\overline{\mathsf{B}})\texttt{.}f|\!)^{\mathsf{A}} = \begin{cases} (\!|id\texttt{@}(\overline{\mathsf{B}})|\!)^{\mathsf{A}}\texttt{.}f & \text{if } \mathsf{A} \in \text{rolesOf}(f) \\ \texttt{Unit.id} & \text{otherwise} \end{cases}$$

$$(\!|\textbf{new}\ \langle\overline{TE}\rangle\, id\texttt{@}(\overline{\mathsf{B}})\langle\overline{TE}\rangle(\overline{Exp})|\!)^{\mathsf{A}} = \begin{cases} \textbf{new}\ \langle\overline{(\!|TE|\!)^{\mathsf{A}}}\rangle(\!|id\texttt{@}(\overline{\mathsf{B}})\langle\overline{TE}\rangle|\!)^{\mathsf{A}}(\overline{(\!|Exp|\!)^{\mathsf{A}}}) & \text{if } \mathsf{A} \in \overline{\mathsf{B}} \\ \texttt{Unit.id}(\overline{(\!|Exp|\!)^{\mathsf{A}}}) & \text{otherwise} \end{cases}$$

The projection $(\!|Stm|\!)^{\mathsf{A}}$ of a statement *Stm* at A is defined following the above intuition, save for the cases of conditionals and selections, which require care to address knowledge of choice (cf. Section 2.3). Specifically, the rule for projecting **if** statements: for the role evaluating the guard (read from its type), it preserves the conditional; for all other roles, the **if** disappears and it is replaced by the projection of the guard (since it might have side-effects) followed by the *merging* $\sqcup$ of the projections of the bodies of the two branches and the projection of the continuation *Stm*.

$$(\!|\textbf{if}(Exp)\{Stm_1\}\textbf{else}\{Stm_2\}Stm|\!)^{\mathsf{A}} = \begin{cases} \textbf{if}((\!|Exp|\!)^{\mathsf{A}})\{(\!|Stm_1|\!)^{\mathsf{A}}\}\textbf{else}\{(\!|Stm_2|\!)^{\mathsf{A}}\}(\!|Stm|\!)^{\mathsf{A}} & \text{if } Exp\texttt{: boolean@A} \\ (\!|Exp|\!)^{\mathsf{A}};\left\{(\!|Stm_1|\!)^{\mathsf{A}} \sqcup (\!|Stm_2|\!)^{\mathsf{A}}\right\}(\!|Stm|\!)^{\mathsf{A}} & \text{otherwise} \end{cases}$$

The merge operator $Stm \sqcup Stm'$ is a partial operator that tries to combine branching code [Carbone et al. 2012], which we adapt to Java for the first time. Essentially, given two Java terms, merging recursively requires them to be equivalent *unless* they are switch statements. Appendix A contains the full definition of merging. Here we report its most interesting case: merging switch statements.

$$\begin{array}{l} \textbf{switch}\ (Exp)\{ \\ \quad \textbf{case}\ id_a\texttt{->}\{Stm_a\} \\ \quad \cdots \\ \quad \underline{\textbf{case}\ id_x\texttt{->}\{Stm_x\}} \\ \quad \textbf{case}\ id_y\texttt{->}\{Stm_y\} \\ \quad \textbf{default}\texttt{->}\{Stm_{d1}\} \\ \}\ Stm \end{array} \sqcup \begin{array}{l} \textbf{switch}\ (Exp)\{ \\ \quad \textbf{case}\ id_a\texttt{->}\{Stm'_a\} \\ \quad \cdots \\ \quad \underline{\textbf{case}\ id_x\texttt{->}\{Stm'_x\}} \\ \quad \textbf{case}\ id_z\texttt{->}\{Stm_z\} \\ \quad \textbf{default}\texttt{->}\{Stm_{d2}\} \\ \}\ Stm' \end{array} = \begin{array}{l} \textbf{switch}\ (Exp \sqcup Exp')\{ \\ \quad \textbf{case}\ id_a\texttt{->}\{Stm_a \sqcup Stm'_a\} \\ \quad \cdots \\ \quad \underline{\textbf{case}\ id_x\texttt{->}\{Stm_x \sqcup Stm'_x\}} \\ \quad \underline{\textbf{case}\ id_y\texttt{->}\{Stm_y\}} \\ \quad \textbf{case}\ id_z\texttt{->}\{Stm_z\} \\ \quad \textbf{default}\texttt{->}\{Stm_{d1} \sqcup Stm_{d2}\} \\ \}\ Stm \sqcup Stm' \end{array}$$

Above, the merging of two switch statements is a switch whose guard is the merging of the original guards ($Exp \sqcup Exp'$). Its cases consist of: for each case present in both the input switches ($SwArg_a, \cdots, SwArg_x$), we get a case in the result whose body merges the respective bodies of the original cases; all cases that are not shared, which are simply put in the result as they are (the lists of cases $\overline{\texttt{case}\ SwArg_y\ \texttt{->}\ Stm_y}$ from the first and $\overline{\texttt{case}\ SwArg_z : Stm_z}$ from the second).

An example of the result of merging was presented for `DistAuth_Client` in Section 3.1, where the cases for `OK` and `KO` are combined from the respective projections for `Client` of the two branches in the source choreographic conditional evaluated by `IP`. These cases are produced by the rule for projecting selections, which applies to statements of the form Exp`;` Stm when Exp calls (possibly in a chain call) a method annotated with `@SelectionMethod`. (Our type checker checks that these annotations are used only for methods that take enumerated types as parameters, cf. Section 2.3.) For compactness, let $S = \overline{Exp}.\langle\overline{TE}\rangle id_1(id_2@\mathsf{A}'.id_3)$ where `@SelectionMethod` $\in$ `annotations`(id_1).

$$
(\!|S;Stm|\!)^{\mathsf{A}} = \begin{cases} \texttt{switch}((\!|S|\!)^{\mathsf{A}}) \left\{ \begin{array}{l} \texttt{case}\ id_3\texttt{->\{}(\!|Stm|\!)^{\mathsf{A}}\texttt{\}} \\ \texttt{default ->\{throw new ...\}} \end{array} \right\} & \text{if } S\colon \texttt{Enum<T>@A} \text{ for some } \texttt{T} \\ (\!|S|\!)^{\mathsf{A}};\ (\!|Stm|\!)^{\mathsf{A}} & \text{otherwise} \end{cases}
$$

For the recipient of the selection (first case), the statement becomes a switch on the projection of the *Exp*ression that will receive the selection, while the projection of the continuation *Stm* becomes the body of the corresponding case in the argument. The projection for the other roles (second case) is standard, projecting the *Exp*ression followed by the projection of the continuation *Stm*.

Our implementation of merging is smart enough to deal with some "non-effectful" usages of `Unit`. For instance, consider the following choreography.

```
if(true@A){System@A.out.println("true"@A);}
```

If we project it at a role different from `A`, say `B`, we obtain the code `Unit.id(Unit.id)` for the then-branch, and [*blank*] for the (missing) else-branch. These fragments are not mergeable, but our compiler uses a *unit-normalising operator*, given in Appendix A, which transforms also the first fragment into [*blank*] by removing the irrelevant usages of `Unit`.

5 TESTING

Testing implementations of choreographies is challenging since the distributed programs of all participants need to be integrated (integration testing). We introduce ChoralUnit, a testing tool that enables the writing of integration tests as simple unit tests for choreographic classes.

Writing Tests. Following standard practice in object-oriented languages and inspired by JUnit, tests in ChoralUnit are defined as methods marked with a `Test` annotation [Gamma and Beck 2006; Hamill 2004]. For example, we can define the following unit test for the `VitalsStreaming` class from Section 3.2.

```
public class VitalsStreamingTest@(Device, Gatherer) {
 @Test
 public static void test1(){
  SymChannel@(Device, Gatherer)<Object> ch =
   TestUtils@(Device, Gatherer).newLocalChannel("VST_channel1"@[Device, Gatherer]);
  new VitalsStreaming@(Device, Gatherer)(ch, new FakeSensor@Device())
   .gather(new PseudoChecker@Gatherer());
}}

class PseudoChecker@R implements Consumer@R<Vitals> {
 public void accept(Vitals@R vitals){
  Assert@R.assertTrue("bad pseudonymisation"@R, isPseudonymised(vitals));
 }
 private static Boolean@R isPseudonymised(Vitals@R vitals) { /* ... */ }
```

Choral Code

```
15 }
16
17 class FakeSensor@R implements Sensor@R { /* ... */ }
```

Choral Code

The test method `test1` checks that data is pseudonymised correctly by `VitalsStreaming`. Test methods must be annotated with `@Test`, be **`static`**, have no parameters, and return no values.

In lines 4–5, we create a channel between the `Device` and the `Gatherer` by invoking the `TestUtils.newLocalChannel` method, which is provided by ChoralUnit as a library to simplify the creation of channels for testing purposes. This method returns an in-memory channel, which both `Device` and `Gatherer` will find by looking it up in a shared map under the key `"VST_channel1"`. Thus, both roles must have the same key in their compiled code, which is guaranteed here by the fact that the expression `"VST_channel1"@[Device,Gatherer]` is syntax sugar for `"VST_channel1"@Device, "VST_channel1"@Gatherer`.

In lines 6–7, we create an instance of `VitalsStreaming` (the choreography we want to test). We use a `FakeSensor` object to simulate a sensor that sends some data containing sensitive information (omitted). We then invoke the `gather` method, passing an implementation of a consumer that checks whether the data received by the `Gatherer` has been pseudonymised correctly.

Given a class like `VitalsStreamingTest`, ChoralUnit will compile it by invoking our compiler with a special flag (`-annotate`). This makes the compiler annotate each generated Java class with a `@Choreography` annotation that contains the name of its source Choral class and the role that the Java class implements. Once compilation is finished, the ChoralUnit tool can be invoked to run the tests in the `VitalsStreamingTest` class. This happens in three steps: (1) ChoralUnit finds all Java classes annotated with a `@Choreography` annotation whose `name` value corresponds to `VitalsStreamingTest`. (2) Each discovered class has a method with the same name for each method in the source Choral test class (`test1` in our example). For each such method that is annotated with `@Test`, ChoralUnit starts a thread running the local implementation of the method by each class generated from the Choral source. (3) The previous step is repeated for all test methods.

In our example, `VitalsStreamingTest` is compiled to a class for `Device` and another for `Gatherer`, each with a `test1` method. Thus, ChoralUnit starts two threads, one running `test1` of the first generated Java class and the other running `test1` of the second generated Java class.

Multiparty Assertions. In the previous example, we have written an assertion (in `PseudoChecker`) that checks a condition at a single role (`Gatherer`). Sometimes, it is useful to assert conditions that involve multiple roles. A typical example is testing the correct implementation of protocols that aim at making two parties agree on a symmetric cryptographic key, like the Diffie-Hellman protocol [Diffie and Hellman 1976]. In particular, after running the protocol, the two participants (say `A` and `B`) should have the same key. We can express this assertion as follows.

```
1 Assert2@(A, B).assertEquals("key mismatch"@B, chAB, keyA@A, keyB@B);
```

Choral Code

Above, method `assertEquals` of class `Assert2` uses the channel `chAB` to communicate the key at `A` (`keyA`) from `A` to `B`, and then checks locally at `B` that it is equal to the key at `B` (`keyB`). If the check fails, an assertion error is raised at `B`.

Class `Assert2` can be user-defined, and likewise developers can define classes that allow for assertions that involve more roles (e.g., `Assert3`, `Assert4`, etc.). In these implementations, the user can also freely code different protocols for communicating the data among the participants.

6 EVALUATION

In Section 3, we explored how Choral can be used to program choreographies for a few realistic scenarios. In this section, we extend the evaluation of our approach in three different directions:

(1) In Section 6.1, we exemplify how one can use Choral to transition existing (Java) programs to choreographies. In Section 6.1.1 we show how Choral aids in transitioning sequential algorithms into concurrent implementations. We consider a Java algorithm and present the necessary steps to transform it into a Choral program that distributes its computation over three nodes—the number of nodes follows naturally from the recursive structure of the algorithm. The steps are straightforward, thanks also to the guidance offered by the Choral compiler. In Section 6.1.2 we consider a complete, three-tier system: RetwisJ[1], a clone of Twitter implemented by the Spring team as an example of integration with the Redis data store. RetwisJ comes as a monolithic application that consists of independent components, which handle interaction with the client, business logic, and interaction with the data store. Following this design, in the transition to Choral (called ChoRetwis), each component appears as a role in the choreography. ChoRetwis is a drop-in replacement for RetwisJ (e.g., we can use RetwisJ and ChoRetwis with the same clients and database). As an advantage of our choreographic refactoring, the architecture is more flexible wrt deployment: all components can be deployed on the same or different machines.
(2) In Section 6.2, we compare Choral to a popular alternative for concurrent and distributed programming: reactive actors. We use the Akka framework for Java as representative for reactive actors: since Choral is essentially an extension of Java, this choice helps in comparing our approach against more standard approaches to concurrent programming at the net of linguistic differences. In addition to the key qualitative advantage that Choral provides choreography compliance, we find that Choral contributes to keeping the codebase smaller. Furthermore, we find that the Java code generated by the Choral compiler is not significantly different in size from manually-written Java code. This opens the door to a quantitative evaluation, which we carry out in Section 6.3.
(3) In Section 6.3, we present a quantitative evaluation of how Choral impacts software development and execution performance. Section 6.3.1 reports some interesting measurements on the performance of the Choral compiler. Our analysis shows that using Choral leads to smaller codebases, and that the numbers of roles and conditionals in a choreography are significant determinants of how much our approach contributes to this aspect. We also observe that both our type checker and our compiler are fast for our example set—feedback is provided and compilation completed in a matter of milliseconds. Thus, our approach does not impact negatively the performance of development toolchains in a significant way. In Section 6.3.2, we look at the runtime performance of the code that we generate. Specifically, we compare the execution times of Choral and Akka implementations of the Karatsuba algorithm presented, respectively, in Section 6.1 and Section 6.2. We find the performance of the two models comparable (with better performance with the Choral alternative in the majority of the tested cases).

Overall, our results are encouraging. In addition to the advantage of choreography compliance, smaller codebases tend to host fewer bugs [Bessey et al. 2010], and Choral appears rather approachable when we consider the context of existing practices.

6.1 From Java to Choral

Thanks to the fact that Choral is based on mainstream abstractions, we can use the implementation of a sequential algorithm in Java as a starting point to obtain a concurrent variation of the algorithm (in Choral).

[1] https://github.com/spring-attic/spring-data-keyvalue-examples/tree/master/retwisj

```java
public static Long multiply(Long n1, Long n2) {
  if (n1 < 10 || n2 < 10) {
    return n1 * n2;
  } else {
    Double m = Math.max(Math.log10(n1), Math.log10(n2)) + 1;
    Integer m2 = Double.valueOf(m / 2).intValue();
    Integer splitter = Double.valueOf(Math.pow(10, m2)).intValue();
    Long h1 = n1 / splitter; Long l1 = n1 % splitter;
    Long h2 = n2 / splitter; Long l2 = n2 % splitter;
    Long z0 = Karatsuba.multiply(l1, l2);
    Long z2 = Karatsuba.multiply(h1, h2);
    Long z1 = Karatsuba.multiply(l1 + h1, l2 + h2) - z2 - z0;
    return z2 * splitter * splitter + z1 * splitter + z0;
  }
}
```

```java
public static Long@A multiply (Long@A n1, Long@A n2,
  SymChannel@(A, B)<Long> ch_AB,
  SymChannel@(B, C)<Long> ch_BC,
  SymChannel@(C, A)<Long> ch_CA) {
  if (n1 < 10@A || n2 < 10@A) {
    ch_AB.<Choice>select(Choice@A.DONE); ch_CA.<Choice>select(Choice@A.DONE);
    return n1 * n2;
  } else {
    ch_AB.<Choice>select(Choice@A.REC); ch_CA.<Choice>select(Choice@A.REC);
    Double@A m = Math@A.max(Math@A.log10(n1), Math@A.log10(n2)) + 1@A;
    Integer@A m2 = Double@A.valueOf(m / 2@A).intValue();
    Integer@A splitter = Double@A.valueOf(Math@A.pow(10@A, m2)).intValue();
    Long@A h1 = n1 / splitter; Long@A l1 = n1 % splitter;
    Long@A h2 = n2 / splitter; Long@A l2 = n2 % splitter;
    Long@A z0 = Karatsuba@(B, C, A)
      .multiply(ch_AB.<Long>com(l1), ch_AB.<Long>com(l2), ch_BC, ch_CA, ch_AB)
      >> ch_AB::<Long>com;
    Long@A z2 = Karatsuba@(C, A, B)
      .multiply(ch_CA.<Long>com(h1), ch_CA.<Long>com(h2), ch_CA, ch_AB, ch_BC)
      >> ch_CA::<Long>com;
    Long@A z1 = Karatsuba@(A, B, C)
      .multiply(l1 + h1, l2 + h2, ch_AB, ch_BC, ch_CA) - z2 - z0;
    return z2 * splitter * splitter + z1 * splitter + z0;
  }
}
```

Fig. 6. Karatsuba algorithm. Left: Java (sequential). Right: Choral (choreographic).

6.1.1 Transforming an algorithm: Karatsuba. Consider the algorithm for fast multiplication by Karatsuba and Ofman [1962]. An implementation of the algorithm in Java is displayed on the left side of Figure 6. Starting from the Java implementation, we can obtain a distributed implementation of the same algorithm in Choral by adding:

- Information on where the data is located.
- Data transmissions for moving the data and implementing knowledge of choice.

The resulting Choral program is displayed on the right side of Figure 6, where the additions are highlighted in yellow.

The Choral program has three roles (`A`, `B`, and `C`), which distribute among themselves the three sub-calculations of the algorithm. In the parameters and return type, we added information on data locality (e.g., `Long@A n1`) and the necessary channels (e.g., `ch_AB`) for moving data in the implementation of the method. Given the original Java code, the type checker of the Choral compiler would assist the programmer by pointing out that data locality information must be added. Likewise, in the implementation of the method, we added information on data locality for constant values and variables (e.g., `Double@A m`). Additionally, we added the necessary data transmissions: selections to implement knowledge of choice for the conditional, and communications of values whenever they should move from a role to another. Again, the Choral compiler aids the programmer by asking for all this information.

6.1.2 Transforming an entire system: RetwisJ (Redis-based Twitter clone). We now focus on the transformation of a complex, real-world system to provide a more comprehensive view of the process. The transformation allows us to illustrate how Choral can help developers transition from monolithic to distributed implementations (à la microservices) while maintaining their options open wrt code reuse, interoperability, and deployment configurations.

Concretely, we took RetwisJ,[2] which is a Java, Spring-based port of Retwis[3], and re-implemented its logic as a distributed application that consists of three separate modules. We report, at the top

[2]Source code available at https://github.com/spring-projects/spring-data-keyvalue-examples, documentation available at https://docs.spring.io/spring-data/data-keyvalue/examples/retwisj/current/.

[3]A Twitter clone originally proposed by the Redis team to illustrate the capabilities of the data store, described at https://redis.io/topics/twitter-clone.

of Section 6.1.2, the simplified class diagram of RetwisJ. The application is a monolith, where the central module "RetwisController" works both as the gateway for serving webpages to the user (the "JSP pages" module) and as the entry-point for user requests (e.g., to log in, to post tweets, etc.). The classes "User" and "Post" model the main entities in the system, while the "RetwisRepository" implements the logic for data persistency and retrieval.

Refactoring RetwisJ in Choral follows naturally well-known patterns from microservice architectures [Dragoni et al. 2017; Newman 2021]: interaction with the client is handled by a "Gateway" component; business logic is managed by a "Controller" component; and data storage and access is managed by a "Storage" component. These components (depicted in the lower part of Section 6.1.2) correspond to roles in our implementation, so we obtain a choreography that defines how these three roles collaborate in order to implement the application.

Each component is implemented by combining the respective code for it compiled from the choreography (for coordination) together with local code that implements the internal functionalities that are out of the scope of the choreography. For such internal functionalities, for example the concrete data read/write operations on Redis, we reuse existing code from RetwisJ. In Section 6.1.2, we display which classes from RetwisJ have been reused as-is (e.g., "User" and "Post"). All three components are loosely-coupled, in the sense that they interact purely via message passing (as instructed by the choreography). Since the choreography uses our abstract channel interfaces, our implementation is more flexible than the original RetwisJ: developers can choose to distribute the components by using, e.g., TCP/IP channels, or to deploy all of them as a single application using in-memory communication channels (which is the only option with RetwisJ).

The choreography is strategically parametric on a few notable aspects.

- The gateway receives API calls through a generic "CommandInterface". This allows us to expose the API over different media. In our concrete example, we implemented an "HTTP-CommandInterface" for exposing a typical REST API (designed by us) and an alternative that acts as drop-in replacement for RetwisJ by implementing the API expected by the JSP pages provided in that project.
- The controller delegates the storage and retrieval of session state to an abstract "Session-Manager". Our implementation stores state locally (in memory), but it can in principle be generalised to storing state on an external distributed store, to allow for replication.
- The storage component relies on an abstract "DatabaseConnection" (a database abstraction layer), which determines how data is concretely represented, read, and written. Our implementation reuses the Redis-based code from RetwisJ. Thus, RetwisJ and our implementation can even be used in parallel. Storage is, however, not limited to using Redis and alternative implementations of "DatabaseConnection" can be provided.

To test our implementation and its consistency with the original RetwisJ, we developed a test suite that programmatically invokes the HTTP APIs of the two systems. The suite performs a series of tests that simulates usage, modifying state (e.g., creating users and posts) and then checking that the results are as expected.

6.2 Programming Paradigms: Choral and Akka

We carry out a brief comparison between Choral and an established framework for concurrent programming: the Akka framework for the Java language. Akka is a popular reactive framework based on actors for the "traditional" way of programming concurrent software; that is, software where each endpoint is programmed from a local viewpoint, in contrast with the global view on the expected interactions of choreographic programming.

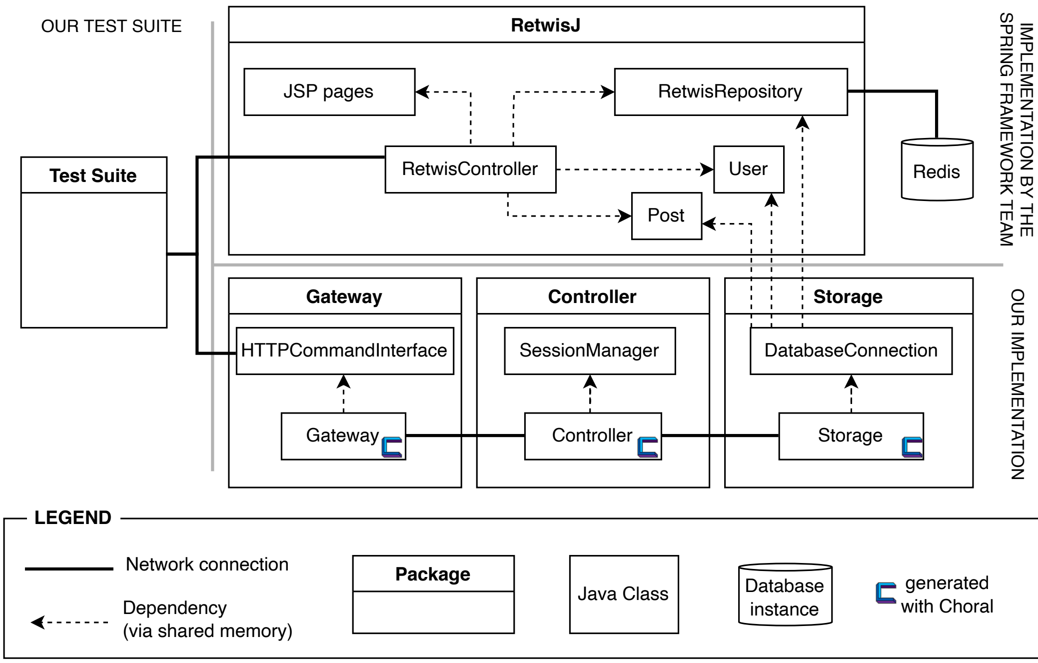

Fig. 7. Diagram of the RetwisJ and ChorRetwis Systems (classes, packages, and deployment).

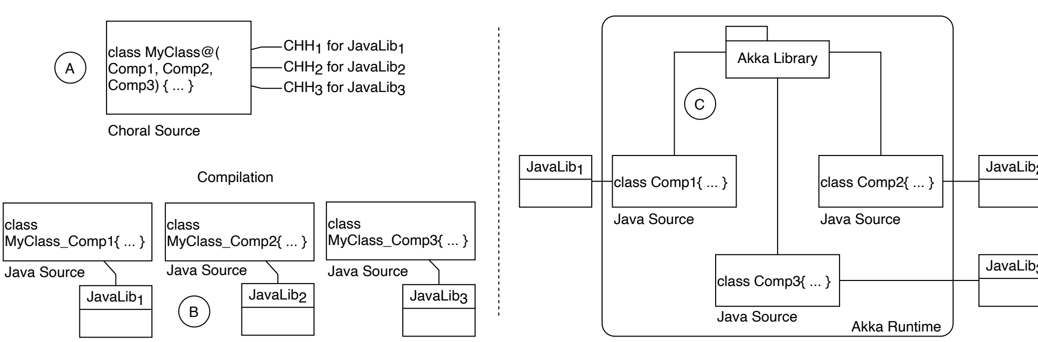

Fig. 8. Depiction of the programming approaches of Choral (left) and Akka (right).

We depict the development processes of Choral and Akka in Figure 8, respectively on the left and on the right sides. The processes are slightly different:

- In Choral, we implement a choreography using a single codebase (Ⓐ). The codebase for each participant is then generated automatically by our compiler (Ⓑ).
- By contrast, in Akka, we implement the behaviour of each participant separately (Ⓒ). There are no components to write choreographies.

Choral provides choreography compliance through its language and compiler. Akka provides no tools to express choreographies, nor to check for compliance—these aspects must be handled manually by the programmer.

Program	Choral (LOC)	Java/Choral-generated (LOC)	Java/Akka (LOC)
DistAuth	56	137	234
MergeSort	63	239	166
Karatsuba	31	92	118

Table 1. Comparison of three codebases implemented in Choral and Akka.

Both Choral and Akka require the programmer to adopt a few principles, respectively: for Choral, our notion of data types with multiple roles; for Akka, the design patterns and APIs expected by the Akka framework for user implementations, and the APIs of the Akka libraries. Notably, Choral does not fix any APIs: the choice of APIs and implementations of channels or other methods are completely up to the user. Thus, Choral leaves more freedom to the developer in choosing what libraries to rely on. Furthermore, Choral requires no runtime library during execution, while Akka requires programmers to adopt the Akka runtime (represented in Figure 8 by the Akka Runtime rectangle that surrounds the components).

Choral and Akka allow for reusing existing Java code and libraries, e.g., database drivers. When Java code involves a single role, using it in Choral is straightforward—we interpret any Java type as a type parametric at a single role. In general, programmers can explicitly coerce Java types to arbitrary Choral types by using special header files called Choral Headers, shortened as CHH ($\mathsf{CHH}_1, \cdots, \mathsf{CHH}_3$ in Figure 8), which isolate code that might be "unsafe" because manually written in Java.

To get more concrete observations and data, we manually implemented in Akka three choreographies presented in this paper, namely DistAuth (see Section 3.1), MergeSort (see Section 3.3), and Karatsuba (see the beginning in this section). In Table 1, we compare the sizes of the three codebases for each example in terms of lines of code (LOC). Specifically, we report: the size of the Choral implementation, the size of the Java code generated from the Choral implementation, and the size of the manually-written Akka implementation.

The comparison between the implementations of DistAuth is straightforward. The main difference lies in the fact that Akka follows a reactive programming style that dictates the usage of fields and messages of different types inside each actor to track the (local) status of the protocol. The flow of interactions becomes thus implicit, and needs to be reconstructed by the expected asynchronous activations of methods at actors—a similar observation has already been made by Weisenburger et al. [2018]. Differently, in Choral the flow of interactions is made explicit by our type system and sequencing of actions. This difference makes the Choral codebase quite shorter, since the fields and message types used to implement causal dependencies in Akka add boilerplate code. The structures of the components between the generated Choral code and the Akka implementation follow the same pattern, i.e., we have one class for each participant (the Client, the Service, and the IP from Section 3.1).

We then moved to implementing the MergeSort and Karatsuba examples, because their recursive nature makes them a good fit for reactive programming as in Akka. Indeed, differently from the DistAuth implementation that had a one-to-one correspondence between the Choral-generated and Akka-based classes, the Akka implementations of MergeSort and Karatsuba rely mainly on a single class that defines the implementation of all roles. We find that the actor paradigm and the programming style we used for Choral in this article promote different ways of dealing with recursion. For the Akka implementation, we followed the idiomatic approach of creating a new actor for each new recursive call of the distributed algorithm. By contrast, in the Choral implementations, it is natural to use the same participants in recursive calls by switching their roles as shown in Sections 3.3 and 6.1.1—this lowers complexity and performance costs wrt coordination, since we avoid creating and managing new participants. Adopting the "one class" style for Akka helps keep the codebase small (the Java implementation generated from Choral is bigger for MergeSort), but makes the implementation tightly coupled.

Interestingly, for the Karatsuba algorithm, the Choral-generated Java implementation is smaller than the manually-written Akka implementation. This is because Karatsuba requires more coordination and local tracking of the distributed state, correspondent to boilerplate code that defines bookkeeping message types and fields in Akka.

What went wrong. There are valuable lessons that we learned from the exercise of writing implementations of the Karatsuba algorithm in Akka and Choral. In particular, writing the Akka implementation was trickier and more error-prone. We illustrate two representative issues, which arise from the lack of a choreographic view in Akka. These issues have been analysed by two of the authors working together to peer-review both the development process and the resulting code.

- For the first Akka implementation, we noticed that its performance was always the same for all tested inputs. It took us some time to notice that no communications were performed at all: all multiplications were resolved locally because there was a typo in the direction of the inequality check that decides whether to perform the `z0`-`z1`-`z2` decomposition of the product (cf. Figure 6) or perform it directly. This was a very subtle bug, because the program was terminating successfully and returned the right results.
  This issue was caused by the fact that the code for the Karatsuba algorithm needs to be distributed across different methods in Akka, which obfuscates the original structure of the algorithm and adds opportunities for banal bugs (e.g., swapping two arguments by mistake). We did not encounter this kind of issues with the Choral implementation, because we just needed to augment the original (correct) Java implementation with channel usage and roles (cf. Figure 6). The distribution of code in the final Java implementation has been carried out by the Choral compiler without mistakes.
- Another bug that we encountered was due to a typo, too, which made the Akka implementation perform the calculation of `z2` instead of the one of `z1`. However, more interestingly, this error manifested itself as a deadlock: an actor was waiting for the arrival of the three sub-calculations for `z0`, `z1`, `z2`, and instead received only those for `z0` and `z2` (the latter twice). Fixing this kind of bugs is tricky in parallel programming, and indeed our first attempt at a fix introduced another bug: we fixed the deadlock, but we started obtaining wrong results because the fix caused a swap of the inputs for `z1` and `z2`.
  The key reason behind these bugs was that, in Akka, we had to manually write code that reads and writes tags in messages to know if they contain the result of computing `z0`, `z1`, or `z2`. In this code, errors can be written in the code that creates messages at the sender and the code that processes them at the receiver. Implementations can therefore fall out of sync. Choreographic programming (hence Choral) prevents this kind of bugs entirely, because it is not possible to write mismatched communications at the choreographic level.

6.3 Microbenchmarks

We now move to a more systematic and quantitative evaluation of how Choral impacts software development—in addition to the key benefit of choreography compliance. First, we evaluate the performance of the Choral compiler with microbenchmarks on 11 Choral programs. Then, since we implemented the same Karatsuba algorithm both in Java, Choral, and Akka, we provide some preliminary runtime benchmarks by contrasting their performance.

Program	Choral (LOC)	# Roles	# Conditionals	Java (LOC)	Size Increase (%)	Type Checking (ms)	Proj. Checking (ms)	Projection (ms)
HelloRoles	9	2	0	14	55%	5.915	0.334	0.187
ConsumeItems	16	2	1	49	206%	9.572	0.861	0.607
BuyerSellerShipper	40	3	2	126	215%	8.204	1.274	1.015
DistAuth	56	3	1	137	144%	11.463	9.097	0.986
VitalsStreaming	47	2	1	78	65%	7.864	1.384	0.417
DiffieHellman	26	2	0	36	38%	5.911	0.232	0.152
MergeSort	63	3	4	239	279%	8.517	7.891	3.723
QuickSort	74	3	3	200	170%	7.213	6.204	2.806
Karatsuba	31	3	1	92	196%	6.491	2.566	1.078
DistAuth5	66	5	1	226	242%	10.581	5.573	1.036
DistAuth10	91	10	1	438	381%	10.576	5.643	3.011

Table 2. Performance results for the Choral compiler.

6.3.1 Compilation Benchmarks. Regarding the performance of the Choral compiler, we report our results in Table 2. There, for each program, we report (left to right): the name of the Choral program, lines of code, number of roles, number of conditionals (**`if`** and **`switch`** blocks), lines of code of the compiled Java code (total for all roles), number of milliseconds to perform type checking, the number of milliseconds to perform the check for projectability, and the number of milliseconds to perform the projection. (Section 4.2). All code is well indented and the numbers of lines just omits empty ones. We collected times on a machine equipped with an Intel Core i5-3570K 3.4 GHz CPU and 12 GB of RAM, running macOS 10.15 and Java 17. The reported times are averages of 1000 runs each, after a warm-up of 1000 prior runs.

Table 2 reports data for programs shown in this article, plus four other programs: BuyerSeller-Shipper is inspired by a recurring e-commerce example found in choreography articles [Carbone et al. 2012; Honda et al. 2016]; the Diffie–Hellman protocol for cryptographic key exchange [Diffie and Hellman 1976]; and DistAuth5 and DistAuth10, which are variants of the DistAuth class from Section 3.1, where we respectively add 3 and 7 roles, 2 and 7 channels, and 4 and 14 selections for coordination.

Our preliminary data from Section 6.2 points out that Choral programs are significantly smaller than the Java implementations compiled from them. This is good in itself: recall that smaller codebases typically host fewer bugs [Bessey et al. 2010]. In our microbenchmarks, compilation leads to an average increase of 181% in codebase size (going from the 38% for `DiffieHellman` up to 381% for `DistAuth10`): the difference between the sizes of the original Choral program and the generated Java code is a rough approximation of code that the programmer has been spared from writing manually. Furthermore, our microbenchmarks suggest that there are two main parameters that affect this benefit:

- The number of roles involved in the source choreography. This is explained by the fact that each statement in Choral involving n roles corresponds to n statements in the generated Java code—one for each role, implementing what the role has to do to follow the choreography. For example, a Choral statement invoking a method to communicate data from `A` and `B` would produce a statement in the code for `A` (for sending) and a statement in the code for `B` (for receiving). There are examples of choreographies with many instructions that involve a single role, like `DiffieHellman`, which results in a smaller expansion.

- The number of conditionals. Conditionals usually require performing selections to handle knowledge of choice (Section 2.3). Then, the (code compiled for the) roles receiving selections has to inspect the type of the received message using a **`switch`** statement, which is automatically added by our compiler and is not present in the original Choral code. For example, MergeSort and QuickSort differ in that the former has 4 conditionals whereas the latter has 3 conditionals, and respectively reach an expansion of 279% and 170%.

As a final remark, we observe that type and projectability checking and projection do not add any significant delay to the development experience: they respectively average ca. 8.391ms, 3.732ms, and 1.365ms. This matches our own subjective programming experience with Choral, where the compiler managed to feel quite responsive in providing quick feedback while coding. Both operations are mostly influenced by the number of conditionals and roles, matching our previous observations.

6.3.2 Runtime Benchmarks. We conclude this section by comparing the execution times of the Choral and Akka implementations of Karatsuba. We show the results in Figure 9. In the figure, we report in each of the six quadrants the average execution time, in nanoseconds, of a sequence of 1000 multiplications. Each quadrant regards a specific "tier" of multiplication, i.e., the 10^9 tier corresponds to the multiplication of two factors of the shape, e.g., $i * 10^5$ and $j * 10^4$, which produce a result of that tier's magnitude. All benchmarked implementations use the same inputs: we generated and used six files, each containing 1000 pairs of random factors. Each file corresponds to a tier. The considered tiers are: 10^9, 10^{11}, 10^{13}, 10^{15}, 10^{17}, and 10^{19}—the latter nears the maximal values managed by the Java `long` data type, but we make sure to never produce overflows. We performed the benchmarks on the same machine used to benchmark the Choral compiler above, with Java 17 and Akka 2.6.9. For each implementation, we run the benchmark two times in sequence, discarding the data of the first run to warm up the JVM and provide more stable results.

To benchmark both Akka and Choral implementations, we used in-memory communications. We wrote an implementation of in-memory channels for Choral, while for Akka we used the default in-memory channels provided by the framework. In Figure 9 we report the execution times of the Choral and Akka implementations both considering their setup (respectively, "ASetup" and "CSetup") and without (respectively, "Akka" and "Choral"). The reason to report setup times is to provide the reader with an indication of the wall-clock times taken by the alternatives. The setup of Akka regards the creation of the `ActorSystem` to execute the Karatsuba behaviour and its closure after having obtained the result. The setup for Choral includes the creation of an `Executor` pool of three threads, the creation of the three in-memory channels, and the closure of the pool after having obtained the result.

The Choral implementation outperforms Akka's in all quadrants, both with and without the overhead from the setup—indeed, although with a slight margin, the performance of the "Choral" implementation *with* the setup times outperforms "Akka" without the corresponding overhead. We also note that, contrarily to Choral, the wall-clock execution times of Akka are largely dominated by its setup times. This phenomenon is testified by the "ASetup" bar in Figure 9, which has almost the same performance irrespective of the magnitude of the computed result, while the alternatives show longer times for greater magnitudes. We attribute this phenomenon to the different ways in which the runtimes of the Choral and Akka in-memory implementations manage concurrency and messaging. Indeed, skimming through the internal code of Akka, we found that the framework puts in place threading models and advanced messaging systems optimised for the execution of many actors that communicate in parallel, which can take a high performance toll when implementing

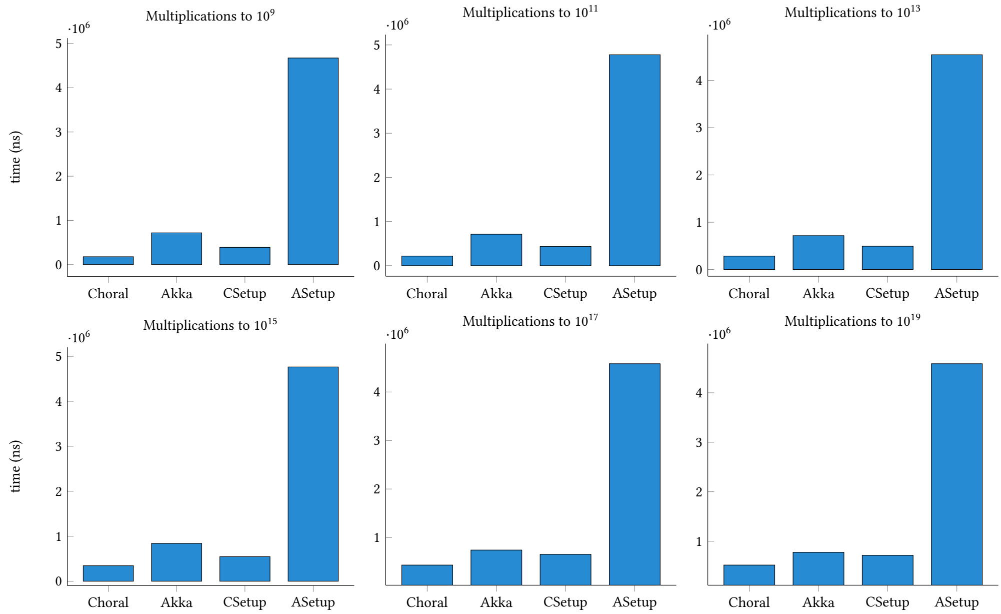

Fig. 9. Benchmarks of Choral (Choral, CSetup), and Akka (Akka, ASetup) implementations of the Karatsuba algorithm.

"lighter" computations, like the one benchmarked here. In the future, we plan to investigate these aspects more in depth, identifying a set of benchmarks sensible to the peculiarities of the chosen Choral and Akka runtimes and able to help shedding light on the trade-offs of either approaches.

7 RELATED WORK, DISCUSSION, AND FUTURE WORK

Choral is a *choreographic programming language*: it makes the flow of interactions and their related computations manifest from a global viewpoint [Montesi 2013]. While Choral suffices already in tackling different kinds of use cases, as we have discussed in this article, the literature on choreographies is vast: it includes features that are generalised by Choral, other features that we have not considered in this work, and open problems. We discuss related work and potential future developments of Choral in the rest of this section.

Previous Implementations of Choreographic Programming. The idea of synthesising local participant specifications that comply with choreographies has been a hot research topic for more than 20 years, and work in this line of research is typically based on automata or process calculi abstractions [Alur et al. 2000; Autili et al. 2018; Basu et al. 2012; Honda et al. 2016; Qiu et al. 2007]. Previous implementations of choreographic programming consist of Chor [Carbone and Montesi 2013] and AIOCJ [Dalla Preda et al. 2017], which are based on process calculi and generate executable Jolie code. Compared to them, Choral solves the modularity problems mentioned in the Introduction, by revisiting choreographies under the light of mainstream abstractions. Another advantage is that the types of channels needed by a choreography are made explicit and can be user-defined [Carbone et al. 2012; Carbone and Montesi 2013; Honda et al. 2016; Qiu et al. 2007].

Other Approaches to Spatially-Distributed Programming. The types that support our choreography-as-objects interpretation have been inspired by ideas found in modal logics for mobile ambients [Cardelli and Gordon 2000] and, later, in the line of work on multitier programming [Cooper et al.

2006; Liu et al. 2009; Murphy VII et al. 2007, 2004; Neubauer and Thiemann 2005; Serrano et al. 2006; Weisenburger et al. 2018]. In the words of Murphy VII et al. [2004], these works represent other approaches to "spatially-distributed computation". For example, in the most recent incarnation of multitier programming (ScalaLoci, by Weisenburger et al. [2018]), a distributed application is essentially defined as a single program that composes different functions, each localised at a single participant. A function can then invoke special primitives to request remote computation by another participant, whose implementation must always be ready for such requests [Weisenburger et al. 2020]. Differently from choreographies, this makes the flow of communications implicit and dependent on an underlying middleware—indeed, multitier programming was not designed to address the problems of defining choreographies and addressing choreography compliance as an aim. Choral generalises data types localised at a single participant to data types localised at many participants (roles), which enables our novel development process for choreography-compliant libraries. Castro-Perez and Yoshida [2020] explored the parallelisation of a simple multitier first-order functional language, for which they can infer abstract (computation is not included) choreographies of the communication flows that these programs can enact; Choral could be a candidate implementation language for this kind of models.

Choral's clear relation to the ideas found in logics for mobile ambients has already proven useful. In particular, in [Giallorenzo et al. 2021], Choral is used to kickstart an investigation of the links and differences between choreographic and multitier programming, by taking Choral and ScalaLoci as representative languages. After identifying the core abstractions that differentiate the two approaches, the authors provide algorithms for translating Choral code into ScalaLoci code and vice versa. Going from a multitier program to a choreographic one requires synthesising one of the many possible protocols (a choreography) that implements the necessary communications to execute the multitier program (which does not specify this aspect). This connection paves the way for joint research and cross-fertilisation between the two communities [Giallorenzo et al. 2021].

Higher-Order Choreographies. Interpreting choreographies as objects enables, for the first time, higher-order composition of choreographies that carry state (the fields of the objects): stateful choreographies (objects) can be passed as arguments. Stateful choreographies have been investigated before without higher-order composition—see, for example, the works [Carbone et al. 2012; Chen and Honda 2012; Cruz-Filipe and Montesi 2020]. Demangeon and Honda [2012] studied how parameters that abstract choreographies can be expanded syntactically when state is not considered, and Dalla Preda et al. [2017] explored a framework to replace choreographies with other choreographies at runtime to achieve adaptability: in neither system the choreographies can carry state, and both systems require a role to act as orchestrator to "enter" into a choreography (whereas in Choral, control is fully distributed). In the setting of multitier programming, Weisenburger and Salvaneschi [2019] introduced a module system to write multitier programs as compositions of submodules. Differently from Choral and the work by Demangeon and Honda [2012], their approach requires to fix roles statically, whereas in our case roles can be freely instantiated at runtime—for example, our merge sort example in Section 3.3 exploits this feature when roles are exchanged in recursive calls. Our new data types might thus be interesting also in the setting of multitier programming.

Selection Inference. Choral requires the programmer to insert the necessary selections to achieve knowledge of choice (Section 2.3). Developing techniques for inferring these selections automatically is an ongoing research topic. Typically, these techniques either modify the source choreography to include extra selections or inject hidden communications in the generated endpoint code [Basu and Bultan 2016; Cruz-Filipe and Montesi 2020; Dalla Preda et al. 2017; Jongmans et al. 2015; Lanese et al. 2013]. However, there is no silver bullet:

(1) In general, it is unfeasible to detect automatically what the optimal selection strategy is. This is a problem for both approaches (modifying the source choreography or injecting hidden communications in the generated code). Say that A needs to inform B and C of a choice by using point-to-point channels. Should A send the first selection to B or to C? That might depend on whether B has a longer task to perform in reaction to the selection compared to C, or vice versa. (Whichever has the longest task to start should get the selection first, to increase parallelism.) And what if multiple channels are available? For example, if A shares a fast channel with B but not with C, and B shares a fast channel with C, then it might be good that A informs B and subsequently B informs C (instead of A informing C directly). These issues become even more sophisticated when considering choreographies with more complex network topologies, scatter/gather channels, recursion, etc.
(2) If a compiler injects hidden communications in the generated code, then the source choreography program does not faithfully represent the communications enacted by the system any more. This makes the choreography less useful when reasoning about, for example, efficiency—network communications like selections are especially a huge performance factor—and security—hidden extra communications might leak information in ways not intended by the designer of the original protocol.

Since both issues are still the object of active investigation, we decided that a first version of Choral should be a base that future work can use to work on them. A promising compromise could be a hybrid assisted way. That is, the programmer should be able to write a choreography including some selections deemed important, but also potentially missing some necessary other selections; then, a tool should detect the missing selections and propose a solution. The programmer could thus decide whether to accept the proposal or improve it manually to achieve their requirements.

In general, we believe that there is a lot of potential in future research on how to optimise communications in Choral. New algorithms might leverage annotations of channels, static analysis, and profiling data. Some algorithms might choose simpler approaches at the expense of parallelism, whereas others might take a more decentralised approach to spreading knowledge of choice to favour parallelism or energy saving (in the Internet of Things, spreading battery consumption evenly or in a focused way might be an advantage depending on the scenario).

Deadlock-Freedom. A typical property in theory of choreographic programming is deadlock-freedom-by-design: the code compiled from a choreography is deadlock-free, because it is not possible to write mismatched communications in choreographies [Carbone and Montesi 2013; Dalla Preda et al. 2017; Hirsch and Garg 2022; Jongmans and van den Bos 2022; Montesi 2022]. The proof of this result relies on strong assumptions: foreign code (in our case, Java code) used within a choreography always terminates, and communication is reliable (messages are never lost, duplicated, or corrupted). Under the same assumptions, Choral provides the same guarantee. However, in the real world, these assumptions usually do not hold and we have to fall back to best-effort strategies. In particular, we have to adopt timeout and supervision mechanisms to avoid divergence and deal with failures, as we have exemplified in Sections 2.5 and 3.2.

Formalising Choral's semantics and type system to investigate their formal properties under different assumptions is interesting future work. Work in this direction [Cruz-Filipe et al. 2021; Hirsch and Garg 2022] started appearing after Choral's first technical report and release [Choral Development Team 2020; Giallorenzo et al. 2020]. In particular, the reader interested in the essential principles of higher-order choreographic composition implemented in Choral can consult [Cruz-Filipe et al. 2021].

Expressivity. We discuss a few interesting directions for future work regarding the expressivity of Choral and of its type system. In general, we believe that our choreographies-as-objects interpretation allows for importing established techniques from type theory to reason statically about roles in useful ways.

Choral can capture a variety of interaction patterns, like scatter-gather and producers-consumers with races. Nevertheless, there are cases where our types can be coarse. For instance, the following could be an interface of a channel where two receivers, `B` and `C`, race to consume a message.

```
interface RaceDiChannel(A,B,C)<T@X> {
 <S@Y extends T@Y> BiPair@(B,C)< Optional<S>, Optional<S> > com(S@A m);
}
```

The return type of method `com` above guarantees that both receivers will have a value of type `Optional<S>` located at them. However, depending on the behaviour that the programmer wishes to model, the interface above could be an over-approximation. For example, implementations that simply discard the message sent from `A` or that deliver the message to both `B` and `C` will satisfy the return type. Currently, we cannot express the type of a channel that forbids such implementations, e.g., a channel guaranteeing that exactly one between `B` and `C` will obtain the message. This means that, in such cases, we have to "pollute" the continuation of the choreography with local checks at both potential receivers.

A way to achieve a more specific type for races could be to extend Choral types with existential quantification over role parameters. For example, we could write

```
S@D with D in [B,C]
```

to express an instance of `S` located at some role `D` in the list `[B,C]` (i.e., `S@D` can be either `S@B` or `S@C`). With this type, we can write a more specific signature: method `com` returns a value of type `S@B` or of type `S@C`, but we cannot statically know which of the two types.

```
interface RaceDiChannel(A,B,C)<T@X> { <S@Y extends T@Y> S@D with D in [B,C] com(S@A m); }
```

Although there is some work on the use of existential quantification in simple choreography languages [Jongmans and Yoshida 2020], its application and integration with a general-purpose language like Choral poses some challenges and design choices. For instance, should roles that lose the race in method `com` be blocked? If so, is this specific of this method or the standard interpretation of every method with an existential return type? These and similar questions beg for a thorough investigation and go beyond Choral. In fact, a satisfactory and general handling of races in choreographic languages is still missing.

Another limitation of the current type system is that the number of role parameters of a choreography is fixed. This limitation is common to many choreographic languages. Deniélou and Yoshida [2011] developed a theory for parameterising choreographies over "collections of roles", whose sizes are determined at runtime. All roles in the same collection must be treated uniformly (e.g., broadcast). We can import that feature to Choral by allowing for role parameters to be collections. For example, we could prefix a role parameter declaration with `*`, as in `*Ds`, to specify that it is a collection of roles. Then, we can write the type of a channel for broadcasting data from `A` to all roles in the collection `Ds` as follows.

```
interface BroadcastDiChannel(A, *Ds)<T@X> { /* ... */ }
```

The method `com` for this channel should take a value of type `S@A` and return a value of type `S@D` for every role `D` in the collection `Ds`. This would require investigating how Choral can be extended with types for distributed data collections, as well.

Error Handling. Choral supports exception handling at a single role, which can then propagate errors to others via knowledge of choice. However, in our experience, it is more convenient to represent failures in return types, like we did in Section 3.1 by using `Optional`. The channel APIs that we showed in this paper are implemented by performing automatic retries. These APIs also have equivalent versions that wrap results in `Result` objects—essentially sum types of the transmitted value type and an error type, as in Go and Rust. Choosing among these implementations is up to the choreography programmer, and programmers might also devise channel implementations with their own strategies (e.g., exponential backoff with bound on the number of retries). Our compiler can, in principle, be extended to synthesise coordination for distributed exceptions, theorised by Carbone et al. [2008].

Asynchronous Programming. The choreographies that we presented here use channel APIs as if they were blocking. This does not mean that an endpoint must dedicate a thread for participating in a choreography: future versions of Java will include fibers and the asynchronous execution of blocking APIs (reactor pattern) [OpenJDK 2020]. Choral is compatible with this direction. Should programmers want to program a choreography explicitly for asynchronous execution by using continuation-passing style, our channel APIs can be extended to take choreographic continuations as parameters.

Fluid APIs from Choreographic Specifications. As we mentioned and discussed in Section 1, previous work explored the automatic generation of fluid APIs that enforce following choreographies [Scalas et al. 2017]. Their choreographic language cannot include computation, so it cannot express any of our use cases. Furthermore, their approach does not support modularity and API control, as we discussed more in detail in the Introduction. Thus, Choral brings two improvements. First, our APIs are more reusable: they change only if the source API is changed, not if the communication behaviour of a method is simply updated. Second, the APIs of our compiled Java code are more idiomatic: they are plain object APIs that look like the typical task-oriented APIs distributed by cloud vendors [Murty 2008; Wilder 2012], which makes Choral a candidate drop-in replacement for current development practices.

Choreography-based Verification. Choreographic languages that are less expressive than Choral (e.g., they cannot include computation) have been used also to verify that interactions among objects respect a protocol. This is obtained by statically checking method invocations, either by using typestates [Kouzapas et al. 2018] or model checking [Scalas et al. 2019]. As noted by Hirsch and Garg [2022], choreographic programming offers a simpler development method: verification approaches require the programmer to design both a choreographic specification and then manually take care of writing a correct implementation of the projection of each role, whereas choreographic programming (and hence Choral) generates the latter automatically. Additionally, since our approach is based on compilation instead of verification, we can provide a more expressive choreographic language.

8 CONCLUSION

With the increased adoption of cloud computing, edge computing, the Internet of Things, and microservices, the need for libraries that implementors can use to participate correctly in choreographies is growing steadily [Atzori et al. 2010; Dragoni et al. 2017; Murty 2008; Wilder 2012]. Building on previous results on the theory of choreographies, choreographic programming came with the promise of aiding in the implementation of choreography-compliant concurrent and distributed software [Montesi 2015]. While the approach has been successfully applied in principle to different scenarios [Cruz-Filipe and Montesi 2016; Dalla Preda et al. 2017; Lluch-Lafuente et al. 2015; López and Heussen 2017; López et al. 2016], the link between choreographic programming and mainstream programming has remained unexplored until now (all implementations rely on the Jolie programming language, which is based on the theory of CCS [Milner 1980; Montesi et al. 2014]). Among the most important consequences, no implementation of the paradigm so far properly supported modularity—generating reusable libraries and controlling their APIs.

In this article, we have taken a fundamental step in the pursuit of the choreographic programming agenda. We have also shown that choreographies can be modelled by extending mainstream abstractions (in our case, objects) and that this leads to a choreographic programming language that supports modularity and can integrate with existing Java code. We have shown that Choral is sufficiently expressive to capture use cases of different kinds, discussed the design of our compiler, and performed a first evaluation which points out that the approach is promising. Choral is thus a step towards equipping programmers with a tool that safely ferries them from the design of choreographies to compliant implementations at the press of the proverbial button.

In the future, Choral could also be a useful vector for the application of research on choreographies based on other paradigms (automata, processes, etc.): researchers could develop translations of their own choreography models to Choral, and then leverage our compiler to obtain library implementations that can be used in mainstream software (in Java). Hopefully, this will lead to more implementations of choreography theories, allowing for their evaluation [Ancona et al. 2016].

ACKNOWLEDGMENTS

This work was partially supported by Villum Fonden, grant no. 29518.

REFERENCES

Rajeev Alur, Kousha Etessami, and Mihalis Yannakakis. 2000. Inference of message sequence charts. In *Proceedings of the 22nd International Conference on on Software Engineering, ICSE 2000, Limerick Ireland, June 4-11, 2000*, Carlo Ghezzi, Mehdi Jazayeri, and Alexander L. Wolf (Eds.). ACM, 304–313. https://doi.org/10.1145/337180.337215

Davide Ancona, Viviana Bono, Mario Bravetti, Joana Campos, Giuseppe Castagna, Pierre-Malo Deniélou, Simon J. Gay, Nils Gesbert, Elena Giachino, Raymond Hu, Einar Broch Johnsen, Francisco Martins, Viviana Mascardi, Fabrizio Montesi, Rumyana Neykova, Nicholas Ng, Luca Padovani, Vasco T. Vasconcelos, and Nobuko Yoshida. 2016. Behavioral Types in Programming Languages. *Foundations and Trends in Programming Languages* 3, 2-3 (2016), 95–230.

Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010. The internet of things: A survey. *Computer networks* 54, 15 (2010), 2787–2805.

Marco Autili, Paola Inverardi, and Massimo Tivoli. 2018. Choreography Realizability Enforcement through the Automatic Synthesis of Distributed Coordination Delegates. *Sci. Comput. Program.* 160 (2018), 3–29. https://doi.org/10.1016/j.scico.2017.10.010

Samik Basu and Tevfik Bultan. 2016. Automated Choreography Repair. In *Fundamental Approaches to Software Engineering - 19th International Conference, FASE 2016, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2016, Eindhoven, The Netherlands, April 2-8, 2016, Proceedings (Lecture Notes in Computer Science, Vol. 9633)*, Perdita Stevens and Andrzej Wasowski (Eds.). Springer, 13–30. https://doi.org/10.1007/978-3-662-49665-7_2

Samik Basu, Tevfik Bultan, and Meriem Ouederni. 2012. Deciding choreography realizability. In *Proceedings of the 39th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2012, Philadelphia, Pennsylvania, USA, January 22-28, 2012*, John Field and Michael Hicks (Eds.). ACM, 191–202. https://doi.org/10.1145/2103656.2103680

Al Bessey, Ken Block, Ben Chelf, Andy Chou, Bryan Fulton, Seth Hallem, Charles Henri-Gros, Asya Kamsky, Scott McPeak, and Dawson Engler. 2010. A few billion lines of code later: using static analysis to find bugs in the real world. *Commun. ACM* 53, 2 (2010), 66–75.

Mario Bravetti and Gianluigi Zavattaro. 2007. Towards a Unifying Theory for Choreography Conformance and Contract Compliance. In *Software Composition, 6th International Symposium, SC 2007, Braga, Portugal, March 24-25, 2007, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 4829)*, Markus Lumpe and Wim Vanderperren (Eds.). Springer, 34–50. https://doi.org/10.1007/978-3-540-77351-1_4

Marco Carbone. 2009. Session-based Choreography with Exceptions. *Electron. Notes Theor. Comput. Sci.* 241 (2009), 35–55. https://doi.org/10.1016/j.entcs.2009.06.003

Marco Carbone, Kohei Honda, and Nobuko Yoshida. 2008. Structured Interactional Exceptions in Session Types. In *CONCUR 2008 - Concurrency Theory, 19th International Conference, CONCUR 2008, Toronto, Canada, August 19-22, 2008. Proceedings (Lecture Notes in Computer Science, Vol. 5201)*, Franck van Breugel and Marsha Chechik (Eds.). Springer, 402–417. https://doi.org/10.1007/978-3-540-85361-9_32

Marco Carbone, Kohei Honda, and Nobuko Yoshida. 2012. Structured Communication-Centered Programming for Web Services. *ACM Trans. Program. Lang. Syst.* 34, 2 (2012), 8:1–8:78. https://doi.org/10.1145/2220365.2220367

Marco Carbone and Fabrizio Montesi. 2013. Deadlock-freedom-by-design: multiparty asynchronous global programming. In *The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13, Rome, Italy - January 23 - 25, 2013*, Roberto Giacobazzi and Radhia Cousot (Eds.). ACM, 263–274. https://doi.org/10.1145/2429069.2429101

Luca Cardelli and Andrew D. Gordon. 2000. Anytime, Anywhere: Modal Logics for Mobile Ambients. In *POPL 2000, Proceedings of the 27th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, Boston, Massachusetts, USA, January 19-21, 2000*, Mark N. Wegman and Thomas W. Reps (Eds.). ACM, 365–377. https://doi.org/10.1145/325694.325742

Giuseppe Castagna, Mariangiola Dezani-Ciancaglini, and Luca Padovani. 2011. On global types and multi-party sessions. In *Formal Techniques for Distributed Systems*. Springer, 1–28.

David Castro-Perez and Nobuko Yoshida. 2020. Compiling first-order functions to session-typed parallel code. In *CC '20: 29th International Conference on Compiler Construction, San Diego, CA, USA, February 22-23, 2020*, Louis-Noël Pouchet and Alexandra Jimborean (Eds.). ACM, 143–154. https://doi.org/10.1145/3377555.3377889

Tzu-Chun Chen and Kohei Honda. 2012. Specifying Stateful Asynchronous Properties for Distributed Programs. In *CONCUR 2012 - Concurrency Theory - 23rd International Conference, CONCUR 2012, Newcastle upon Tyne, UK, September 4-7, 2012. Proceedings (Lecture Notes in Computer Science, Vol. 7454)*, Maciej Koutny and Irek Ulidowski (Eds.). Springer, 209–224. https://doi.org/10.1007/978-3-642-32940-1_16

Choral Development Team. 2020. Choral Language Website. https://www.choral-lang.org.

Ezra Cooper, Sam Lindley, Philip Wadler, and Jeremy Yallop. 2006. Links: Web Programming Without Tiers. In *Formal Methods for Components and Objects, 5th International Symposium, FMCO 2006, Amsterdam, The Netherlands, November 7-10, 2006, Revised Lectures (Lecture Notes in Computer Science, Vol. 4709)*, Frank S. de Boer, Marcello M. Bonsangue, Susanne Graf, and Willem P. de Roever (Eds.). Springer, 266–296. https://doi.org/10.1007/978-3-540-74792-5_12

Luís Cruz-Filipe, Eva Graversen, Lovro Lugovic, Fabrizio Montesi, and Marco Peressotti. 2021. Choreographies as Functions. *CoRR* abs/2111.03701 (2021). arXiv:2111.03701 https://arxiv.org/abs/2111.03701

Luís Cruz-Filipe and Fabrizio Montesi. 2016. Choreographies in Practice. In *Formal Techniques for Distributed Objects, Components, and Systems - 36th IFIP WG 6.1 International Conference, FORTE 2016, Held as Part of the 11th International Federated Conference on Distributed Computing Techniques, DisCoTec 2016, Heraklion, Crete, Greece, June 6-9, 2016, Proceedings (Lecture Notes in Computer Science, Vol. 9688)*, Elvira Albert and Ivan Lanese (Eds.). Springer, 114–123. https://doi.org/10.1007/978-3-319-39570-8_8

Luís Cruz-Filipe and Fabrizio Montesi. 2020. A core model for choreographic programming. *Theor. Comput. Sci.* 802 (2020), 38–66. https://doi.org/10.1016/j.tcs.2019.07.005

Mila Dalla Preda, Maurizio Gabbrielli, Saverio Giallorenzo, Ivan Lanese, and Jacopo Mauro. 2017. Dynamic Choreographies: Theory And Implementation. *Logical Methods in Computer Science* 13, 2 (2017).

Romain Demangeon and Kohei Honda. 2012. Nested Protocols in Session Types. In *CONCUR 2012 - Concurrency Theory - 23rd International Conference, CONCUR 2012, Newcastle upon Tyne, UK, September 4-7, 2012. Proceedings (Lecture Notes in Computer Science, Vol. 7454)*, Maciej Koutny and Irek Ulidowski (Eds.). Springer, 272–286. https://doi.org/10.1007/978-3-642-32940-1_20

Pierre-Malo Deniélou and Nobuko Yoshida. 2011. Dynamic multirole session types. In *Proceedings of the 38th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2011, Austin, TX, USA, January 26-28, 2011*, Thomas Ball and Mooly Sagiv (Eds.). ACM, 435–446. https://doi.org/10.1145/1926385.1926435

Whitfield Diffie and Martin E. Hellman. 1976. New directions in cryptography. *IEEE Trans. Inf. Theory* 22, 6 (1976), 644–654. https://doi.org/10.1109/TIT.1976.1055638

Nicola Dragoni, Saverio Giallorenzo, Alberto Lluch-Lafuente, Manuel Mazzara, Fabrizio Montesi, Ruslan Mustafin, and Larisa Safina. 2017. Microservices: Yesterday, Today, and Tomorrow. In *Present and Ulterior Software Engineering*, Manuel Mazzara and Bertrand Meyer (Eds.). Springer, 195–216. https://doi.org/10.1007/978-3-319-67425-4_12

Simon Fowler, Sam Lindley, J. Garrett Morris, and Sára Decova. 2019. Exceptional asynchronous session types: session types without tiers. *Proc. ACM Program. Lang.* 3, POPL (2019), 28:1–28:29. https://doi.org/10.1145/3290341

Erich Gamma and Kent Beck. 2006. JUnit.

Saverio Giallorenzo, Fabrizio Montesi, and Marco Peressotti. 2020. Choreographies as Objects. *CoRR* abs/2005.09520 (2020). arXiv:2005.09520 https://arxiv.org/abs/2005.09520

Saverio Giallorenzo, Fabrizio Montesi, Marco Peressotti, David Richter, Guido Salvaneschi, and Pascal Weisenburger. 2021. Multiparty Languages: The Choreographic and Multitier Cases. In *35th European Conference on Object-Oriented Programming (ECOOP 2021) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 194)*, Anders Møller and Manu Sridharan (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 22:1–22:27. https://doi.org/10.4230/LIPIcs.ECOOP.2021.22

Saverio Giallorenzo, Fabrizio Montesi, Larisa Safina, and Stefano Pio Zingaro. 2019. Ephemeral Data Handling in Microservices. In *2019 IEEE International Conference on Services Computing, SCC 2019, Milan, Italy, July 8-13, 2019*, Elisa Bertino, Carl K. Chang, Peter Chen, Ernesto Damiani, Michael Goul, and Katsunori Oyama (Eds.). IEEE, 234–236. https://doi.org/10.1109/SCC.2019.00048

Paul A Grassi, James L Fenton, EM Newton, RA Perlner, AR Regenscheid, WE Burr, JP Richer, NB Lefkovitz, JM Danker, Yee-Yin Choong, et al. 2017. NIST Special Publication 800-63b: Digital Identity Guidelines. *Enrollment and Identity Proofing Requirements. url: https://pages. nist. gov/800-63-3/sp800-63a. html* (2017).

Martin Grotzke. acc. May 2020. Kryo. https://github.com/EsotericSoftware/kryo.

Paul Hamill. 2004. *Unit test frameworks: tools for high-quality software development.* O'Reilly Media, Inc.

Andrew K. Hirsch and Deepak Garg. 2022. Pirouette: higher-order typed functional choreographies. *Proc. ACM Program. Lang.* 6, POPL (2022), 1–27. https://doi.org/10.1145/3498684

Kohei Honda, Nobuko Yoshida, and Marco Carbone. 2016. Multiparty Asynchronous Session Types. *J. ACM* 63, 1 (2016), 9. https://doi.org/10.1145/2827695 Also: POPL, pages 273–284, 2008.

Urs Hunkeler, Hong Linh Truong, and Andy Stanford-Clark. 2008. MQTT-S—A publish/subscribe protocol for Wireless Sensor Networks. In *2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08).* IEEE, 791–798.

Hans Hüttel, Ivan Lanese, Vasco T. Vasconcelos, Luís Caires, Marco Carbone, Pierre-Malo Deniélou, Dimitris Mostrous, Luca Padovani, António Ravara, Emilio Tuosto, Hugo Torres Vieira, and Gianluigi Zavattaro. 2016. Foundations of Session Types and Behavioural Contracts. *ACM Comput. Surv.* 49, 1 (2016), 3:1–3:36. https://doi.org/10.1145/2873052

Intl. Telecommunication Union. 1996. Recommendation Z.120: Message Sequence Chart.

Sung-Shik Jongmans and Petra van den Bos. 2022. A Predicate Transformer for Choreographies - Computing Preconditions in Choreographic Programming. In *Programming Languages and Systems - 31st European Symposium on Programming, ESOP 2022, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2022, Munich, Germany, April 2-7, 2022, Proceedings (Lecture Notes in Computer Science, Vol. 13240)*, Ilya Sergey (Ed.). Springer, 520–547. https://doi.org/10.1007/978-3-030-99336-8_19

Sung-Shik Jongmans and Nobuko Yoshida. 2020. Exploring Type-Level Bisimilarity towards More Expressive Multiparty Session Types. In *Programming Languages and Systems - 29th European Symposium on Programming, ESOP 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, Dublin, Ireland, April 25-30, 2020, Proceedings (Lecture Notes in Computer Science, Vol. 12075)*, Peter Müller (Ed.). Springer, 251–279. https://doi.org/10.1007/978-3-030-44914-8_10

Sung-Shik T. Q. Jongmans, Francesco Santini, and Farhad Arbab. 2015. Partially distributed coordination with Reo and constraint automata. *Serv. Oriented Comput. Appl.* 9, 3-4 (2015), 311–339. https://doi.org/10.1007/s11761-015-0177-y

Anatolii Alekseevich Karatsuba and Yu P Ofman. 1962. Multiplication of many-digital numbers by automatic computers. In *Doklady Akademii Nauk*, Vol. 145. Russian Academy of Sciences, 293–294.

Donald Knuth. 1998. Section 5.2. 4: Sorting by merging. *The Art of Computer Programming* 3 (1998), 158–168.

Dimitrios Kouzapas, Ornela Dardha, Roly Perera, and Simon J. Gay. 2018. Typechecking protocols with Mungo and StMungo: A session type toolchain for Java. *Sci. Comput. Program.* 155 (2018), 52–75. https://doi.org/10.1016/j.scico.2017.10.006

Ivan Lanese, Fabrizio Montesi, and Gianluigi Zavattaro. 2013. Amending Choreographies. In *Proceedings 9th International Workshop on Automated Specification and Verification of Web Systems, WWV 2013, Florence, Italy, 6th June 2013 (EPTCS, Vol. 123)*, António Ravara and Josep Silva (Eds.). 34–48. https://doi.org/10.4204/EPTCS.123.5

Tanakorn Leesatapornwongsa, Jeffrey F. Lukman, Shan Lu, and Haryadi S. Gunawi. 2016. TaxDC: A Taxonomy of Non-Deterministic Concurrency Bugs in Datacenter Distributed Systems. In *Proc. of ASPLOS.* 517–530.

Jed Liu, Michael D. George, K. Vikram, Xin Qi, Lucas Waye, and Andrew C. Myers. 2009. Fabric: a platform for secure distributed computation and storage. In *Proceedings of the 22nd ACM Symposium on Operating Systems Principles 2009, SOSP 2009, Big Sky, Montana, USA, October 11-14, 2009*, Jeanna Neefe Matthews and Thomas E. Anderson (Eds.). ACM, 321–334. https://doi.org/10.1145/1629575.1629606

Alberto Lluch-Lafuente, Flemming Nielson, and Hanne Riis Nielson. 2015. Discretionary Information Flow Control for Interaction-Oriented Specifications. In *Logic, Rewriting, and Concurrency (Lecture Notes in Computer Science, Vol. 9200).* Springer, 427–450.

Hugo A. López and Kai Heussen. 2017. Choreographing cyber-physical distributed control systems for the energy sector. In *SAC*. ACM, 437–443.

Hugo A. López, Flemming Nielson, and Hanne Riis Nielson. 2016. Enforcing Availability in Failure-Aware Communicating Systems. In *FORTE (Lecture Notes in Computer Science, Vol. 9688).* Springer, 195–211.

Shan Lu, Soyeon Park, Eunsoo Seo, and Yuanyuan Zhou. 2008. Learning from mistakes: a comprehensive study on real world concurrency bug characteristics. In *Proc. of ASPLOS.* 329–339.

Robin Milner. 1980. *A Calculus of Communicating Systems.* Lecture Notes in Computer Science, Vol. 92. Springer. https://doi.org/10.1007/3-540-10235-3

Fabrizio Montesi. 2013. *Choreographic Programming.* Ph.D. Thesis. IT University of Copenhagen. http://www.fabriziomontesi.com/files/choreographic_programming.pdf.

Fabrizio Montesi. 2015. Kickstarting Choreographic Programming. In *Web Services, Formal Methods, and Behavioral Types - 11th International Workshop, WS-FM 2014, Eindhoven, The Netherlands, September 11-12, 2014, and 12th International Workshop, WS-FM/BEAT 2015, Madrid, Spain, September 4-5, 2015, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 9421)*, Thomas T. Hildebrandt, António Ravara, Jan Martijn E. M. van der Werf, and Matthias Weidlich (Eds.). Springer, 3–10. https://doi.org/10.1007/978-3-319-33612-1_1

Fabrizio Montesi. 2022. *Introduction to Choreographies.* Cambridge University Press. To appear.

Fabrizio Montesi, Claudio Guidi, and Gianluigi Zavattaro. 2014. Service-Oriented Programming with Jolie. In *Web Services Foundations*, Athman Bouguettaya, Quan Z. Sheng, and Florian Daniel (Eds.). Springer, 81–107. https://doi.org/10.1007/978-1-4614-7518-7_4

Fabrizio Montesi and Marco Peressotti. 2017. Choreographies meet Communication Failures. *CoRR* abs/1712.05465 (2017). arXiv:1712.05465 http://arxiv.org/abs/1712.05465

Adriaan Moors, Frank Piessens, and Martin Odersky. 2008. Generics of a higher kind. In *Proceedings of the 23rd Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2008, October 19-23, 2008, Nashville, TN, USA*, Gail E. Harris (Ed.). ACM, 423–438. https://doi.org/10.1145/1449764.1449798

Tom Murphy VII, Karl Crary, and Robert Harper. 2007. Type-Safe Distributed Programming with ML5. In *Trustworthy Global Computing, Third Symposium, TGC 2007, Sophia-Antipolis, France, November 5-6, 2007, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 4912)*, Gilles Barthe and Cédric Fournet (Eds.). Springer, 108–123. https://doi.org/10.1007/978-3-540-78663-4_9

Tom Murphy VII, Karl Crary, Robert Harper, and Frank Pfenning. 2004. A Symmetric Modal Lambda Calculus for Distributed Computing. In *19th IEEE Symposium on Logic in Computer Science (LICS 2004), 14-17 July 2004, Turku, Finland, Proceedings.* IEEE Computer Society, 286–295. https://doi.org/10.1109/LICS.2004.1319623

James Murty. 2008. *Programming amazon web services: S3, EC2, SQS, FPS, and SimpleDB.* "O'Reilly Media, Inc.".

Maurice Naftalin and Philip Wadler. 2007. *Java generics and collections.* " O'Reilly Media, Inc.".

Matthias Neubauer and Peter Thiemann. 2005. From sequential programs to multi-tier applications by program transformation. In *Proceedings of the 32nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2005, Long Beach, California, USA, January 12-14, 2005*, Jens Palsberg and Martín Abadi (Eds.). ACM, 221–232. https://doi.org/10.1145/1040305.1040324

Sam Newman. 2021. *Building microservices.* "O'Reilly Media, Inc.".

Rumyana Neykova and Nobuko Yoshida. 2017. Let it recover: multiparty protocol-induced recovery. In *Proceedings of the 26th International Conference on Compiler Construction, Austin, TX, USA, February 5-6, 2017*, Peng Wu and Sebastian Hack (Eds.). ACM, 98–108. http://dl.acm.org/citation.cfm?id=3033031

Object Management Group. 2011. Business Process Model and Notation. http://www.omg.org/spec/BPMN/2.0/.

Peter W. O'Hearn. 2018. Experience Developing and Deploying Concurrency Analysis at Facebook. In *Static Analysis - 25th International Symposium, SAS 2018, Freiburg, Germany, August 29-31, 2018, Proceedings (Lecture Notes in Computer Science, Vol. 11002)*, Andreas Podelski (Ed.). Springer, 56–70. https://doi.org/10.1007/978-3-319-99725-4_5

OpenID Foundation. 2014. OpenID Specification. https://openid.net/developers/specs/.

OpenJDK. acc. May 2020. Loom - Fibers, Continuations and Tail-Calls for the JVM. https://openjdk.java.net/projects/loom/.

Tomas Petricek and Jon Skeet. 2009. *Real World Functional Programming: With Examples in F# and C.* Manning Publications Co.

Zongyan Qiu, Xiangpeng Zhao, Chao Cai, and Hongli Yang. 2007. Towards the theoretical foundation of choreography. In *WWW.* IEEE Computer Society Press, United States, 973–982.

Alceste Scalas, Ornela Dardha, Raymond Hu, and Nobuko Yoshida. 2017. A Linear Decomposition of Multiparty Sessions for Safe Distributed Programming. In *31st European Conference on Object-Oriented Programming, ECOOP 2017, June 19-23, 2017, Barcelona, Spain (LIPIcs, Vol. 74)*, Peter Müller (Ed.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 24:1–24:31. https://doi.org/10.4230/LIPIcs.ECOOP.2017.24

Alceste Scalas, Nobuko Yoshida, and Elias Benussi. 2019. Verifying message-passing programs with dependent behavioural types. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2019, Phoenix, AZ, USA, June 22-26, 2019*, Kathryn S. McKinley and Kathleen Fisher (Eds.). ACM, 502–516. https://doi.org/10.1145/3314221.3322484

Manuel Serrano, Erick Gallesio, and Florian Loitsch. 2006. Hop: a language for programming the web 2.0. In *Companion to the 21th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2006, October 22-26, 2006, Portland, Oregon, USA*, Peri L. Tarr and William R. Cook (Eds.). ACM, 975–985. https://doi.org/10.1145/1176617.1176756

Manu Sporny, Toby Inkster, Henry Story, Bruno Harbulot, and Reto Bachmann-Gmür. 2011. Webid 1.0: Web identification and discovery. *W3C Editors Draft* (2011).

K. Narendra Swaroop, Kavitha Chandu, Ramesh Gorrepotu, and Subimal Deb. 2019. A health monitoring system for vital signs using IoT. *Internet of Things* 5 (2019), 116 – 129. https://doi.org/10.1016/j.iot.2019.01.004

W3C. 2004. WS Choreography Description Language. http://www.w3.org/TR/ws-cdl-10/.

Pascal Weisenburger, Mirko Köhler, and Guido Salvaneschi. 2018. Distributed system development with ScalaLoci. *Proc. ACM Program. Lang.* 2, OOPSLA (2018), 129:1–129:30. https://doi.org/10.1145/3276499

Pascal Weisenburger and Guido Salvaneschi. 2019. Multitier Modules. In *33rd European Conference on Object-Oriented Programming, ECOOP 2019, July 15-19, 2019, London, United Kingdom (LIPIcs, Vol. 134)*, Alastair F. Donaldson (Ed.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 3:1–3:29. https://doi.org/10.4230/LIPIcs.ECOOP.2019.3

Pascal Weisenburger, Johannes Wirth, and Guido Salvaneschi. 2020. A Survey of Multitier Programming. *ACM Comput. Surv.* 53, 4 (2020), 81:1–81:35. https://doi.org/10.1145/3397495

Bill Wilder. 2012. *Cloud architecture patterns: using microsoft azure*. "O'Reilly Media, Inc.".

Derek Wyatt. 2013. *Akka Concurrency*. Artima Incorporation, Sunnyvale, CA, USA.

A PROJECTION TO JAVA

A.1 Projection

We omit modifiers (MD) and annotations (AN), they are rendered by the projection as they are.

$$
\begin{array}{rl}
(\mathit{Enum}) & \llparenthesis \mathtt{enum}\ id@\mathsf{A}\ \{\overline{id}\} \rrparenthesis = \mathtt{enum}\ id\ \{\overline{id}\} \\
(\mathit{Interface}) & \llparenthesis \mathtt{interface}\ id@(\overline{\mathsf{A}})\ \langle\overline{FTP}\rangle\ \mathtt{extends}\ TE\ \overline{\&\ TE}\ \{\overline{MDef;}\} \rrparenthesis = \\
& \left[\ \mathtt{interface}\ \mathrm{name}(id, \mathsf{A}, \overline{\mathsf{A}})\ \langle\overline{\llparenthesis FTP \rrparenthesis}\rangle\ \mathtt{extends}\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}\ \overline{\&\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}}\ \{\overline{\llparenthesis MDef \rrparenthesis^{\mathsf{A}};}\}\ \middle|\ \mathsf{A} \in \overline{\mathsf{A}}\ \right] \\
(\mathit{Class}) & \llparenthesis \mathtt{class}\ id@(\overline{\mathsf{A}})\ \langle\overline{FTP}\rangle\ \mathtt{extends}\ TE\ \mathtt{implements}\ TE\overline{\mathtt{,}\ TE}\ \{\overline{CField}\ \overline{CConst}\ \overline{MDef;} \\
& \overline{MDef\{Stm\}}\} \rrparenthesis = \left[\ \mathtt{class}\ \mathrm{name}(id, \mathsf{A}, \overline{\mathsf{A}})\ \mathtt{extends}\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}\ \mathtt{implements}\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}\ \overline{\mathtt{,}\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}} \right. \\
& \left. \{\overline{\llparenthesis CField \rrparenthesis^{\mathsf{A}}}\ \overline{\llparenthesis CConst \rrparenthesis^{\mathsf{A}}}\ \overline{\llparenthesis MDef \rrparenthesis^{\mathsf{A}};}\ \overline{\llparenthesis MDef \rrparenthesis^{\mathsf{A}}\{[\![\llparenthesis Stm \rrparenthesis^{\mathsf{A}}]\!]\}}\}\ \middle|\ \mathsf{A} \in \overline{\mathsf{A}}\ \right] \\
(\mathit{FTP}) & \llparenthesis id@(\overline{\mathsf{A}})\ \mathtt{extends}\ TE\ \overline{\mathtt{,}\ TE} \rrparenthesis = \begin{cases} [\ id\_\mathsf{A}\ \mathtt{extends}\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}\ \overline{\&\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}}\ \mid \mathsf{A} \in \overline{\mathsf{A}}\ ] & \text{if } |\overline{\mathsf{A}}| \geq 1 \\ id\ \mathtt{extends}\ \llparenthesis TE \rrparenthesis^{\mathsf{A}} & \text{otherwise} \end{cases} \\
(\mathit{TE}) & \llparenthesis \mathtt{void} \rrparenthesis^{\mathsf{A}} = \mathtt{void} \\
& \llparenthesis id@(\overline{\mathsf{B}})\mathtt{<}\overline{TE}\mathtt{>} \rrparenthesis^{\mathsf{A}} = \begin{cases} id\mathtt{<}\overline{\llparenthesis TE \rrparenthesis^{\mathsf{A}}}\mathtt{>} & \overline{\mathsf{B}} = \mathsf{A} \\ id\_A'\mathtt{<}\overline{\llparenthesis TE \rrparenthesis^{\mathsf{A}}}\mathtt{>} & \mathsf{A} \text{ is the } i\text{-th element of } \overline{\mathsf{B}} \text{ and } \mathtt{roleName}(id, i) = A' \\ \mathtt{Unit} & \text{otherwise} \end{cases} \\
(\mathit{MDef}) & \llparenthesis \langle\overline{FTP}\rangle\ TE\ id\ (\overline{TE\ id}) \rrparenthesis^{\mathsf{A}} = \langle\overline{\llparenthesis FTP \rrparenthesis}\rangle\ \llparenthesis TE \rrparenthesis^{\mathsf{A}}\ id\ (\overline{\llparenthesis TE\ id \rrparenthesis^{\mathsf{A}}}) \\
(\mathit{CField}) & \llparenthesis \overline{TE\ id;} \rrparenthesis^{\mathsf{A}} = \begin{cases} \llparenthesis TE \rrparenthesis^{\mathsf{A}} id; & \text{if } \mathsf{A} \in \mathrm{rolesOf}(TE) \\ [blank] & \text{otherwise} \end{cases} \\
(\mathit{CConst}) & \llparenthesis id\ (\overline{TE\ id})\{Stm\} \rrparenthesis^{\mathsf{A}} = id\_\mathsf{A}(\overline{\llparenthesis TE\ id \rrparenthesis^{\mathsf{A}}})\{[\![\llparenthesis Stm \rrparenthesis^{\mathsf{A}}]\!]\} \\
(\mathit{Stm}) & \llparenthesis nil \rrparenthesis^{\mathsf{A}} = \llparenthesis [blank] \rrparenthesis^{\mathsf{A}} = [blank] \\
& \llparenthesis \mathtt{return}\ Exp; \rrparenthesis^{\mathsf{A}} = \mathtt{return}\ \llparenthesis Exp \rrparenthesis^{\mathsf{A}}; \\
& \llparenthesis Exp; Stm \rrparenthesis^{\mathsf{A}} = \begin{cases} \mathtt{switch}(\llparenthesis Exp \rrparenthesis^{\mathsf{A}})\{ & \text{if } \mathrm{typeOf}(Exp) \mathrel{\mathtt{<:}} \mathtt{Enum@}\mathsf{A}, \\ \quad \mathtt{case}\ id_3\mathtt{->}\{\llparenthesis Stm \rrparenthesis^{\mathsf{A}}\} & \quad Exp = \overline{Exp'}.\langle\overline{TE}\rangle id_1(id_2@\mathsf{A}'.id_3) \text{ and} \\ \quad \mathtt{default->\{throw\ \ldots\}\}} & \quad \mathtt{@SelectionMethod} \in \mathtt{annotOf}(id_1) \\ \llparenthesis Exp \rrparenthesis^{\mathsf{A}}; \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{if } \mathsf{A} \in \mathrm{rolesOf}(Exp) \\ \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{otherwise} \end{cases} \\
& \llparenthesis TE\ id\mathtt{=}Exp; Stm \rrparenthesis^{\mathsf{A}} = \begin{cases} \llparenthesis TE \rrparenthesis^{\mathsf{A}}\mathtt{=}\llparenthesis Exp \rrparenthesis^{\mathsf{A}}; \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{if } \mathsf{A} \in \mathrm{rolesOf}(TE) \\ \llparenthesis Exp \rrparenthesis^{\mathsf{A}}; \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{if } \mathsf{A} \in \mathrm{rolesOf}(Exp) \setminus \mathrm{rolesOf}(TE) \\ \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{otherwise} \end{cases} \\
& \llparenthesis Exp_1\ AsgOp\ Exp_2; Stm \rrparenthesis^{\mathsf{A}} = \begin{cases} \llparenthesis Exp_1 \rrparenthesis^{\mathsf{A}} AsgOp \llparenthesis Exp_2 \rrparenthesis^{\mathsf{A}}; \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{if } \mathsf{A} \in \mathrm{rolesOf}(\mathrm{typeOf}(Exp)) \\ \llparenthesis Exp_1 \rrparenthesis^{\mathsf{A}}\mathtt{.id(}\llparenthesis Exp_2 \rrparenthesis^{\mathsf{A}}\mathtt{)}; \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{if } \mathsf{A} \in \mathrm{rolesOf}(Exp_1, Exp_2) \\ \llparenthesis Stm \rrparenthesis^{\mathsf{A}} & \text{otherwise} \end{cases} \\
& \llparenthesis \mathtt{if}(Exp)\{Stm_1\}\mathtt{else}\{Stm_2\}Stm \rrparenthesis^{\mathsf{A}} =
\end{array}
$$

$$\begin{cases} \texttt{if}(\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}})\{\llparenthesis \mathit{Stm}_1 \rrparenthesis^{\mathsf{A}}\}\texttt{else}\{\llparenthesis \mathit{Stm}_2 \rrparenthesis^{\mathsf{A}}\}\llparenthesis \mathit{Stm} \rrparenthesis^{\mathsf{A}} & \text{if typeOf}(\mathit{Exp}) = \texttt{boolean@A} \\ \llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}; \left\{ [\![\llparenthesis \mathit{Stm}_1 \rrparenthesis^{\mathsf{A}}]\!] \sqcup [\![\llparenthesis \mathit{Stm}_2 \rrparenthesis^{\mathsf{A}}]\!] \right\} \llparenthesis \mathit{Stm} \rrparenthesis^{\mathsf{A}} & \text{otherwise} \end{cases}$$

$$\llparenthesis \{\mathit{Stm}_1\}\mathit{Stm}_2 \rrparenthesis^{\mathsf{A}} = \{\llparenthesis \mathit{Stm}_1 \rrparenthesis^{\mathsf{A}}\}\llparenthesis \mathit{Stm}_2 \rrparenthesis^{\mathsf{A}}$$

$$\llparenthesis \texttt{try}\{\mathit{Stm}\}\overline{\texttt{catch}(\mathit{TE}\ \mathit{id})\{\mathit{Stm}\}}\ \mathit{Stm} \rrparenthesis^{\mathsf{A}} = \texttt{try}\{\llparenthesis \mathit{Stm} \rrparenthesis^{\mathsf{A}}\}\overline{\llparenthesis \texttt{catch}(\mathit{TE}\ \mathit{id})\{\mathit{Stm}\} \rrparenthesis^{\mathsf{A}}}\ \llparenthesis \mathit{Stm} \rrparenthesis^{\mathsf{A}}$$

$$\llparenthesis \texttt{catch}(\mathit{TE}\ \mathit{id})\{\mathit{Stm}\} \rrparenthesis^{\mathsf{A}} = \begin{cases} \texttt{catch}(\llparenthesis \mathit{TE} \rrparenthesis^{\mathsf{A}}\ \mathit{id})\{\llparenthesis \mathit{Stm} \rrparenthesis^{\mathsf{A}}\} & \text{if } \mathsf{A} \in \text{rolesOf}(\mathit{TE}) \\ [\mathit{blank}] & \text{otherwise} \end{cases}$$

$$(\mathit{Exp}) \quad \llparenthesis \mathit{lit} \rrparenthesis^{\mathsf{A}} = \begin{cases} l & \text{if } \mathit{lit} = l\texttt{@}(\overline{\mathsf{B}}) \text{ and } \mathsf{A} \in \overline{\mathsf{B}} \\ \texttt{Unit.id} & \text{otherwise} \end{cases}$$

$$\llparenthesis \mathit{Exp}\ \mathit{BinOp}\ \mathit{Exp}' \rrparenthesis^{\mathsf{A}} = \begin{cases} \llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}} \mathit{BinOp}\ \llparenthesis \mathit{Exp}' \rrparenthesis^{\mathsf{A}} & \text{if } \begin{pmatrix} \mathit{BinOp} \in \{\texttt{\&\&}, \texttt{||}\} \\ \wedge\ \text{rolesOf}(\mathit{Exp}') = \{\mathsf{A}\} \end{pmatrix} \\ & \vee \begin{pmatrix} \mathit{BinOp} \notin \{\texttt{\&\&}, \texttt{||}\} \\ \wedge\ \mathsf{A} \in \text{rolesOf}(\text{typeOf}(\mathit{Exp})) \end{pmatrix} \\ \texttt{Unit.id}(\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}, \llparenthesis \mathit{Exp}' \rrparenthesis^{\mathsf{A}}) & \text{if } \begin{pmatrix} \mathit{BinOp} \in \{\texttt{\&\&}, \texttt{||}\} \\ \wedge\ \text{rolesOf}(\mathit{Exp}') = \{\mathsf{A'}\} \end{pmatrix} \\ & \vee \mathit{BinOp} \notin \{\texttt{\&\&}, \texttt{||}\} \end{cases}$$

let $\text{head}(\mathit{Exp}.\mathit{Exp}') = \mathit{Exp}_h$, $\text{rest}(\mathit{Exp}.\mathit{Exp}') = \mathit{Exp}_r$

$$\llparenthesis \mathit{Exp}.\mathit{Exp}' \rrparenthesis^{\mathsf{A}} = \begin{cases} \llparenthesis \mathit{Exp}_h \rrparenthesis^{\mathsf{A}}\texttt{.id}(\llparenthesis \mathit{Exp}_r \rrparenthesis^{\mathsf{A}}) & \text{if } \llparenthesis \mathit{Exp}_h \rrparenthesis^{\mathsf{A}} \in \{\texttt{Unit.}\bullet\} \\ \texttt{Unit.id}(\llparenthesis \mathit{Exp}_h \rrparenthesis^{\mathsf{A}})\texttt{.}\overline{\mathit{Exp}_1} & \text{if } \llparenthesis \mathit{Exp}_r \rrparenthesis^{\mathsf{A}} \in \{\texttt{Unit.}\overline{\mathit{Exp}_1}\} \\ \llparenthesis \mathit{Exp}_h \rrparenthesis^{\mathsf{A}}.\llparenthesis \mathit{Exp}_r \rrparenthesis^{\mathsf{A}} & \text{otherwise} \end{cases}$$

$$\llparenthesis \langle\overline{\mathit{TE}}\rangle \mathit{id}(\overline{\mathit{Exp}}) \rrparenthesis^{\mathsf{A}} = \begin{cases} \langle\overline{\llparenthesis \mathit{TE} \rrparenthesis^{\mathsf{A}}}\rangle \mathit{id}(\overline{\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}}) & \text{if } \mathsf{A} \in \text{rolesOf}(\text{typeOf}(\langle\overline{\mathit{TE}}\rangle \mathit{id}(\overline{\mathit{Exp}}))) \\ \texttt{Unit.id}(\overline{\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}}) & \text{otherwise} \end{cases}$$

$$\llparenthesis \mathit{id}\texttt{@}(\overline{\mathsf{B}})\texttt{.}\langle\overline{\mathit{TE}}\rangle \mathit{id}(\overline{\mathit{Exp}}) \rrparenthesis^{\mathsf{A}} = \begin{cases} \llparenthesis \mathit{id}\texttt{@}(\overline{\mathsf{B}}) \rrparenthesis^{\mathsf{A}}\texttt{.}\langle\overline{\llparenthesis \mathit{TE} \rrparenthesis^{\mathsf{A}}}\rangle \mathit{id}(\overline{\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}}) & \mathsf{A} \in \overline{\mathsf{B}} \\ \texttt{Unit.id}(\overline{\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}}) & \text{otherwise} \end{cases}$$

$$\llparenthesis \texttt{new}\ \langle\overline{\mathit{TE}}\rangle \mathit{id}\langle\overline{\mathit{TE}}\rangle(\overline{\mathit{Exp}}) \rrparenthesis^{\mathsf{A}} = \begin{cases} \texttt{new}\ \langle\overline{\llparenthesis \mathit{TE} \rrparenthesis^{\mathsf{A}}}\rangle \llparenthesis \mathit{id}\texttt{@}(\overline{\mathsf{B}})\langle\overline{\mathit{TE}}\rangle \rrparenthesis^{\mathsf{A}}(\overline{\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}}) & \mathsf{A} \in \overline{\mathsf{B}} \\ \texttt{Unit.id}(\overline{\llparenthesis \mathit{Exp} \rrparenthesis^{\mathsf{A}}}) & \text{otherwise} \end{cases}$$

$$(\mathit{FAcc}) \quad \llparenthesis \mathit{id} \rrparenthesis^{\mathsf{A}} = \begin{cases} \mathit{id} & \mathsf{A} \in \text{rolesOf}(\mathit{id}) \\ \texttt{Unit.id} & \text{otherwise} \end{cases}$$

$$\llparenthesis \mathit{id}\texttt{@}(\overline{\mathsf{B}})\texttt{.}f \rrparenthesis^{\mathsf{A}} = \begin{cases} \llparenthesis \mathit{id}\texttt{@}(\overline{\mathsf{B}}) \rrparenthesis^{\mathsf{A}}\texttt{.}f & \mathsf{A} \in \text{rolesOf}(f) \\ \texttt{Unit.id} & \text{otherwise} \end{cases}$$

A.2 Merging

$$\dot{\bigsqcup}\ \overline{Stm} = \dot{\bigsqcup}(Stm_1, \cdots, Stm_n) = [\![Stm_1]\!] \sqcup \cdots \sqcup [\![Stm_n]\!]$$

Statements

$$\texttt{return}\ Exp \sqcup \texttt{return}\ Exp' = \texttt{return}\ Exp \sqcup Exp'$$

$$TE\ id; Stm \sqcup TE\ id; Stm' = TE\ id; (Stm \sqcup Stm')$$

$$(Exp_1\ AsgOp\ Exp_2; Stm) \sqcup (Exp'_1\ AsgOp\ Exp'_2; Stm')$$

$$= (Exp_1 \sqcup Exp'_1)\ AsgOp\ (Exp_2 \sqcup Exp'_2); (Stm \sqcup Stm')$$

$$(Exp; Stm) \sqcup (Exp'; Stm') = (Exp \sqcup Exp'); (Stm \sqcup Stm')$$

$$\{Stm_1\}\ Stm_2 \sqcup \{Stm'_1\}\ Stm'_2 = \{Stm_1 \sqcup Stm'_1\}\ (Stm_2 \sqcup Stm'_2)$$

$$\texttt{if}(Exp)\{Stm_1\}\texttt{else}\{Stm_2\}Stm \sqcup \texttt{if}(Exp')\{Stm'_1\}\texttt{else}\{Stm'_2\}Stm'$$

$$= \texttt{if}(Exp \sqcup Exp')\{Stm_1 \sqcup Stm'_1\}\texttt{else}\{Stm_2 \sqcup Stm'_2\}(Stm \sqcup Stm')$$

$$\begin{array}{l} \texttt{switch}\ (Exp)\{ \\ \quad \texttt{case}\ id_a\texttt{->}\{Stm_a\} \\ \quad \dots \\ \quad \underline{\texttt{case}\ id_x\texttt{->}\{Stm_x\}} \\ \quad \texttt{case}\ id_y\texttt{->}\{Stm_y\} \\ \quad \texttt{default->}\{Stm_{d1}\} \\ \}\ Stm \end{array} \sqcup \begin{array}{l} \texttt{switch}\ (Exp)\{ \\ \quad \texttt{case}\ id_a\texttt{->}\{Stm'_a\} \\ \quad \dots \\ \quad \underline{\texttt{case}\ id_x\texttt{->}\{Stm'_x\}} \\ \quad \texttt{case}\ id_z\texttt{->}\{Stm_z\} \\ \quad \texttt{default->}\{Stm_{d2}\} \\ \}\ Stm' \end{array} = \begin{array}{l} \texttt{switch}\ (Exp \sqcup Exp')\{ \\ \quad \texttt{case}\ id_a\texttt{->}\{Stm_a \sqcup Stm'_a\} \\ \quad \dots \\ \quad \underline{\texttt{case}\ id_x\texttt{->}\{Stm_x \sqcup Stm'_x\}} \\ \quad \underline{\texttt{case}\ id_y\texttt{->}\{Stm_y\}} \\ \quad \texttt{case}\ id_z\texttt{->}\{Stm_z\} \\ \quad \texttt{default->}\{Stm_{d1} \sqcup Stm_{d2}\} \\ \}\ Stm \sqcup Stm' \end{array}$$

$$\begin{array}{l} \texttt{try}\ \{Stm_1\}\ \overline{\texttt{catch}\ (TE\ id)\ \{Stm\}}\ Stm_2 \sqcup \texttt{try}\ \{Stm_3\}\ \overline{\texttt{catch}\ (TE\ id)\ \{Stm'\}}\ Stm_4 \\ \quad = \texttt{try}\ \{Stm_1 \sqcup Stm_3\}\ \overline{\texttt{catch}\ (TE\ id)\ \{Stm \sqcup Stm'\}}\ Stm_2 \sqcup Stm_4 \end{array}$$

Expressions

$$\text{let } \bullet \in \{nil, [blank], \texttt{null}, \texttt{this}, \texttt{super}, id\},\ \bullet \sqcup \bullet = \bullet$$

$$\text{let } \bullet \in \{\texttt{new}\ id\ \langle\overline{TE}\rangle, id\ \langle\overline{TE}\rangle\ \},\ \bullet\ (\overline{Exp}) \sqcup \bullet\ (\overline{Exp'}) = \bullet\ (\overline{Exp \sqcup Exp'})$$

$$(Exp_1\ BinOp\ Exp_2) \sqcup (Exp'_1\ BinOp\ Exp'_2;) = (Exp_1 \sqcup Exp'_1)\ BinOp\ (Exp_2 \sqcup Exp'_2)$$

$$Exp_1.Exp_2 \sqcup Exp_3.Exp_4 = (Exp_1 \sqcup Exp_3)(.Exp_2 \sqcup .Exp_4)$$

$$.id \sqcup .id = .id \qquad .id\langle\overline{TE}\rangle(\overline{Exp}) \sqcup .id\langle\overline{TE}\rangle(\overline{Exp'}) = .id\langle\overline{TE}\rangle(\overline{Exp \sqcup Exp'})$$

$$.Exp_1.Exp_2 \sqcup .Exp_3.Exp_4 = (.Exp_1 \sqcup .Exp_3)(.Exp_2 \sqcup .Exp_4)$$

A.3 Normaliser

Statements

$\llbracket nil \rrbracket = nil \qquad \llbracket \texttt{return}\ Exp; \rrbracket = \texttt{return}\ \llbracket Exp \rrbracket; \qquad \llbracket TE\ id; Stm \rrbracket = TE\ id; \llbracket Stm \rrbracket$

$\llbracket Exp\ AsgOp\ Exp'; Stm \rrbracket = \llbracket Exp \rrbracket\ AsgOp\ \llbracket Exp' \rrbracket; \llbracket Stm \rrbracket$

$\llbracket \{Stm\}\ Stm \rrbracket = \{\llbracket Stm \rrbracket\}\ \llbracket Stm \rrbracket$

$$\textsc{noop}(Exp) = \begin{cases} [blank] & \text{if } Exp \in \{\texttt{Unit.id},\ id\ \overline{.id}, \texttt{this}, \texttt{null}\} \\ Exp & \text{otherwise} \end{cases}$$

$$\llbracket Exp; Stm \rrbracket = \begin{cases} \llbracket Stm \rrbracket & \text{if } \textsc{noop}(\llbracket Exp \rrbracket) = [blank] \\ \llbracket Exp \rrbracket; \llbracket Stm \rrbracket & \text{otherwise} \end{cases}$$

$\llbracket \texttt{if}(Exp)\{Stm_1\}\texttt{else}\{Stm_2\}Stm \rrbracket = \texttt{if}(\llbracket Exp \rrbracket)\{\llbracket Stm_1 \rrbracket\}\texttt{else}\{\llbracket Stm_2 \rrbracket\}\llbracket Stm \rrbracket$

$\llbracket \texttt{switch}(Exp)\{\overline{\texttt{case}\ id\texttt{->}\{Stm\}}\ \texttt{default->}Stm'\}\ Stm \rrbracket$

$\quad = \texttt{switch}(\llbracket Exp \rrbracket)\{\overline{\texttt{case}\ id\texttt{->}\{\llbracket Stm \rrbracket\}}\ \texttt{default->}\llbracket Stm' \rrbracket\}\ \llbracket Stm \rrbracket$

$\llbracket \texttt{try}\ \{Stm\}\ \overline{\texttt{catch}\ (TE\ id)\ \{Stm\}}\ Stm \rrbracket = \texttt{try}\ \{\llbracket Stm \rrbracket\}\ \overline{\texttt{catch}\ (TE\ id)\ \{\llbracket Stm \rrbracket\}}\ \llbracket Stm \rrbracket$

Expressions

$\llbracket \texttt{null} \rrbracket = \texttt{null} \quad \llbracket \texttt{null} \rrbracket^\star = \langle \text{false}, \texttt{null} \rangle \quad \llbracket \texttt{this} \rrbracket = \texttt{this} \quad \llbracket \texttt{this} \rrbracket^\star = \langle \text{false}, \texttt{this} \rangle$

$\llbracket id \rrbracket = id \qquad \llbracket id \rrbracket^\star = \langle \text{false}, id \rangle \qquad \text{let } \llbracket id\langle \overline{TE} \rangle(\overline{Exp}) \rrbracket^\star = \langle \bullet, \diamond \rangle,\ \llbracket id\langle \overline{TE} \rangle(\overline{Exp}) \rrbracket = \diamond$

$\text{let } \overline{\llbracket Exp \rrbracket^\star} = \overline{\langle \bullet, \diamond \rangle},\ \llbracket id\langle \overline{TE} \rangle(\overline{Exp}) \rrbracket^\star = \langle \bigvee \overline{\bullet}, id\langle \overline{TE} \rangle(\overline{\diamond}) \rangle$

$\text{let } \llbracket \texttt{new}\ id\langle \overline{TE} \rangle(\overline{Exp}) \rrbracket^\star = \langle \bullet, \diamond \rangle,\ \llbracket \texttt{new}\ id\langle \overline{TE} \rangle(\overline{Exp}) \rrbracket = \diamond$

$\text{let } \overline{\llbracket Exp \rrbracket^\star} = \overline{\langle \bullet, \diamond \rangle},\ \llbracket \texttt{new}\ id\langle \overline{TE} \rangle(\overline{Exp}) \rrbracket^\star = \langle \bigvee \overline{\bullet}, \texttt{new}\ id\langle \overline{TE} \rangle(\overline{\diamond}) \rangle$

$\llbracket Exp\ BinOp\ Exp' \rrbracket = \llbracket Exp \rrbracket\ BinOp\ \llbracket Exp' \rrbracket$

$$\text{let } \llbracket Exp.Exp' \rrbracket^\star = \langle \bullet, \diamond \rangle,\ \llbracket Exp.Exp' \rrbracket = \begin{cases} \llbracket \diamond \rrbracket & \text{if } \bullet = \text{true} \\ \diamond & \text{otherwise} \end{cases}$$

$$\llbracket Exp.Exp' \rrbracket^\star = \begin{cases} \langle \text{true}, \texttt{Unit.id}(\overline{Exp}) \rangle & \text{if } Exp.Exp' = \texttt{Unit.id.id}(\overline{Exp}) \\ \langle \text{true}, Exp \rangle & \text{if } Exp.Exp' = \texttt{Unit.id}(Exp) \\ \langle \text{false}, \texttt{Unit.id} \rangle & \text{if } Exp = \texttt{Unit} \text{ and } \llbracket .Exp' \rrbracket^\star = \langle \bullet, [blank] \rangle \\ \langle \bullet \vee \bullet', \diamond\ \diamond' \rangle & \text{otherwise, let } \llbracket Exp \rrbracket^\star = \langle \bullet, \diamond \rangle \\ & \ \text{and } \llbracket .Exp' \rrbracket^\star = \langle \bullet', \diamond' \rangle \end{cases}$$

$\llbracket .id \rrbracket^\star = \langle \text{false}, .id \rangle \qquad \text{let } \llbracket .Exp \rrbracket^\star = \langle \bullet, \diamond \rangle,\ \llbracket .id.Exp \rrbracket^\star = \langle \bullet, .id\ \diamond \rangle$

$$\text{let } \overline{\llbracket Exp \rrbracket^\star} = \overline{\langle \bullet, \diamond \rangle},\ \llbracket .id(\overline{Exp}) \rrbracket^\star = \begin{cases} \langle \bigvee \overline{\bullet}, .id(\overline{\diamond}) \rangle & \text{if } .id \neq \texttt{.id} \\ \langle \text{true}, [blank] \rangle & \text{if } \overline{\textsc{noop}(\diamond)} = \overline{[blank]} \\ \langle \bigvee \overline{\bullet} \vee |\overline{\diamond}| \neq |\overline{\star}|, \texttt{.id}(\star) \rangle & \text{otherwise, let } \overline{\textsc{noop}(\diamond)} = \overline{\star} \end{cases}$$

$\text{let } \llbracket .id\langle \overline{TE} \rangle(\overline{Exp}) \rrbracket^\star = \langle \bullet, \diamond \rangle \text{ and } \llbracket .Exp \rrbracket^\star = \langle \bullet', \diamond' \rangle,$

$\qquad \llbracket .id\langle \overline{TE} \rangle(\overline{Exp}).Exp \rrbracket^\star = \langle \bullet \vee \bullet', \diamond\ \diamond' \rangle$