Montana Smart Pointers:
They're Smart, and They're Pointers

Jennifer Hamilton

Department of Computer Science
University Of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

Technical Report CS-97-17

April 4, 1997

Abstract

The Montana C++ programming environment provides an API interface to the compiler, which allows the compilation process to be extended through programmer-supplied tools. This paper investigates the feasibility of that interface, using smart pointers as an example. Smart pointers are a powerful feature of the C++ language that enable a variety of applications, such as garbage collection, persistence, and distributed objects. However, while smart pointers can be used in much the same way as built-in pointers, they are not interchangeable. Using the Montana API, smart pointer functionality can be introduced for built-in pointers, thus enabling built-in pointers that act like smart pointers. We provide an overview of the Montana programming environment and describes how smart pointers can be implemented using the Montana API.

Acknowledgements

I would like to thank my supervisor, Dr. Jacob Slonim, and Dr. Gordon Cormack, for their time and energy spent in providing guidance, advice, and comments throughout the project. Financial assistance from the IBM Toronto Lab is gratefully acknowledged.

Contents

1. Introduction ... 1
2. The Montana C++ Programming Environment ... 1
3. Smart Pointers ... 2
   3.1 Usage Differences ... 2
      3.1.1 Class Hierarchies ... 3
         3.1.1.1 User-Defined Conversions ... 3
         3.1.1.2 Parallel Hierarchy ... 4
      3.1.2 Qualified Pointers ... 4
      3.1.3 Comparison to null ... 5
4. Making Built-In Pointers Smart ... 5
   4.1 A Reference Counting Smart Pointer ... 5
   4.2 Which Built-in Pointers Become Smart? ... 6
5. Montana CodeStore Architecture ... 6
   5.1 The Montana Incorporation Process ... 7
   5.2 Transformation ... 7
6. Implementing Smart Pointers with Montana ... 8
   6.1 Creating a Transformation Incorporation Extension ... 9
   6.2 Dynamically Loading a Transformation Incorporation Extension ... 9
   6.3 The SmartPointerTransformer class ... 12
   6.4 CS_SmartPointerTransformer::modify ... 12
   6.5 SmartPointerTransformerImplementation class ... 14
   6.6 ReferenceCounterTransformerImplementation class ... 16
      6.6.1 Transforming Pointer Assignments ... 17
      6.6.2 Transforming Initialization Expressions ... 18
7. An Example ... 18
8. Further Work ... 21
9. Conclusion ... 21
10. References ... 22
11. Special References ... 23
12. Appendix - Complete Implementation ... 24
   12.1 SmartPointer.h ... 24
   12.2 ReferenceCounter.h ... 26
   12.3 ReferenceCounterInterface.h ... 26
   12.4 SmartPointer.C ... 27
   12.5 ReferenceCounter.C ... 30

List of Figures

Figure 1 Simple Smart Pointer Class 2
Figure 2 Template Smart Pointer class 2
Figure 3 Class Hierarchies with Smart Pointers 3
Figure 4 Ambiguous Function Call 3
Figure 5 Member Template 4
Figure 6 Reference Counter Functions 5
Figure 7 Built-in Pointer Operations 6
Figure 8 `CS_Transformer` class 7
Figure 9 `CS_ImplementationComponentFactory` class 8
Figure 10 Relationship Between Classes 8
Figure 11 SmartPointerImplementationComponentFactory implementation 9
Figure 12 `SmartPointerImplementationComponentFactory` class 9
Figure 13 ice File for Smart Pointer Extension 10
Figure 14 Montana Configuration File 10
Figure 15 Extension Dynamic Load Point 10
Figure 16 `SmartPointer` class 10
Figure 17 `incorporationStartup` method 11
Figure 18 `SmartPointerTransformer` class 11
Figure 19 `SmartPointerTransformer transform` methods 12
Figure 20 `modifyAssignExpression` method 13
Figure 21 `modifyImplicitInitializer` method 13
Figure 22 `modifyExpressionInitializer` method 13
Figure 23 `modifyDestructorStateChangeExpression` methods 14
Figure 24 SmartPointerTransformerImplementation class 14
Figure 25 `transformerImplementation` method 15
Figure 26 ReferenceCounter class 15
Figure 27 ReferenceCounterTransformerImplementation class 16
Figure 28 modifyAssignExpression method 16
Figure 29 modifyExpressionInitializer method 17
Figure 30 Test Program Output 18
Figure 31 Test program 19
Figure 32 Transformed Expressions Without Smart Pointer Extension 19
Figure 33 Transformed Expressions With Extension 20

1. Introduction

The Montana[1] C++ programming environment is a joint development effort between IBM's Software Solutions and Research Divisions, and will be the base for a future release of IBM's VisualAge C++ product. Montana provides many unique features over traditional C++ compilers, most notably support for complete incremental compilation and an API interface [Nac96].

The purpose of this paper is to assess the feasibility of the Montana API interface for extending the compilation process to augment built-in language syntax. We have chosen the C++ smart pointer support as a basis of comparison. In this paper we present a partial smart pointer implementation using a Montana extension, where built-in pointer operations are modified as part of the compilation process, and summarize the results.

2. The Montana C++ Programming Environment

The Montana project grew from the recognition that current C++ development environments, while improving, were lacking in many areas, especially compared with those available for languages such as Smalltalk. One of the major frustrations in developing large C++ applications is the build turnaround time. The goal for Montana is to provide extremely fast incremental compilation, so that recompilation time required is proportional to the size of the change. In particular, changing a header files should not force recompilation of all files that happen to include it.

The design goal for the Montana architecture is that it can be extended in a variety of ways. A good example is the Montana object model[2] support. Most C++ compilers support a single native object model, the semantics for which are entrenched in the compiler itself, making it difficult to support different object, such as DirectToSOM C++ [Ham96] or other industry object models. Montana, however, was designed so that the object model is supported through a well-defined interface. A new object model can be added without requiring massive changes throughout the compiler. At the time of writing, the author was responsible for the design and development of such non-native object models.

Montana is designed around a system called *CodeStore* [BCCKLN94]. CodeStore consists of a C++ parser, a database that contains the compiled C++ program representation, and a class library that provides an API interface to the compilation process and program representation. Using this class library interface, C++-knowledgeable tools such as browsers can query the program representation of a compiled C++ program. In addition, CodeStore tools called *extensions* can be written that interact with the compilation process.

There are three types of extensions [Sor96]: 1) *CodeStore extensions*, which add data to the CodeStore and have incremental update capability, 2) *incorporation*[3] *extensions*, that modify or observe the incorporation process directly and 3) *user interface extensions*, which allow additional artifacts such as buttons and menus to be added to the user interface display. An example of the first type of extension is a separate compiler that is triggered as part of the compilation process to handle different file types, while an example of an incorporation extension is a tool that interacts directly with the compilation process itself, querying or updating the result. In this paper, we will concentrate on the second form of extension.

---

[1] The name "Montana" originated from an architecture meeting in which the idea of developing a new compiler with a clean slate was referred to as a "blue sky" approach. Since "blue sky" was thought to be the motto for the state of Montana (it's actually "big sky"), that became the name of the project. [Nac96]

[2] By object model, we mean issues such as how objects are laid out in memory and the strategy used to support virtual functions and bases. See [Lip96] for a detailed discussion.

[3] *Incorporation* is the Montana term for recompiling a program, in which the changes to the source will be incorporated into the CodeStore database.

3. Smart Pointers

Smart pointers are a powerful feature of the C++ language that enable a variety of applications, such as garbage collection, persistence, and distributed objects. They are used to augment the functionality of C++ pointer operations, allowing the programmer to perform additional work when pointers are created and used.

Smart pointers essentially allow a user-defined exit added to be pointer operations. A smart pointer [Stro89] itself is an instance of a class that wraps a built-in pointer, for which the dereference operator `->` has been overloaded. Such objects can be used in much the same way as a built-in pointer, but have additional functionality provided through *operator overloading*[4]. In much the same way as inheritance, smart pointers can be used in C++ to extend the functionality of a class. However, while inheritance extends the functionality of class instances themselves, smart pointers are used to extend the environment containing the instance. In other words, smart pointers are used to modify how the programming environment operates on an object, rather than how the object operates on itself. Smart pointers have a wide variety of uses, from simple applications such as detecting null dereferences, debugging, and read-only pointers [Alg95], to more complex applications such as garbage collection [GC96], [Ede92a], and persistence [Coh96].

```
#include <iostream.h>

struct S {
    int i;
};

class SP {
    S *_p;
  public:
    SP(S *p) : _p(p) {}
    S* operator->() {
        cout << "dereferencing" << endl;
        return _p;
    }
};

int main()
{
    SP sp(new S);

    sp->i = 10;        // sp.operator->()->i
}
```

Figure 1 Simple Smart Pointer Class

```
template<class T> class SP {
   T *_p;
  public:
    SP(T *p) : _p(p) {}
    T* operator->() {
        cout << "dereferencing" << endl;
        return _p;
    }
};

int main()
{
    SP<S> sp(new S);

    sp->i = 10;  // sp.operator->()->i
}
```

Figure 2 Template Smart Pointer class

Figure 1 shows a very simple definition of a smart pointer class. The `operator->()` function is a unary postfix operator which must be followed by a function or data member name. When applied to an object, the result is reapplied to the given member name. Thus, `sp->i` in the main function becomes `(sp.operator->())->i`. The expression `sp->i = 10` invokes the `SP::operator->` member function, returning the value of `_p` that will be reapplied and used in the actual dereferencing operation with the member `i`.

Smart pointers are more typically defined through a template class however, as shown in Figure 2, allowing multiple pointer types to be supported.

3.1 Usage Differences

In general, smart pointers can be used in exactly the same way as built-in pointers, however, as described in [Ede92b], there are some important differences between the two with respect to implicit type conversions performed by the compiler. These fall into two major categories: 1) class hierarchies and 2) types qualified with `const` or `volatile`. A further issue, described in [Mey96a] and [Mey96b], is testing for nullness.

[4] Using the keyword `operator`, the C++ language allows programmers to define class methods that are language operators, such as `+`.

When using built-in pointers, the compiler implicitly performs a variety of conversions between pointer types. Examples are `T*` to `const T*`, `Derived*` to `Base*`, and `T*` to `void*`. These implicit conversions are not directly supported by smart pointer types.

3.1.1 Class Hierarchies

The conversion from a derived class pointer to that of a base class is an important one for C++, as this is how polymorphism is supported in the language, using virtual functions. Even if the template class definition is used, there is no type relationship between the two pointers `SP<Base>` and `SP<Derived>`. Given the code in Figure 3, the first assignment in function `main` is valid because the compiler can implicitly convert a `Derived*` to a `Base*`. However, the second assignment is invalid, because there is no implicit conversion between `SP<Derived>` and `SP<Base>`.

Given that smart pointers are to be used in the same way as built-in pointers, an explicit cast conversion from `SP<Derived>` and `SP<Base>` would not suffice. What is necessary is to provide an implicit standard compiler or a user-defined conversion between the two types.

[Ede92b] describes two approaches to handling this problem. The first is to supply user-defined conversions between to all bases and the second is to define a parallel class hierarchy.

```
class Base {
};

class Derived : public Base {
};

int main()
{
    Base *baseP;
    Derived *derivedP;

    SP<Base> baseSP(new Base);
    SP<Derived> derivedSP(new Derived);

    baseP = derivedP;          // ok
    baseSP = derivedSP;        // error
}
```

Figure 3 Class Hierarchies with Smart Pointers

3.1.1.1 User-Defined Conversions

Note that because user-defined conversions are not implicitly chained by the compiler, conversions must be supplied to all bases, both direct and indirect. In other words, providing a conversion from class from `C` to its direct base `B` and from class `B` to its direct base `A` still does not provide a conversion from class `C` to its indirect base `A`. There are several drawbacks to the approach of providing user-defined conversions, the first being that for deep class hierarchies, providing such conversions can be quite tedious. The second is that the user-defined conversion does not provide the same semantics as a standard conversion, in which conversion to a direct base is preferred over conversion to an indirect base.

For example, consider the class hierarchy shown in Figure 4. If conversions were defined from `SP<Derived2>` to both `SP<Base>` and `SP<Derived>`, then the call to function `foo` in the `main` function would be ambiguous because the call matches both functions. With built-in pointers, the call would not be ambiguous because a conversion to a direct base class is preferred over a conversion to an indirect base.

```
class Base {};
class Derived : public Base {};
class Derived2 : public Derived {};

void foo(SP<Base>);
void foo(SP<Derived>);

int main()
{
    SP<Derived2> derived2SP(new Derived2);

    foo(derived2SP);
}
```

Figure 4 Ambiguous Function Call

A third problem with this approach is that one must specialize the smart pointer template class in order to add the conversions, which diminishes the value of using a template in the first place [Mey96a].

A relatively recent addition to the ANSI standard, *member function templates* [Stro94], [Mey96a], provides a partial

solution to these problems. A member function template allows the programmer to declare a member function that has formal template arguments, which provides a means for converting between smart pointer types.

Figure 5 shows a simple member template declaration. If the compiler needs to convert an `SP<Derived>` object to an `SP<Base>` object, it will first look for any user-defined operators that support the desired conversion. If none are found, the compiler will then look for a member template function that would support the conversion through instantiation. This is provided by the `template<class T2> operator SP<T2>()` member function template shown above. The member template can successfully be instantiated only if the statement `return SP<T2>(_p);` can be validly instantiated as `return SP<Base>(_p);`

```
template<class T> class SP {
    T *_p;
  public:
    SP(T *p) : _p(p) {}
    T* operator->() {
        cout << "dereferencing" << endl;
        return _p;
    }
    template<class T2> operator SP<T2>()
        { return SP<T2>(_p); }
};
```

Figure 5 Member Template

This is correct because `_p` is of type `Derived *`, which can be implicitly converted to type `Base *`. In other words, the member function template defined above provides a conversion from `SP<Type1>` to `SP<Type2>` for any `Type1 *` that can be implicitly converted to a `Type2 *`.

Using member function templates to implement smart pointers solves two of the three problems associated with providing user-defined conversions, namely the need to explicitly define every conversion and to specialize the template class to do so. However, it still suffers from the same problem in that the semantics of a smart pointer conversion do not match those of a built-in pointer conversion, such as preference to convert to a direct base class. In addition, relatively few compilers support member function templates at this point.

3.1.1.2 Parallel Hierarchy

The second means of dealing with the base class conversion problems is to define a hierarchy of smart pointers that parallels that of the original hierarchy. The first issue that arises with this approach is where the wrapped built-in pointer is defined. If a pointer is defined in each class in the parallel hierarchy, there is first of all a significant storage overhead. Secondly, care must be taken when an assignment is made to a smart pointer to update all pointers in any base classes so that proper base class conversions can take place.

Rather than duplicating the wrapped built-in pointer in each base class, the underlying pointer could be declared in an abstract virtual base class. This would avoid the storage overhead by having the pointer stored only once. However, this approach suffers from conversion problems with base classes at non-zero offsets.

3.1.2 Qualified Pointers

Another problem that occurs when dealing with smart pointers is in handling `const` and `volatile` qualifications. With built-in pointers, both the pointer and the referent can be qualified. For example, `const T*` means that the referent is `const`, whereas `T* const` means that the pointer itself is `const`, and `const T* const` means that both the referent and the pointer are `const`. The C++ language supports implicit conversions from non-`const` to `const` types. For example, a `T*` can be assigned to a `const T*` or a `T* const`.

Smart pointers are restricted to being declared only as `const SP<T>`, which makes the pointer itself const. The only way to declare a smart pointer so that object pointed to is const is by including const with the type, as in `SP<const T>`. This however suffers from the same problem as the class hierarchies in that `SP<T>` and `SP<const T>` are considered unrelated types by the compiler. Thus an `SP<T>` cannot be implicitly converted to an `SP<const T>`. If the compiler supports it, the member function template described earlier will handle this conversion, otherwise an explicit user-defined conversion function must be supplied.

Volatile-qualified pointers suffer from the same issues and restrictions.

3.1.3 Comparison to null

When dealing with pointers in C++, invariably one needs to determine if the pointer is null. With built-in pointers, one can simply compare the value to 0, in which case the compiler implicitly converts the constant expression 0 to the null pointer for comparison purposes. Thus one can write `if (ptr)` or `if (!ptr)` to test for `ptr` being null. However, with smart pointers, there is no implicit conversion to the null pointer, so such expressions cannot be used. This can be alleviated by providing a function to test for nullness, however this would be contrary to standard programming idioms. Simply providing a conversion to `void *` for a smart pointer will solve the problem, however it introduces new problems in that mixed-type conversions would be allowed that would not be valid for built-in pointers. For example, given two unrelated types `Type1` and `Type2`, a `Type1 *` cannot be compared to a `Type2 *`. However, if the `SP` class provided an implicit conversion to null, then an `SP<Type1>` could be compared to an `SP<Type2>` or to any other pointer type, such as `int *`. One solution to this, proposed in [Mey96a], is to overload operator! to return 1 if the smart pointer is null, and 0 otherwise. This allows for tests such as `if (!ptr)`, however, it still doesn't allow `if (ptr)`. Two additional approaches to handling this problem, not without their own drawbacks, can be found in [Mey 96b].

4. Making Built-In Pointers Smart

In the previous section, we described the differences between smart and built-in pointers and the problems that result from these differences. If, however, the "smarts" of a smart pointer could be added to a built-in pointer, these problems would be alleviated. In this section, we describe the changes that would be needed to built-in pointer expressions in order that they operate as smart pointers. For the purpose of this example, we will implement a reference counting smart pointer. In subsequent sections, we will describe how to implement this model using a Montana incorporation extension.

4.1 A Reference Counting Smart Pointer

The basic model for a reference-counting smart pointer is as follows:

1) Whenever a new reference is made to a given object, the reference count for that object should be incremented.
2) If a reference to an object is removed, the reference count for that object should be decremented. If the reference count for an object goes to zero, delete the object.

These rules are illustrated by the functions `increment` and `decrement` shown in Figure 6.

In order to add reference counting smart pointer functionality to built-in pointers operations, the following expression transformations are required:

<u>Pointer assignment:</u> Whenever an assignment is made to a designated smart built-in pointer, the reference count for the object originally pointed to should be decremented and that of the object now pointed to should be incremented. Thus, the expression `p1 = p2` becomes:

```
(p1 == p2 ? 0 : decrement(p1), p1 =
p2, increment(p1), p1)
```

```
void decrement(ReferenceCounter *sp)
{
    if (!sp)
        return;
    if (! --sp->rc)
        delete sp;
}

void increment(ReferenceCounter *sp)
{
    if (!sp)
        return;
    ++sp->rc;
}
```

Figure 6 Reference Counter Functions

<u>Pointer initialization:</u> A designated smart built-in pointer must always be either explicitly initialized to a value, or to zero. (If a pointer were not initialized to zero and contained

non-zero garbage, a subsequent assignment to that pointer using the previous expression would likely result in an exception).

The statement `SPC* p1;` becomes:

```
SPC* p1 = 0;
```

and `SPC** p1 = new SPC*;` becomes:

```
SPC** p1=new SPC*; p1 ? *p1=0 : 0;
```

If a smart pointer is initialized to a value, the reference count for the underlying object must be incremented. So the statement `SPC* p1 = p2;` becomes:

```
SPC* p1 = p2; increment(p1);
```

Object Initialization: When a designated smart built-in pointer object is created, the reference count must be initialized. For dynamically-created objects, the count should be initialized to 0, and for static or automatic objects, the reference count should be initialized to 1 so that the object can be used in reference counting contexts, but will never be deleted.

Pointer destruction: When a designated smart built-in pointer is destroyed the reference count for the referenced object must be decremented. There are several ways that a smart pointer will be destroyed, the most common being that it goes out of scope. Other possibilities are that a dynamically allocated smart pointer is deleted, or an exception occurs in which the containing block is unwound from the stack. Only the deletion of a dynamically-allocated smart pointer consists of an expression that can be transformed. The other two require modifications to the function itself so that the scope termination and exception handling code will include the decrement of any smart pointers declared therein.

4.2 Which Built-in Pointers Become Smart?

The above discussion raises the question of how to determine which built-in pointer operations should be transformed into smart pointer operations. One could blindly apply the transformation to all built-in pointer operations, but this would certainly be overkill. Rather, we would like to select only specific pointers for the transformation The approach that we have chosen is to define a special base class, `ReferenceCounter`. Expressions involving objects declared of, or pointers to, a class derived from `ReferenceCounter` will be transformed as described above.

For example, consider the built-in pointers declared of type `C*` in Figure 7. Because class `C` is derived from ReferenceCounter, several transformations should take place. `cp1` and `cp2` should be implicitly initialized to 0 at the point of declaration, and the assignment from `cp2` to `cp1` should be transformed as described earlier.

```
class C : public ReferenceCounter {};

int main()
{
    C *cp1, *cp2;

    cp1 = cp2;
}
```

Figure 7 Built-in Pointer Operations

5. Montana CodeStore Architecture

The previous section described the necessary transformations to built-in pointer operations in order to implement a reference counting smart pointer. This section provides an overview of the Montana CodeStore architecture and describes in a generic sense how a transformation extension can be added. This will form the basis for the remainder of the paper, which describes our implementation of smart built-in pointers in Montana using an incorporation extension.

5.1 The Montana Incorporation Process

As part of the Montana incremental compilation process, the compiler separates a source file into *regions*, where each region consists of approximately one declaration. If a region, or something that region depends upon, has changed since the last incorporation, it is re-incorporated. Re-incorporation involves a number of standard steps: parsing, semantic analysis, transformation, error checking, code generation, and incremental linking. In addition, dependency arcs are added between CodeStore elements so that a change in one region can trigger a re-incorporation of a dependent region. For example, a region containing a derived class declaration will have a dependency on each region containing one of its base classes.

The Montana class `CS_CodeStore` is used to represent the underlying CodeStore database. This class supports a variety of routines to create, query and update the CodeStore. An application that operates on a CodeStore will contain exactly one instance of the `CS_CodeStore` class. If an incorporation is currently taking place against the database, this `CS_CodeStore` instance will contain a reference to an object of type `CS_IncorporationState`, which represents the current state of the incorporation.

5.2 Transformation

The transformation step involves simplifying expressions and statements into a C-like representation. In Montana, it is implemented through the class `CS_Transformer`, which is shown in Figure 8. `CS_Transformer` has three versions of the method `transform`, corresponding to the different types of transformations that are supported. Most calls to the `CS_Transformer::transform` methods are for statements or initializations. Expression transformations typically occur as part of the transformation of their containing statement.

When the compiler needs to transform an item, it obtains a transformer object from the CodeStore's incorporation state object. The incorporation state in turn retrieves the transformer from an implementation component factory (see Figure 9). The incorporation state maintains a list of implementation component factories, and selects the transformer returned by the front element in the list.

```
class CS_Transformer : public CS_IncorporationComponentBase<CS_DepthFirstModifier>
{
public:
   CS_Transformer(CS_IncorporationState& s) :
CS_IncorporationComponentBase<CS_DepthFirstModifier>(s) { }

   // Transform a statement tree
   //
   virtual CS_bool transform(CS_Statement*& stmt, CS_bool emitMessages)
      { modifyStatement(*stmt); return CS_true; }

   // Transform a variable initializer
   //
   virtual CS_bool
         transform(CS_Initializer*& init, CS_VariableDeclaration& var, CS_bool emitMessages)
      { init = &modifyInitializer(*init, &var.typeDescriptor(), &var); return CS_true; }

   // Transform an expression tree
   //
   virtual CS_Expression& transform(CS_Expression& expr, CS_bool emitMessages)
      { return modifyExpression(expr); }
};
```

Figure 8 `CS_Transformer`

```
class CS_ImplementationComponentFactory : public CS_Link<CS_ImplementationComponentFactory>
{
public:
   virtual CS_TypeAnalyzer&   typeAnalyzer()  { return next()->typeAnalyzer();  }
   virtual CS_Diagnostician&  diagnostician() { return next()->diagnostician(); }
   virtual CS_Optimizer&      optimizer()     { return next()->optimizer();     }
   virtual CS_Transformer&    transformer()   { return next()->transformer();   }
};
```

Figure 9 `CS_ImplementationComponentFactory` class

An implementation component can be one of four types: a type analyzer, a diagnostician, a transformer, or an optimizer. Each of these implementation components take part in a specific portion of the incorporation process, and can be overridden to modify the compilation process. Applications can provide custom implementation components by subclassing the implementation component factory class and providing overrides for the methods of interest. By inserting this new class at the front of the incorporation state list, the incorporation state will select the overridden component provided.

For example, to provide a transformer (incorporation) extension, the class `CS_ImplementationComponentFactory` would be derived from, supplying a transformer method that would return the custom transformer object. The incorporation state method `prependImplementation-ComponentFactory` would be called to add this factory to the front of the list. Any factory methods that are not overridden would return the result of invoking that method against the next factory in the list, as shown in Figure 9 with the method invocation against the result of the `next` method.

Montana supplies a default implementation component factory that provides the standard implementations for each component. When no extensions have been introduced, this factory will be at the front of the incorporation state's list. Figure 10 shows the relationship between the various classes discussed in this section.

6. Implementing Smart Pointers with Montana

In this section, we will present our implementation of smart pointers through built-in pointers using a Montana transformation incorporation extension. The complete implementation is included in the Appendix A, and we have extracted specific pieces to clarify the explanation. We will first describe how to add our specific transformation extension, and then present our implementation for reference counting smart pointers based on the required expression transformations discussed earlier.

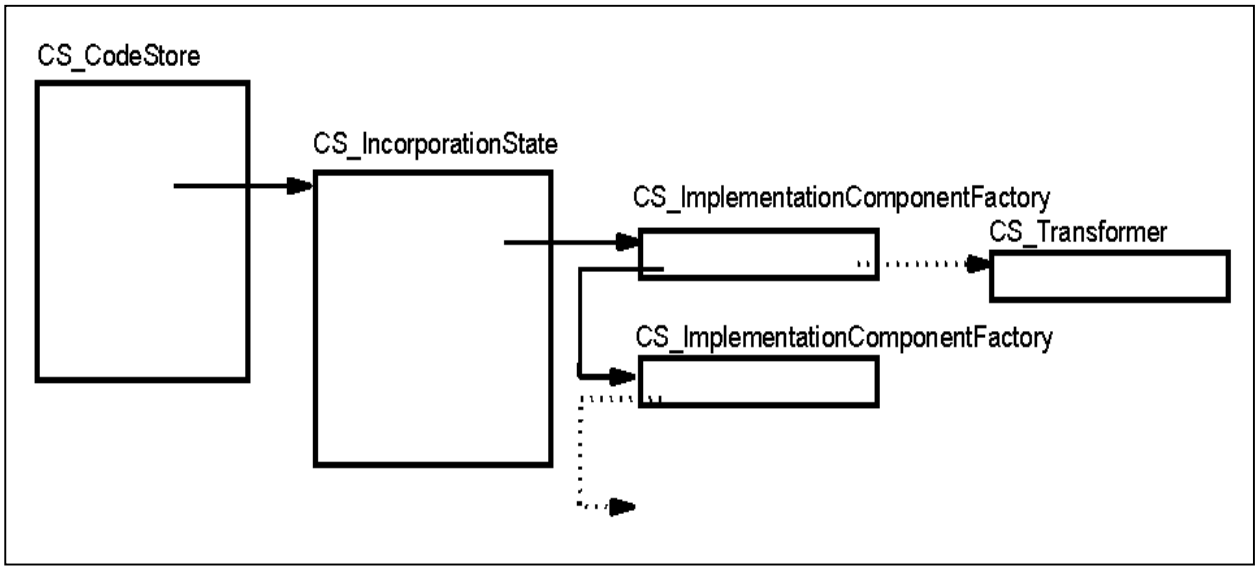

Figure 10 Relationship Between Classes

6.1 Creating a Transformation Incorporation Extension

As described in the previous section, in order to implement a transformation extension, we must create a subclass of the `CS_ImplementationComponentFactory` class and insert an object of this new type at the front of the incorporation state's factory list. Figure 12 shows the definition of the class `SmartPointerImplementationComponentFactory` and Figure 11 shows the corresponding implementation. The method transformer is overridden to return our custom smart pointer transformer extension. The constructor for the factory initializes the `_transformer` member by creating a new object of class `SmartPointerTransformer`. This latter class will implement our smart pointer transformer extension, and will be discussed in more detail subsequently. Note that the transformer extension constructor is passed the incorporation state and the current transformer object, obtained from the front of the factory list.

6.2 Dynamically Loading a Transformation Incorporation Extension

We now have a factory implementation that will return our custom transformation extension. The next issue to address is how the factory object will be created and added to the front of the incorporation state's factory list. This will be achieved by loading a dynamic link library (DLL) that contains a static variable whose initialization will cause the factory to be created and inserted into the list. Then the question is, how is the DLL loaded? We will now examine the Montana support for defining and loading extensions.

```
SmartPointerImplementationComponentFactory::
        SmartPointerImplementationComponentFactory(CS_IncorporationState& s) :
    _state(s),
    // The second argument to this constructor comes from pull on the
    // chain of components stored in the IncorporationState.
    //
    _transformer(new SmartPointerTransformer(_state,
                 _state.implementationComponentFactory().transformer()))
{
    assume(_transformer);
}

CS_Transformer& SmartPointerImplementationComponentFactory::
        transformer()
{
    assume(_transformer);
    return *_transformer;
}

SmartPointerImplementationComponentFactory::
        ~SmartPointerImplementationComponentFactory()
{
    delete _transformer;
}
```

Figure 11 SmartPointerImplementationComponentFactory implementation

```
class SmartPointerImplementationComponentFactory
   : public CS_ImplementationComponentFactory {
  public:
    SmartPointerImplementationComponentFactory(CS_IncorporationState&);
    virtual CS_Transformer& transformer();
    virtual ~SmartPointerImplementationComponentFactory();

  private:
    CS_IncorporationState& _state;
    CS_Transformer* _transformer;
};
```

Figure 12 `SmartPointerImplementationComponentFactory` class

Externally, a Montana extension is introduced using an *Incremental C++ Extension*, or *ice*, file. Montana searches for and applies ice files at load time according to a defined search order. Figure 13 shows an ice file which defines an extension called `SmartPointer`, for which the corresponding DLL to load is `smartp.dll`. The suffix and prefix information is used to associate a specific file type with a given extension. This interface is for CodeStore extensions, but is used as a temporary measure until the final interface for incorporation extensions is defined.

```
[SmartPointer]

type=extension
description=Smart Pointer
Extension
dll=smartp.dll
suffixes=sp SP
prefix=dummy
```

Figure 13 ice File for Smart Pointer Extension

Montana programs are compiled by providing a *configuration file* which supplies the various options for the compilation. For our purposes, the configuration file shown in Figure 14 is used.

```
source type(cpp) src0 = "t.cpp"
target "t.exe" { source src0 }

source type(sp) src1 = "dummy.sp"
```

Figure 14 Montana Configuration File

This configuration file indicates that the source file to be compiled is `t.cpp`, the target executable will be `t.exe`, and that an additional source file called `dummy.sp` will also be processed. This latter source file, being an unsupported file type, will cause Montana to search the ice files for an appropriate extension that handles this file type, and load the extension DLL `smartp.dll`.

The next step is to register an *extension dynamic load point* using a statically-defined variable in the `smartp.dll` extension DLL as shown in Figure 15. An extension dynamic load point is used to register an extension with the compiler. For an incorporation extension, the final parameter to the extension dynamic load point constructor, the *incorporation startup function pointer* is most important. This function will be run at the start of every incorporation and can be used by an extension to plug in components into the incorporation state.

```
CS_ExtensionDynamicLoadPoint

SmartPointer::extension_load_point(
        SmartPointer::className(),
        SmartPointer::update,
        SmartPointer::isChanged,

SmartPointer::processOptions,
        EXTENSION_PRIORITY,

SmartPointer::incorporationStartup);
```

Figure 15 Extension Dynamic Load Point

```
class SmartPointer : public CS_InterfaceBase
{
public:
   static const char* className();

   static void SmartPointer::incorporationStartup(
      CS_ExtensionDynamicLoadPointLink&, CS_IncorporationState&);

   static CS_DependencyNode::UpdateResult
        update(CS_ExtensionSource* me, CS_IncorporationState& state,
               CS_bool emitMessages);

   static CS_bool isChanged(CS_ExtensionSource* me);

   static void processOptions(CS_ExtensionSource* me, CS_OptionList& options);

private:
   static CS_ExtensionDynamicLoadPoint  extension_load_point;
};
```

Figure 16 `SmartPointer` class

```
void SmartPointer::incorporationStartup(
        CS_ExtensionDynamicLoadPointLink&, CS_IncorporationState& state)
{
    cout << __FUNCTION__ << endl;

    SmartPointerImplementationComponentFactory* fac =
        new SmartPointerImplementationComponentFactory(state);
    assume(fac);  // (our version of "assert")

    // Push our new factory with its new Transformer onto the chain
    // stored in the IncorporationState.
    //
    state.prependImplementationComponentFactory(*fac);

    return;
}
```

Figure 17 `incorporationStartup` method

The main effect then, of constructing the static member variable `SmartPointer::extension_load_point` is that the method `SmartPointer::incorporationStartup` will be called prior to each incorporation. The class `SmartPointer` and the `incorporationStartup` method are shown in Figure 16 and Figure 17. In the `incorporationStartup` method the newly-created `SmartPointerImplementationComponentFactory` object is added to the front of the incorporation state's factory list (see Figure 10). Adding this custom factory object to the front of the queue will cause any requests made of the incorporation state for a transformer object to return the custom transformer, `SmartPointerTransformer`.

```
class SmartPointerTransformer : public CS_Transformer {
  public:
    SmartPointerTransformer(CS_IncorporationState& s, CS_Transformer& p)
       : CS_Transformer(s), _parent(p),
         _referenceCounterTransformerImplementation(0) {
    }

    ~SmartPointerTransformer();

    virtual CS_bool transform(CS_Statement*& stmt, CS_bool emitMessages);
    virtual CS_bool transform(CS_Initializer*&, CS_VariableDeclaration&, CS_bool);
    virtual CS_Expression& transform(CS_Expression&, CS_bool);

    virtual CS_Expression& modifyAssignExpression(CS_BinaryExpression&);
    virtual CS_Initializer& modifyExpressionInitializer(
        CS_ExpressionInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);
    virtual CS_Initializer& modifyImplicitInitializer(
        CS_ImplicitInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);
    virtual CS_Expression& modifyDestructorStateChangeExpression(
        CS_DestructorStateChangeExpression&);

    CS_Expression& typeAnalyze(CS_Expression&);
    CS_Initializer& typeAnalyze(CS_Initializer&);

  private:
    CS_Transformer& _parent;

    // classes for each smart pointer implementation
    ReferenceCounterTransformerImplementation* _referenceCounterTransformerImplementation;

    // Return the smart pointer implementation, if any, for the expression
    // The expression must be a pointer to a class derived from a SmartPointer class
    SmartPointerTransformerImplementation* transformerImplementation(CS_TypeDescriptor&);
};
```

Figure 18 `SmartPointerTransformer` class

6.3 The SmartPointerTransformer class

At this point, whenever a transformation takes place, the `SmartPointerTransformer` class (see Figure 18) will have control. One of the three overridden `transform` methods shown at the beginning of the class will be called depending upon the type of transformation taking place: a statement, initialization, or an expression. The overridden versions of the `SmartPointerTransformer` methods are shown in Figure 19. These `transform` methods have fairly standard implementations. They first call an appropriate `modify` method, and then invoke the `transform` method of the previous element in the component chain (given by member variable `_parent.`)

```
CS_bool SmartPointerTransformer::transform(CS_Statement*& stmt, CS_bool emitMessages)
{
    modifyStatement(*stmt);
    _parent.transform(stmt, emitMessages);
    return CS_true;
}

CS_bool SmartPointerTransformer::
        transform(CS_Initializer*& init, CS_VariableDeclaration& var, CS_bool emitMessages)
{
    init = &modifyInitializer(*init, &var.typeDescriptor(), &var);
    _parent.transform(init, var, emitMessages);
    return CS_true;
}

CS_Expression& SmartPointerTransformer::transform(CS_Expression& expr, CS_bool emitMessages)
{
    CS_Expression *expr2 = &modifyExpression(expr);
    return _parent.transform(*expr2, emitMessages);
}
```

Figure 19 `SmartPointerTransformer transform` methods

Recall that the constructor for the `SmartPointerTransformer` class is passed the current transformer object, which is used to initialize the data member `_parent`. Calling the parent `transform` method allows the standard compiler transformations to take place after the extension has been run.

It is when the `modify` method is called that the transformer extension has an opportunity to modify the transformed expression. The compiler-supplied `modify` methods step through the underlying item and calls an appropriate `modifyxxx` method for each entity encountered. By overriding methods corresponding to expression of interest, the transformer extension can modify these expressions. In this case, we have overloaded `modifyAssignExpression`, `modifyExpressionInitializer`, `modifyImplicitInitializer`, and `modifyDestructorStateChangeExpression` (these methods will be explained in more detail in the next section). If the underlying expression or statement corresponds to one of these four, the overloaded method will be called. Each of these methods determines if any further expression transformation is necessary, based on the type of the object being operated on. If so, the expression is transformed according to the model described earlier for transforming built-in pointer operations into smart pointer operations.

6.4 CS_SmartPointerTransformer::modify

Now we will discuss the implementation of the `CS_SmartPointerTransformer::modify` methods. Each of these methods uses the `SmartPointerTransformer::transformerImplementation` method to determine if the current expression or statement deals with an object of interest. This method returns a `SmartPointerTransformationImplementation` that will perform implementation-specific transformations, depending upon the smart pointer type. We will discuss this latter class in the next section.

The `modifyAssignExpression` method is shown in Figure 20. For our smart pointer implementation, we want to detect any pointers assignments where the underlying type is derived from a special base class such as

`ReferenceCounter.modifyAssignExpression` is passed a reference to a `CS_BinaryExpression` object, representing the assignment expression currently being transformed. If the type descriptor for the lhs of this expression, given by `CS_BinaryExpression:: expression1()`, is not a pointer, then no further work is necessary. If it is a pointer, then the type to which it points, given by `CS_TypeDescriptor::next()` is passed to `transformerImplementation` to check if it points to an object derived from one of the special base classes. If a non-null value is returned, the `modifyAssignExpression` of the returned implementation object is called to modify the expression.

The `modifyExpressionInitializer` and `modifyImplicitInitializer` methods (Figure 21 and Figure 22) perform similar functions, but must handle both pointer and non-pointer initializations. These methods call the `transformerImplementation` with the `CS_TypeDescriptor` for the non-pointer variable being initialized. If the variable being initialized is a pointer, the `CS_TypeDescriptor` for the type pointed to, given by `td->next()`, is passed instead.

```
CS_Expression& SmartPointerTransformer::modifyAssignExpression(CS_BinaryExpression& binary)
{
    if (! binary.expression1().typeDescriptor().isPointer())
        return binary;

    SmartPointerTransformerImplementation *ti =
        transformerImplementation(*binary.expression1().typeDescriptor().next());

    return ti ? typeAnalyze(ti->modifyAssignExpression(binary)) : binary;
}
```

Figure 20 `modifyAssignExpression` method

```
CS_Initializer& SmartPointerTransformer::modifyExpressionInitializer(
        CS_ExpressionInitializer& init, CS_TypeDescriptor* td, CS_VariableDeclaration* var)
{
    if (!td)
        return init;

    // need to handle both pointer and object initialization
    SmartPointerTransformerImplementation *ti =
        transformerImplementation(td->isPointer() ? *td->next() : *td);

    if (! ti)
        return init;

    return typeAnalyze(ti->modifyExpressionInitializer(init, td, var));
}
```

Figure 21 `modifyExpressionInitializer` method

```
CS_Initializer& SmartPointerTransformer::modifyImplicitInitializer(
        CS_ImplicitInitializer& init, CS_TypeDescriptor* td, CS_VariableDeclaration* var)
{
    if (!td)
        return init;

    // need to handle both pointer and object initialization
    SmartPointerTransformerImplementation *ti =
        transformerImplementation(td->isPointer() ? *td->next() : *td);

    if (! ti)
        return init;

    return typeAnalyze(ti->modifyImplicitInitializer(init, td, var));
}
```

Figure 22 `modifyImplicitInitializer` method

```
CS_Expression& SmartPointerTransformer::
        modifyDestructorStateChangeExpression(CS_DestructorStateChangeExpression& dsce)
{
    return SmartPointerTransformerImplementation::
        modifyDestructorStateChangeExpression(dsce);
}

CS_Expression& SmartPointerTransformerImplementation::
    modifyDestructorStateChangeExpression(CS_DestructorStateChangeExpression& dsce)
{
    // save most recent state table entry
    if (dsce.tableEntry() && dsce.tableEntry()->asDestructorStateTableEntry())
        _currentDestructorStateTableEntry = dsce.tableEntry()->asDestructorStateTableEntry();

    return dsce;
}
```

Figure 23 `modifyDestructorStateChangeExpression` methods

The `modifyDestructorStateChangeExpression` method, shown in Figure 23, does not actually perform any modification against the given expression. Rather, it calls the static method `SmartPointerTransformationImplementation::modifyDestructorStateChangeExpression`, which simply saves the most recently seen state table entry if that entry is for a destructor. This value will be used later to add information to the state table in the appropriate location.

6.5 SmartPointerTransformerImplementation class

As discussed earlier, the model for our smart pointer implementation is that a built-in pointer will be transformed into a smart pointer if the underlying type inherits from a special base class. In order to provide multiple smart pointer transformations, we have defined a common base class called `SmartPointerTransformerImplementation` (Figure 24), from which a derived class will be defined for each smart pointer implementation. This derived class will handle the transformations specific to that smart pointer implementation.

```
class SmartPointerTransformerImplementation : public
CS_IncorporationComponentBase<CS_InterfaceBase> {
  public:
    SmartPointerTransformerImplementation(CS_IncorporationState &state,
                                          SmartPointerTransformer &transformer) :
        CS_IncorporationComponentBase<CS_InterfaceBase>(state), _transformer(transformer) {}

    virtual CS_Expression& modifyAssignExpression(CS_BinaryExpression&) = 0;
    virtual CS_Initializer& modifyExpressionInitializer(
        CS_ExpressionInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*) = 0;
    virtual CS_Initializer& modifyImplicitInitializer(
        CS_ImplicitInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*) = 0;

    static CS_Expression& modifyDestructorStateChangeExpression(
        CS_DestructorStateChangeExpression& dsce);
    CS_DestructorStateTableEntry* currentDestructorStateTableEntry()
        { return _currentDestructorStateTableEntry; }
    void currentDestructorStateTableEntry(CS_DestructorStateTableEntry *ste)
        { _currentDestructorStateTableEntry=ste; }

    SmartPointerTransformer& transformer() { return _transformer; }

  private:

    SmartPointerTransformer& _transformer;
    static CS_DestructorStateTableEntry *_currentDestructorStateTableEntry;
};
```

Figure 24 SmartPointerTransformerImplementation class

```
SmartPointerTransformerImplementation *SmartPointerTransformer::
        transformerImplementation(CS_TypeDescriptor& td)
{
    if (! td.isNamedType() ||
            ! td.declaration().declarationKind() == CS_Declaration::IsClass)
        return NULL;
    CS_ClassDeclaration &decl = *td.declaration().asClassDeclaration();

    CS_ClassDeclaration *referenceCounter =
        ReferenceCounterTransformerImplementation(state(), *this).findClassDeclaration();
    if (referenceCounter &&
            CS_ANSI_Queries::isBaseClassOf(*referenceCounter, decl)) {
        cout << "got a pointer to class derived from "
             << referenceCounter->signature() << endl;
        if (! _referenceCounterTransformerImplementation) {
            _referenceCounterTransformerImplementation =
                new ReferenceCounterTransformerImplementation(state(), *this);
        }
        return _referenceCounterTransformerImplementation;
    }
    // insert code to look for other smart pointer class implementations here
    return NULL;
}
```

Figure 25 `transformerImplementation` method

When a method needs to determine if a smart pointer implementation applies to a given expression, it calls the method `SmartPointerTransformer::transformerImplementation`, shown in Figure 25. The `transformerImplementation` method is passed a reference to an object of type `CS_TypeDescriptor`, which describes the type of the object being operated on. If the object is of a class type, a reference to the associated `CS_ClassDeclaration` is assigned to the variable `decl`. A `CS_ClassDeclaration` provides complete information about a class declaration, such as the class name, members. etc. So at this point, `decl` will reference the class declaration for the object of interest

Next, the `findClassDeclaration` method of the `ReferenceCounterTransformerImplementation` class is invoked. This method returns a pointer to the `CS_ClassDeclaration` object representing the class `ReferenceCounter,` if that declaration has been encountered in the program, and null otherwise. If non-null is returned, the method uses the `CS_ANSI_Queries::isBaseClassOf` method to determine if the class declaration for the object of interest is derived from `ReferenceCounter`. The `CS_ANSI_Queries` class provides a variety of functions that support querying of class declarations. If `ReferenceCounter` is a base class, the `_referenceCounterTransformerImplementation` data member will be initialized with a new `ReferenceCounterTranformerImplementation` object if it has not yet been initialized.

In the current implementation, we have defined one special base class, `ReferenceCounter`, and one corresponding specialization of `SmartPointerTransformerImplementation`. The programmer would include the declaration of `ReferenceCounter` class (see Figure 26) and derive from it to introduce reference-counting functionality for pointer operations against objects of that derived class. (The `name` member in `ReferenceCounter` is used for debugging purposes and will be discussed later). Additional smart pointer implementations could be introduced by inserting code at the end of the `transformerImplementation` method where indicated.

```
class ReferenceCounter {
  private:
    int rc;
    char *_name;

    static void dtor(ReferenceCounter **sp, int);
    static void decrement(ReferenceCounter *sp);
    static void increment(ReferenceCounter *sp);

  public:
    ReferenceCounter(char *name) : _name(name) { rc = 0; }
    char *name() { return _name; }
    virtual ~ReferenceCounter();
};
```

Figure 26 ReferenceCounter class

6.6 ReferenceCounterTransformerImplementation class

The ReferenceCounterTranformerImplementation (see Figure 27) class provides the transformer extension implementation for a reference counting smart pointer. It is a specialization of the `SmartPointerTransformerImplementation`, and contains overrides of the `SmartPointerTransformerImplementation::modify` methods, along with additional methods specific to implementing a reference counting smart pointer.

```
class ReferenceCounterTransformerImplementation :
        public SmartPointerTransformerImplementation {
  public:
    ReferenceCounterTransformerImplementation(
        CS_IncorporationState& state, SmartPointerTransformer& transformer) :
        SmartPointerTransformerImplementation(state, transformer) {};

    CS_ClassDeclaration* findClassDeclaration();

    CS_Expression& modifyAssignExpression(CS_BinaryExpression& binary);
    CS_Initializer& modifyExpressionInitializer(
        CS_ExpressionInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);
    CS_Initializer& modifyImplicitInitializer(
        CS_ImplicitInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);

    private:
        const CS_Atom& referenceCountMember();
        virtual CS_Expression& decrementReferenceCounterExpression(CS_Expression &);
        virtual CS_Expression& incrementReferenceCounterExpression(CS_Expression &);
        virtual CS_Expression& createStateChangeExpression(
            CS_Expression&, CS_VariableDeclaration&);
        virtual CS_FunctionDeclaration& findOrCreateDecrement();
        virtual CS_FunctionDeclaration& findOrCreateIncrement();
        virtual CS_FunctionDeclaration& findOrCreateDtor();
        virtual CS_FunctionDeclaration& findOrCreateMemberFunction(char *);
        virtual CS_DestructorStateTableEntry&createDestructorStateTableEntry(
            CS_VariableDeclaration&, CS_TreeNode&);
        virtual void addDestructorCalls(
            CS_VariableDeclaration&, CS_DestructorStateTableEntry&, CS_TokenLocation&);
};
```

Figure 27 ReferenceCounterTransformerImplementation class

```
CS_Expression&
ReferenceCounterTransformerImplementation::modifyAssignExpression(
    CS_BinaryExpression &binary)
{
    // Don't transform expressions on temporaries
    CS_Expression &e = binary.expression1();
    if (e.expressionKind() == CS_Expression::IsName &&
            e.asNameExpression()->name().declaration() &&
            ! e.asNameExpression()->name().declaration()->mapsToASourceLocation()) {
        return binary;
    }

    CS_TokenLocation loc =
        binary.sourceLocation().sourceRegion()->tokenLocation();

    CS_Expression &expr =
        ef().createCommaExpression(loc,
            decrementReferenceCounterExpression(binary.expression1()),
            ef().createCommaExpression(loc,
                binary,
                incrementReferenceCounterExpression(binary.expression1())));
    return expr;

}
```

Figure 28 modifyAssignExpression method

6.6.1 Transforming Pointer Assignments

As discussed earlier, when an assignment to a reference-counting smart pointer occurs, we want to transform an expression such as `x = y`, where x points to a type derived from `ReferenceCounter,` to the following:

```
(x == y ? 0 :
decrement(x), x = y,
increment(x), x)
```

This is achieved through the `modifyAssignExpression` method shown in Figure 28. The first thing the method does is check if the left-hand-side (lhs) of the expression (given by `binary.expression1()`), is a compiler-generated temporary. Unlike programmer-declared variables, the declaration of a temporary will not have a corresponding source location. Our implementation does not currently handle temporaries, and assignments to such are ignored. Temporaries are discussed in more detail in section 8.

```
CS_Initializer& ReferenceCounterTransformerImplementation::
        modifyExpressionInitializer(CS_ExpressionInitializer& init,
                                    CS_TypeDescriptor* td, CS_VariableDeclaration* var)
{
    assume(var);

    CS_TokenLocation loc =
        init.expression().sourceLocation().sourceRegion()->tokenLocation();

    if (td->isPointer()) {
        CS_BinaryExpression *be = init.expression().asBinaryExpression();
        assume(be);
        // C *c1(x) will already be transformed to C *c1 = x;
        assume(be->binaryExpressionKind() == CS_BinaryExpression::opAssign);

        CS_Expression &expr1 = be->expression1();

        // change C *c1; c1 = x to
        //        C *c1; c1 = (c1=x, c1 != 0 ? c1->rc++: 0, c1)

        init.setExpression(
            ef().createAssignExpression(loc,
                expr1,
            ef().createCommaExpression(loc,
                ef().createAssignExpression(loc,
                    ic().cloneExpression(expr1),
                    transformer().modifyExpression(be->expression2())),
                ef().createCommaExpression(loc,
                    incrementReferenceCounterExpression(
                        expr1),
                    ef().createCommaExpression(loc,
                        createStateChangeExpression(expr1, *var),
                    ic().cloneExpression(expr1))))));
        return init;
    }
    // For non-dynamic variables, initialize reference count
    // to 1 so that never get collected.

    init.setExpression(
        ef().createCommaExpression(loc,
            transformer().modifyExpression(init.expression()),
            ef().createAssignExpression(loc,
                ef().createDotExpression(loc,
                    ef().createNameExpression(loc, *var),
                    referenceCountMember()),
                ef().createLiteralExpression(
                    cs(), loc, intType(), 1))));
    return init;
}
```

Figure 29 modifyExpressionInitializer method

If the target of the assignment is not a temporary, a new comma expression is created using the result of the `decrementReferenceCounter` and `incrementReferenceCounter` methods along with the current assignment expression. The `findOrCreateDecrement` method called in the `decrementRefereneCounterExpression` method locates or creates a declaration corresponding to the `ReferenceCounter::increment` method declared earlier. Note that the right-hand-side (rhs) value is used three times in the resulting expression, in decrement, assignment, and increment expressions. We should generate a temporary to hold the rhs value so that expressions containing side-effects are not executed multiple times. Further, we currently do not generate code to handle the initial check for the lhs being equal to the rhs, which requires a temporary for both the lhs and the rhs, or the final expression containing just the rhs for the assignment value. This support would be added when temporaries are handled by our implementation (see section 8).

6.6.2 Transforming Initialization Expressions

The `modifyExpressionInitializer` method, shown in Figure 29, transforms initialization expressions corresponding to the model described earlier. Much of this method is fairly self-explanatory. For pointers, however, there is an additional action performed, which is to add state change information. If a pointer goes out of scope, either due to an exception or control implicitly returning from the function, the appropriate reference count decrement must take place. This is achieved through calling the `createStateChangeExpression` method, which will use the saved state change variable to create a state change node and insert it in the table in the appropriate place, based on the most recent state change that occurred within the function. This will cause code to be inserted at the end of the function on implicit scope termination to call the `dtor` method defined for the class `ReferenceCounter`, which will decrement the reference count. See the appendix for details.

7. An Example

To demonstrate the smart pointer transformer extension in action, Figure 31 shows a simple program containing several pointer declarations and assignments. Figure 30 shows the resulting execution output after compiling the program with the transformation extension. The built-in pointers act like smart pointers!

```
C c1;
C *cp1 = &c1;
>> Incrementing count for c1 to 2
C *cp2 = new C("new C 1");
>> Incrementing count for new C 1 to 1
cp2 = 0;
>> Decrementing count for new C 1 to 0
>> Deleting new C 1 with 0 references
cp1 = cp2;
>> Decrementing count for c1 to 1
C *cp3 = new C("new C 2");
>> Incrementing count for new C 2 to 1
C c2;
cp1 = &c2;
>> Incrementing count for c2 to 2
>> Deleting c2 with 2 references
>> Decrementing count for new C 2 to 0
>> Deleting new C 2 with 0 references
>> Decrementing count for c2 to -1
>> Deleting c1 with 1 references
```

Figure 30 Test Program Output

```
#include "ReferenceCounterInterface.h"

class C : public ReferenceCounter {
  public:
    int i;
    C(char *name) : ReferenceCounter(name) {}
};

int main()
{
    cout << "C c1;" << endl;
    C c1("c1");

    cout << "C *cp1 = &c1;" << endl;
    C *cp1 = &c1;

    cout << "C *cp2 = new C(\"new C 1\");" << endl;
    C *cp2 = new C("new C 1");

    cout << "cp2 = 0;" << endl;
    cp2 = 0;

    cout << "cp1 = cp2;" << endl;
    cp1 = cp2;

    cout << "C *cp3 = new C(\"new C 2\");" << endl;
    C *cp3 = new C("new C 2");

    cout << "C c2;" << endl;
    C c2("c2");

    cout << "cp1 = &c2;" << endl;
    cp1 = &c2;

    return 0;
}
```

Figure 31 Test program

```
transformed tree for: int main()
{
  __ef __fsm_tab = { 0xBEEFDEAD, 4, { { <offset of c1 + 0>, &C::__dftdt, 1, 16, 0, 0 }, {
<offset of @1 + 0>, &operator delete, -3, 16, 0, 1 }, { <offset of @2 + 0>, &operator delete, -
3, 16, 0, 2 }, { <offset of c2 + 0>, &C::__dftdt, 1, 16, 0, 3 } } };
  __est __es = { 0, 0, &__fsm_tab, (long int *) 0, 0 };
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C c1;"), endl);
  C c1; *C::C(&c1, "c1") , __es.__s = 1;
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C *cp1 = &c1;"), endl);
  C *cp1; cp1 = &c1;
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C *cp2 = new C(\"new C 1\");"),
endl);
  C *cp2; cp2 = (( @1 = ::operator new(16) ? __es.__s = 2 , C::C(@1, "new C 1") , __es.__s = 1 :
0 ) , @1);
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "cp2 = 0;"), endl);
  cp2 = 0;
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "cp1 = cp2;"), endl);
  cp1 = cp2;
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C *cp3 = new C(\"new C 2\");"),
endl);
  C *cp3; cp3 = (( @2 = ::operator new(16) ? __es.__s = 3 , C::C(@2, "new C 2") , __es.__s = 2 :
0 ) , @2);
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C c2;"), endl);
  C c2; *C::C(&c2, "c2") , __es.__s = 4;
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "cp1 = &c2;"), endl);
  cp1 = &c2;
  return @3 = 0 , (__es.__s = 3 , C::~C(&c2, 2, 0) , (__es.__s = 0 , C::~C(&c1, 2, 0))) , @3;
}
```

Figure 32 Transformed Expressions Without Smart Pointer Extension

Without the transformer extension, the transformed tree for the main function would be as shown in Figure 32. The first line of the transformed function contains a finite state machine table used for exception handling. Each of the 4 entries in the table specifies the action to take should an exception occur during execution of the function. There is an entry for the two local variables requiring destruction, along with the dynamically-allocated storage, in the order they occur in the function. Most of the remaining transformed function is fairly self-explanatory. The state variable `__es.__es` is updated to indicate the progress made, (in other words the state of the function), should an exception occur. The final line handles local destructors, updating the state as each destructor is called in reverse order of declaration. Note that there are no state table entries, state changes, or final destruction code, corresponding to the built-in pointers, as there is no cleanup necessary or possible for them.

Figure 33 shows the transformed tree with the reference counter transformer extension. The first difference is that the state table contains entries for each of the three built-in pointers, calling the `ReferenceCounter::dtor` method. In addition, state changes have been added throughout the function for the built-in pointer declarations. Each declaration or pointer assignment now includes the additional smart pointer functionality that was added as part of the transformation extension. And finally, at the end of the function, the built-in pointers are decremented as they go out of scope, in reverse order of declaration.

```
transformed tree for: int main()
{
  __ef __fsm_tab = { 0xBEEFDEAD, 7, { { <offset of c1 + 0>, &C::__dftdt, 1, 16, 0, 0 }, {
<offset of cp1 + 0>, &ReferenceCounter::dtor, 1, 4, 0, 1 }, { <offset of cp2 + 0>,
&ReferenceCounter::dtor, 1, 4, 0, 2 }, { <offset of @0 + 0>, &operator delete, -3, 16, 0, 3 }, {
<offset of cp3 + 0>, &ReferenceCounter::dtor, 1, 4, 0, 4 }, { <offset of @1 + 0>, &operator
delete, -3, 16, 0, 5 }, { <offset of c2 + 0>, &C::__dftdt, 1, 16, 0, 6 } } };
  __est __es = { 0, 0, &__fsm_tab, (long int *) 0, 0 };
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C c1;"), endl);
  C c1; *C::C(&c1, "c1") , __es.__s = 1 , c1.rc = 1;
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C *cp1 = &c1;"), endl);
  C *cp1; cp1 = (cp1 = &c1 , (ReferenceCounter::increment(static_cast<ReferenceCounter *> (cp1))
, (cp1 , __es.__s = 2 , cp1)));
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C *cp2 = new C(\"new C 1\");"),
endl);
  C *cp2; cp2 = (cp2 = (( @0 = ::operator new(16) ? __es.__s = 4 , C::C(@0, "new C 1") ,
__es.__s = 3 : 0 ) , @0) , (ReferenceCounter::increment(static_cast<ReferenceCounter *> (cp2)) ,
(cp2 , __es.__s = 3 , cp2)));
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "cp2 = 0;"), endl);
  ReferenceCounter::decrement(static_cast<ReferenceCounter *> (cp2)) , (cp2 = 0 ,
ReferenceCounter::increment(static_cast<ReferenceCounter *> (cp2)));
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "cp1 = cp2;"), endl);
  ReferenceCounter::decrement(static_cast<ReferenceCounter *> (cp1)) , (cp1 = cp2 ,
ReferenceCounter::increment(static_cast<ReferenceCounter *> (cp1)));
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C *cp3 = new C(\"new C 2\");"),
endl);
  C *cp3; cp3 = (cp3 = (( @1 = ::operator new(16) ? __es.__s = 6 , C::C(@1, "new C 2") ,
__es.__s = 5 : 0 ) , @1) , (ReferenceCounter::increment(static_cast<ReferenceCounter *> (cp3)) ,
(cp3 , __es.__s = 5 , cp3)));
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "C c2;"), endl);
  C c2; *C::C(&c2, "c2") , __es.__s = 7 , c2.rc = 1;
  foo();
  *ostream::operator<<(ostream::operator<<((ostream *) &cout, "cp1 = &c2;"), endl);
  ReferenceCounter::decrement(static_cast<ReferenceCounter *> (cp1)) , (cp1 = &c2 ,
ReferenceCounter::increment(static_cast<ReferenceCounter *> (cp1)));
  return @2 = 0 , (__es.__s = 6 , C::~C(&c2, 2, 0) , (__es.__s = 5 ,
ReferenceCounter::decrement(static_cast<ReferenceCounter *> (cp3))) , (__es.__s = 3 ,
ReferenceCounter::decrement(static_cast<ReferenceCounter *> (cp2))) , (__es.__s = 2 ,
ReferenceCounter::decrement(static_cast<ReferenceCounter *> (cp1))) , (__es.__s = 0 , C::~C(&c1,
2, 0))) , @2;
}
```

Figure 33 Transformed Expressions With Extension

8. Further Work

There are several areas that were not covered by this work, most notably temporaries. The problem with temporaries is that the API currently does not support them very well. There is no model for detecting when a temporary goes out of use, which is necessary in order to correctly apply reference counting. The current model does not include temporaries in the reference count, which is sufficient for some, but not all, cases. Consider an expression such as `cp2 = cp1++;` The initial value of `cp2` must be saved in a temporary prior to the increment of `cp2` in order to be assigned to `cp1,` so the expression would be transformed as follows:

```
cp2 = (@0 = cp1 , cp1 = cp1 + 1 , @0);
```

Applying the smart pointer transformation without taking into account the temporary would yield an incorrect result if the decrement against `cp1` caused the underlying storage to be deleted, resulting in a dangling reference being assigned to `cp2.` The API needs a mechanism for allowing a transformation to determine when a temporary goes out of use so that, for this example, the appropriate reference counting can take place. When such support for temporaries is avaiable, the smart pointer implementation must also be updated to generate temporaries for the modified expressions to avoid multiple evaluation of expressions containing side-effects, as discussed earlier.

Further work also needs to be done in the area of non-implicit scope termination, (for example through a return statement), and exception handling. The current transformation implementation handles reference decrementing only for implicit scope termination, through the state table additions. However, the API does support the capability to handle explicit scope termination, by detecting return statements and using the state information to determine what needs to be done. With respect to exception handling, the current extension implementation does work for simple examples, but not for all cases. Due to time constraints, we did not pursue these areas, but anticipate that the implementation would be fairly straightforward.

9. Conclusion

C++ smart pointers, while similar to built-in pointers, cannot be used interchangeably. Most notably, implicit compiler conversions are not supported for smart pointers. We have proposed that the “smarts” of smart pointers be added to built-in pointers, and presented the expression transformations that would be necessary to implement a reference-counting built-in pointer. Using the Montana API, we have demonstrated a working example of these ideas.

The Montana API interface has proven to be quite complete for the purposes of adding transformation extensions. Other work in this area [Car97] supports this conclusion. We found the API interface and design to be reasonably straightforward and understandable, particularly given the complexity of the problem we were attempting to solve, that of modifying compiler-generated expressions. Nonetheless, adding a transformation extension is not a trivial undertaking, and is more likely to be expected of a class library vendor rather than a casual programmer.

While there is some additional work necessary to allow full support for a reference counting smart pointer implementation, it is clear that the interface is quite capable of handling such language-level extensions. The API definitely need better support for temporaries, both those generated by the compiler and by extensions such as the one we have demonstrated. However, given the flexibility of the interface, this does not seem like a difficult design issue.

10. References

[Alg95] Alger, Jeff; *Secrets of the C++ Masters*, London, England: Academic Press, 1995.

[BCCKLN94] Barton, John; Charles, Phillipe; Chee, Yi-Min; Karasick, Michael; Lieber, Derek; and Nackman, Lee; "CodeStore: Infrastructure for C++-Knowlegeable Tools", Unpublished position paper submitted to OOPSLA 1994 Workshop on Object-Oriented Compilation. http://www.research.ibm.com/softwaretechnology/papers/oopsla94_codestore/position.html

[Car97] Carmichael, Ian; Unpublished working example of using a Montana transformation extension and global program knowledge to de-virtualize virtual function calls.

[Chu94] Churchill, Steve; "Exception Recovery with Smart Pointers", *C++ Report*, January 1994.

[Coh96] Cohen, Shimon; "Lightweight Persistence in C++", *C++ Report*, May 1996.

[Det92] Detlefs, David; "Garbage Collections and Run-time Typing as a C++ Library", *Usenix C++ Technical Conference Proceedings*, June 1992.

[Ede92a] Edelson, Daniel R.; "A Mark-and-Sweep Collector for C++", *Conference Record of ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, January 1992.

[Ede92b] Edelson, Daniel R.; "Smart Pointers: They're Smart, but They're Not Pointers", *USENIX C++ Conference Proceedings*, 1992.

[Ede92c] Edelson, Daniel R.; "Precompiling C++ for Garbage Collection", I*nternational Workshop on Memory Management Proceedings*, September 1992.

[GC96] "Great Circle Technology Overview"; Geodesic Systems, 1996.

[Ham96] Hamilton, Jennifer; *Programming with DirectoToSOM C++*, New York, NY: John Wiley, 1996.

[Har96] Harvey, David; "Smart Pointer Templates in C++", http://web.ftech.net/~honeyg/articles/smartp.html, 1996.

[Lip96] Lippman, Stanley; *Inside the C++ Object Model*, Reading, Mass: Addison-Wesley, 1996.

[Mey96a] Meyers, Scott; *More Effective C++*, Reading, Mass: Addison-Wesley, 1996.

[Mey96b] Meyers, Scott; "Refinements to Smart Pointers", *C++ Report*, November-December, 1996.

[Mil96] Miller, Justin; "Clean Up: C++ Garbage Collection", *Byte*, January 1996.

[Stro91] Stroustrup, Bjarne; *The C++ Programming Language, 2nd Edition*, Reading, Mass. Addison-Wesley, 1991.

[Stro89] Stroustrup, Bjarne; "The Evolution of C++: 1985 - 1989", *Computing Systems*, Summer 1989.

[Stro94] Stroustrup, Bjarne; *The Design and Evolution of C++*, Reading, Mass: Addison-Wesley, 1994.

11. Special References

The following referenced papers are IBM documents that are expected to be made publicly available as part of the Montana product. We provide a brief synopsis here.

[Kar96] Karasick, Michael; "So You Want To Write a Tool?", IBM document, 1996. *Introduces the structure of CodeStore tools that use the query interface and explains the CodeStore programming conventions and idioms.*

[Nac96] Nackman, Lee; "An Overview of Montana", IBM document, 1996. *Provides an introduction to Montana, it's history, goals, and architecture.*

[Sor96] Soroker, Danny; "Montana White Paper: Extensions", IBM document, 1996. *Describes the Montana interface for defining and loading extensions.*

[Stre96] Streeter, Dave; "Building Programs with Montana", IBM document, 1996. *Describes how to use Montana to compile programs and build the Montana database.*

12. Appendix - Complete Implementation

12.1 SmartPointer.h

```
#ifndef SmartPointerH
#define SmartPointerH

#include "interface/Extension.h"
#include "interface/ExtensionOption.h"
#include "interface/ExtensionSource.h"
#include "interface/InterfaceBase.h"

#include "common/AtomStore.h"

#include "incorporator/ExtensionDynamicLoadPoint.h"
#include "incorporator/ImplementationComponentFactory.h"
#include "incorporator/IncorporationState.h"
#include "incorporator/Transformer.h"
#include "incorporator/TypeAnalyzer.h"

#define EXTENSION_PRIORITY CS_DependencyNode::beforeLibrariesPriority

class CS_String;
class CS_ExtensionOptionBase;
class SmartPointerTransformer;
class ReferenceCounterTransformerImplementation;

//
// The SmartPointer class provides the call-back methods used by the kernel.
//
class SmartPointer : public CS_InterfaceBase
{
public:
   static const char* className();

   static void SmartPointer::incorporationStartup(CS_ExtensionDynamicLoadPointLink&,
CS_IncorporationState&);

   static CS_DependencyNode::UpdateResult
        update(CS_ExtensionSource* me, CS_IncorporationState& state, CS_bool emitMessages);

   static CS_bool isChanged(CS_ExtensionSource* me);

   static void processOptions(CS_ExtensionSource* me, CS_OptionList& options);

private:
   static CS_ExtensionDynamicLoadPoint  extension_load_point;
};

//
// The SmartPointerImplementationComponentFactory class
// registers the SmartPointerTransformer class as a new transformation phase.
//
class SmartPointerImplementationComponentFactory
   : public CS_ImplementationComponentFactory {
  public:
    SmartPointerImplementationComponentFactory(CS_IncorporationState&);
    virtual CS_Transformer& transformer();
    virtual ~SmartPointerImplementationComponentFactory();

  private:
    CS_IncorporationState& _state;
    CS_Transformer* _transformer;
};

//
// This is the common base class for all smart pointer implementations
```

```
//
class SmartPointerTransformerImplementation : public
CS_IncorporationComponentBase<CS_InterfaceBase> {
  public:
    SmartPointerTransformerImplementation(CS_IncorporationState &state,
                                          SmartPointerTransformer &transformer) :
        CS_IncorporationComponentBase<CS_InterfaceBase>(state), _transformer(transformer) {}

    virtual CS_Expression& modifyAssignExpression(CS_BinaryExpression&) = 0;
    virtual CS_Initializer& modifyExpressionInitializer(
        CS_ExpressionInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*) = 0;
    virtual CS_Initializer& modifyImplicitInitializer(
        CS_ImplicitInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*) = 0;

    static CS_Expression& modifyDestructorStateChangeExpression(
        CS_DestructorStateChangeExpression& dsce);
    CS_DestructorStateTableEntry* currentDestructorStateTableEntry()
        { return _currentDestructorStateTableEntry; }
    void currentDestructorStateTableEntry(CS_DestructorStateTableEntry *ste)
        { _currentDestructorStateTableEntry=ste; }

    SmartPointerTransformer& transformer() { return _transformer; }

  private:

    SmartPointerTransformer& _transformer;
    static CS_DestructorStateTableEntry *_currentDestructorStateTableEntry;
};

//
// This class handles the smart pointer transformation phase
//
// use -showTransformedGraph to view completed transformations
//
class SmartPointerTransformer : public CS_Transformer {
  public:
    SmartPointerTransformer(CS_IncorporationState& s, CS_Transformer& p)
       : CS_Transformer(s), _parent(p),
         _referenceCounterTransformerImplementation(0) {
    }

    ~SmartPointerTransformer();

    virtual CS_bool transform(CS_Statement*& stmt, CS_bool emitMessages);
    virtual CS_bool transform(CS_Initializer*&, CS_VariableDeclaration&, CS_bool);
    virtual CS_Expression& transform(CS_Expression&, CS_bool);

    virtual CS_Expression& modifyAssignExpression(CS_BinaryExpression&);
    virtual CS_Initializer& modifyExpressionInitializer(
        CS_ExpressionInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);
    virtual CS_Initializer& modifyImplicitInitializer(
        CS_ImplicitInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);
    virtual CS_Expression& modifyDestructorStateChangeExpression(
        CS_DestructorStateChangeExpression&);

    CS_Expression& typeAnalyze(CS_Expression&);
    CS_Initializer& typeAnalyze(CS_Initializer&);

  private:
    CS_Transformer& _parent;

    // classes for each smart pointer implementation
    ReferenceCounterTransformerImplementation* _referenceCounterTransformerImplementation;

    // Return the smart pointer implementation, if any, for the expression
    // The expression must be a pointer to a class derived from a SmartPointer class
    SmartPointerTransformerImplementation* transformerImplementation(CS_TypeDescriptor&);
};
#endif
```

12.2 ReferenceCounter.h

```
#include "SmartPointer.h"

class ReferenceCounterTransformerImplementation :
        public SmartPointerTransformerImplementation {
  public:
    ReferenceCounterTransformerImplementation(
        CS_IncorporationState& state, SmartPointerTransformer& transformer) :
        SmartPointerTransformerImplementation(state, transformer) {};

    CS_ClassDeclaration* findClassDeclaration();

    CS_Expression& modifyAssignExpression(CS_BinaryExpression& binary);
    CS_Initializer& modifyExpressionInitializer(
        CS_ExpressionInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);
    CS_Initializer& modifyImplicitInitializer(
        CS_ImplicitInitializer&, CS_TypeDescriptor*, CS_VariableDeclaration*);

    private:
        const CS_Atom& referenceCountMember();
        virtual CS_Expression& decrementReferenceCounterExpression(CS_Expression &);
        virtual CS_Expression& incrementReferenceCounterExpression(CS_Expression &);
        virtual CS_Expression& createStateChangeExpression(
            CS_Expression&, CS_VariableDeclaration&);
        virtual CS_FunctionDeclaration& findOrCreateDecrement();
        virtual CS_FunctionDeclaration& findOrCreateIncrement();
        virtual CS_FunctionDeclaration& findOrCreateDtor();
        virtual CS_FunctionDeclaration& findOrCreateMemberFunction(char *);
        virtual CS_DestructorStateTableEntry&createDestructorStateTableEntry(
            CS_VariableDeclaration&, CS_TreeNode&);
~
        virtual void addDestructorCalls(
            CS_VariableDeclaration&, CS_DestructorStateTableEntry&, CS_TokenLocation&);
};
```

12.3 ReferenceCounterInterface.h

```
#include <iostream.h>

class ReferenceCounter;

class ReferenceCounter {
  private:
    int rc;
    char *_name;

    static void dtor(ReferenceCounter **sp, int);
    static void decrement(ReferenceCounter *sp);
    static void increment(ReferenceCounter *sp);

  public:
    ReferenceCounter(char *name) : _name(name) { rc = 0; }
    char *name() { return _name; }
    virtual ~ReferenceCounter();
};

ReferenceCounter::~ReferenceCounter()
{
    cout << "Deleting " << _name <<
            " with " << rc << " references" << endl;
    rc = 0;
}

void ReferenceCounter::dtor(ReferenceCounter **sp, int)
{
    // can assume sp not null
```

```
    decrement(*sp);
}

void ReferenceCounter::decrement(ReferenceCounter *sp)
{
    if (!sp)
        return;
    cout << "Decrementing count for " << sp->name() << " to " << sp->rc-1 << endl;
    if (! --sp->rc)
        delete sp;
}

void ReferenceCounter::increment(ReferenceCounter *sp)
{
    if (!sp)
        return;
    cout << "Incrementing count for " << sp->name() << " to " << sp->rc+1 << endl;
    ++sp->rc;
}
```

12.4 SmartPointer.C

```
#include <sys/stat.h>
#include <iostream.h>

#include "SmartPointer.h"
#include "ReferenceCounter.h"

CS_ExtensionDynamicLoadPoint
   SmartPointer::extension_load_point(
        SmartPointer::className(),
        SmartPointer::update,
        SmartPointer::isChanged,
        SmartPointer::processOptions,
        EXTENSION_PRIORITY,
        SmartPointer::incorporationStartup);

//--------- SmartPointer --------

// This method is called when incorporation begins. Is is hooked
// in through the initialization fo extension_load_point with the
// DLL is statically loaded.
void SmartPointer::incorporationStartup(
        CS_ExtensionDynamicLoadPointLink&, CS_IncorporationState& state)
{
    cout << __FUNCTION__ << endl;

    SmartPointerImplementationComponentFactory* fac =
        new SmartPointerImplementationComponentFactory(state);
    assume(fac);  // (our version of "assert")

    // Push our new factory with its new Transformer onto the chain
    // stored in the IncorporationState.
    //
    state.prependImplementationComponentFactory(*fac);

    return;
}

CS_DependencyNode::UpdateResult SmartPointer::
        update(CS_ExtensionSource* me, CS_IncorporationState& state,
               CS_bool emitMessages)
{
   return CS_DependencyNode::success;
}

CS_bool SmartPointer::isChanged(CS_ExtensionSource* me)
{
   return (me->timestampHasChanged());
```

```
}

void SmartPointer::
        processOptions(CS_ExtensionSource* me, CS_OptionList& options)
{
}

const char* SmartPointer::className()
{
   return "SmartPointer";
}


//--------- SmartPointerImplementationComponentFactory --------

// This method is called from SmartPointer::incporporationStartup to
// register a new transformation phase
SmartPointerImplementationComponentFactory::
        SmartPointerImplementationComponentFactory(CS_IncorporationState& s) :
    _state(s),
    // The second argument to this constructor comes from pull on the
    // chain of components stored in the IncorporationState.
    //
    _transformer(new SmartPointerTransformer(_state,
                  _state.implementationComponentFactory().transformer()))
{
    assume(_transformer);
}

CS_Transformer& SmartPointerImplementationComponentFactory::
        transformer()
{
    assume(_transformer);
    return *_transformer;
}

SmartPointerImplementationComponentFactory::
        ~SmartPointerImplementationComponentFactory()
{
    delete _transformer;
}

//--------- SmartPointerTransformer --------

CS_bool SmartPointerTransformer::
        transform(CS_Statement*& stmt, CS_bool emitMessages)
{
    modifyStatement(*stmt);
    _parent.transform(stmt, emitMessages);
    return CS_true;
}

CS_bool SmartPointerTransformer::
        transform(CS_Initializer*& init, CS_VariableDeclaration& var, CS_bool emitMessages)
{
    init = &modifyInitializer(*init, &var.typeDescriptor(), &var);
    _parent.transform(init, var, emitMessages);
    return CS_true;
}

CS_Expression& SmartPointerTransformer::
        transform(CS_Expression& expr, CS_bool emitMessages)
{
    CS_Expression *expr2 = &modifyExpression(expr);
    return _parent.transform(*expr2, emitMessages);
}

CS_Expression& SmartPointerTransformer::
        modifyAssignExpression(CS_BinaryExpression& binary)
{
    if (! binary.expression1().typeDescriptor().isPointer())
        return binary;
```

```
    SmartPointerTransformerImplementation *ti =
        transformerImplementation(*binary.expression1().typeDescriptor().next());

    return ti ? typeAnalyze(ti->modifyAssignExpression(binary)) : binary;
}

CS_Initializer& SmartPointerTransformer::
        modifyExpressionInitializer(CS_ExpressionInitializer& init,
                                    CS_TypeDescriptor* td,
                                    CS_VariableDeclaration* var)
{
    if (!td)
        return init;

    // need to handle both pointer and object initialization
    SmartPointerTransformerImplementation *ti =
        transformerImplementation(td->isPointer() ? *td->next() : *td);

    if (! ti)
        return init;

    return typeAnalyze(ti->modifyExpressionInitializer(init, td, var));
}

CS_Initializer& SmartPointerTransformer::
        modifyImplicitInitializer(CS_ImplicitInitializer& init,
                                  CS_TypeDescriptor* td,
                                  CS_VariableDeclaration* var)
{
    if (!td)
        return init;

    // need to handle both pointer and object initialization
    SmartPointerTransformerImplementation *ti =
        transformerImplementation(td->isPointer() ? *td->next() : *td);

    if (! ti)
        return init;

    return typeAnalyze(ti->modifyImplicitInitializer(init, td, var));
}

CS_Expression& SmartPointerTransformer::
        modifyDestructorStateChangeExpression(CS_DestructorStateChangeExpression& dsce)
{
    if (dsce.tableEntry() && dsce.tableEntry()->asDestructorStateTableEntry())
        cerr << " for: " << dsce.tableEntry()->
                asDestructorStateTableEntry()->variable().signature();
    if (dsce.tableEntry())
        cerr << ", tableEntry address: " << (void *)dsce.tableEntry();
    cerr << ", expression address: " << (void *)&dsce;
    cerr << endl;

    return SmartPointerTransformerImplementation::
        modifyDestructorStateChangeExpression(dsce);
}

// to call _parent transform.
CS_Expression& SmartPointerTransformer::typeAnalyze(CS_Expression& expr)
{
    CS_Expression *exprp = &expr;
    // perform type analysis on the expression
    CS_bool result = state().typeAnalyzer().analyze(exprp, 0, CS_false);
    assume(result);
    return *exprp;
}

CS_Initializer& SmartPointerTransformer::typeAnalyze(CS_Initializer& init)
{
    if (! init.asExpressionInitializer())
        return init;
```

```
    CS_Expression &ei = init.asExpressionInitializer()->expression();
    CS_TokenLocation loc = ei.sourceLocation().sourceRegion()->tokenLocation();
    CS_Initializer& newInit =
        ef().createExpressionInitializer(loc, typeAnalyze(ei));
    // indicate var should be initialized, not set for statics as BE will handle
    newInit.setInitialization(CS_true);
    return newInit;
}

SmartPointerTransformerImplementation *SmartPointerTransformer::
        transformerImplementation(CS_TypeDescriptor& td)
{
    if (! td.isNamedType() ||
            ! td.declaration().declarationKind() == CS_Declaration::IsClass)
        return NULL;
    CS_ClassDeclaration &decl = *td.declaration().asClassDeclaration();

    CS_ClassDeclaration *referenceCounter =
        ReferenceCounterTransformerImplementation(state(), *this).findClassDeclaration();
    if (referenceCounter &&
            CS_ANSI_Queries::isBaseClassOf(*referenceCounter, decl)) {
        cout << "got a pointer to class derived from "
             << referenceCounter->signature() << endl;
        if (! _referenceCounterTransformerImplementation) {
            _referenceCounterTransformerImplementation =
                new ReferenceCounterTransformerImplementation(state(), *this);
        }
        return _referenceCounterTransformerImplementation;
    }
    // insert code to look for other smart pointer class implementations here
    return NULL;
}

SmartPointerTransformer::~SmartPointerTransformer()
{
    delete _referenceCounterTransformerImplementation;
}

CS_DestructorStateTableEntry *SmartPointerTransformerImplementation::
    _currentDestructorStateTableEntry = 0;

CS_Expression& SmartPointerTransformerImplementation::
    modifyDestructorStateChangeExpression(
        CS_DestructorStateChangeExpression& dsce)
{
    // save most recent state table entry
    if (dsce.tableEntry() && dsce.tableEntry()->asDestructorStateTableEntry())
        _currentDestructorStateTableEntry = dsce.tableEntry()->asDestructorStateTableEntry();

    return dsce;
}
```

12.5 ReferenceCounter.C

```
#include "interface/LexicalBlockStatement.h"
#include "incorporator/StatementFactory.h"

#include "ReferenceCounter.h"
//
// change x = &a to x != 0 ? (x->rc-- == 0 ? delete x : 0) : 0,
//                  x = &a,
//                  x != 0 ? x->rc++ : 0
//
CS_Expression&
ReferenceCounterTransformerImplementation::modifyAssignExpression(
    CS_BinaryExpression &binary)
{
    // Don’t transform expressions on temporaries
```

```
    CS_Expression &e = binary.expression1();
    if (e.expressionKind() == CS_Expression::IsName &&
            e.asNameExpression()->name().declaration() &&
            ! e.asNameExpression()->name().declaration()->mapsToASourceLocation()) {
        return binary;
    }

    CS_TokenLocation loc =
        binary.sourceLocation().sourceRegion()->tokenLocation();

    CS_Expression &expr =
        ef().createCommaExpression(loc,
            decrementReferenceCounterExpression(binary.expression1()),
            ef().createCommaExpression(loc,
                binary,
                incrementReferenceCounterExpression(binary.expression1())));
    return expr;
}

CS_Initializer& ReferenceCounterTransformerImplementation::
        modifyExpressionInitializer(CS_ExpressionInitializer& init,
                                    CS_TypeDescriptor* td, CS_VariableDeclaration* var)
{
    assume(var);

    CS_TokenLocation loc =
        init.expression().sourceLocation().sourceRegion()->tokenLocation();

    if (td->isPointer()) {
        CS_BinaryExpression *be = init.expression().asBinaryExpression();
        assume(be);
        // C *c1(x) will already be transformed to C *c1 = x;
        assume(be->binaryExpressionKind() == CS_BinaryExpression::opAssign);

        CS_Expression &expr1 = be->expression1();

        // change C *c1; c1 = x to
        //        C *c1; c1 = (c1=x, c1 != 0 ? c1->rc++: 0, c1)

        init.setExpression(
            ef().createAssignExpression(loc,
                expr1,
            ef().createCommaExpression(loc,
                ef().createAssignExpression(loc,
                    ic().cloneExpression(expr1),
                    transformer().modifyExpression(be->expression2())),
                ef().createCommaExpression(loc,
                    incrementReferenceCounterExpression(
                        expr1),
                    ef().createCommaExpression(loc,
                        createStateChangeExpression(expr1, *var),
                    ic().cloneExpression(expr1))))));
        return init;
    }

    // For non-dynamic variables, initialize reference count
    // to 1 so that never get collected.

    init.setExpression(
        ef().createCommaExpression(loc,
            transformer().modifyExpression(init.expression()),
            ef().createAssignExpression(loc,
                ef().createDotExpression(loc,
                    ef().createNameExpression(loc, *var),
                    referenceCountMember()),
                ef().createLiteralExpression(
                    cs(), loc, intType(), 1))));
    return init;
}

CS_Initializer& ReferenceCounterTransformerImplementation::
        modifyImplicitInitializer(CS_ImplicitInitializer& init,
```

```
                                    CS_TypeDescriptor* td, CS_VariableDeclaration* var)
{
    return init;
}

// when a pointer goes out of scope, decrement it's reference count
CS_Expression &ReferenceCounterTransformerImplementation::
    createStateChangeExpression(CS_Expression &expr, CS_VariableDeclaration &var)
{
    CS_TokenLocation loc =
        expr.sourceLocation().sourceRegion()->tokenLocation();

    CS_DestructorStateTableEntry& ste =
        createDestructorStateTableEntry(var, expr);
    addDestructorCalls(var, ste, loc);
    currentDestructorStateTableEntry(&ste);

    CS_Expression &expr2 = transformer().typeAnalyze(expr);

    CS_DestructorStateChangeExpression &sce =
        ef().createDestructorStateChangeExpression(loc,
            CS_DestructorStateChangeExpression::IsConstruction,
            &ste,
            &expr2);
    sce.setTypeDescriptor(var.typeDescriptor());
    ste.setConstruction(sce);
    return sce;
}

CS_FunctionDeclaration&
ReferenceCounterTransformerImplementation::findOrCreateMemberFunction(char *func)
{
    const CS_Atom &name =
        codeStore().atomStore().findOrCreate(func);
    CS_Declaration *dec =
        findClassDeclaration()->memberDeclarationStore().findFirstDeclaration(name);
    assume(dec);
    assume(dec->asFunctionDeclaration());
    return *(dec->asFunctionDeclaration());
}

CS_FunctionDeclaration& ReferenceCounterTransformerImplementation::findOrCreateDtor()
{
    return findOrCreateMemberFunction("dtor");
}

CS_FunctionDeclaration& ReferenceCounterTransformerImplementation::findOrCreateDecrement()
{
    return findOrCreateMemberFunction("decrement");
}

CS_FunctionDeclaration& ReferenceCounterTransformerImplementation::findOrCreateIncrement()
{
    return findOrCreateMemberFunction("increment");
}

CS_DestructorStateTableEntry &ReferenceCounterTransformerImplementation::
    createDestructorStateTableEntry(CS_VariableDeclaration &var, CS_TreeNode& n)
{
    CS_FunctionDeclaration& dtor = findOrCreateDtor();
    CS_DestructorStateTableEntry& ste =
        ef().createDestructorStateTableEntry(
            CS_DestructorStateTableEntry::IsLocalVariable,
            // next state table entry
            currentDestructorStateTableEntry() ?
                currentDestructorStateTableEntry() : NULL,
            var,
            NULL, // class type
            1,        // number of elements
            0,  // offset
            dtor);
```

```
    assume(outermostLexicalBlockStatement(n));

    CS_LexicalBlockStatement &block = *outermostLexicalBlockStatement(n);

    block.insertStateTableEntryAfter(ste, currentDestructorStateTableEntry());

    return ste;
}

void ReferenceCounterTransformerImplementation::
    addDestructorCalls(CS_VariableDeclaration &var, CS_DestructorStateTableEntry &dste,
CS_TokenLocation &loc)
{
    // For each state table entry, an ealier phase will already have figured
    // out where destructor calls (ie at return/break/goto) are required.
    // Need to insert a destructor expression for each
    // point at which the previous state table entry required
    // a destructor. This handles the timing of destructors so that
    // our destructor gets called in the correct relative order
    // If the variable for the previous state change expression is not
    // in the same lexical block, then do nothing, as no timing concerns.
    // Instead must add info to return/break/goto statements as encounter
    // them (using modifyxxx).

    if (! dste.next())
        return;

    // find the first compiler-generated DestructorStateTableEntry - it will have
    // the destructor calls that we want to append to
    for (CS_StateTableEntry *nextSte = dste.next(); nextSte; nextSte=nextSte->next()) {
        if (! nextSte->asDestructorStateTableEntry())
            continue;
        if (nextSte->asDestructorStateTableEntry()->destructorCalls().size() > 0)
            break;
    }

    if (! nextSte)
        return;
    CS_DestructorStateTableEntry &nextCompilerDtse = *nextSte->asDestructorStateTableEntry();

    // make sure the DTSE we found is in the same scope
    if (dste.variable().enclosingDeclarationStore() !=
                nextCompilerDtse.variable().enclosingDeclarationStore())
        return;

    CS_List<CS_DestructorStateChangeExpression>& dsceList =
        nextCompilerDtse.destructorCalls();

    CS_ListIterator<CS_DestructorStateChangeExpression> iter(dsceList);
    for (CS_DestructorStateChangeExpression *dsce = iter.first();
            dsce; dsce = iter.next()) {
        CS_TreeNode* parent = dsce->parent();
        assume(parent);

        CS_FunctionCallExpression &dtorCall =
            ef().createFunctionCallExpression(loc,
                ef().createNameExpression(loc, findOrCreateDecrement()));
        dtorCall.appendArgument(
            ef().createExpressionInList(
                ef().createNameExpression(loc, var)));
        CS_Expression &dtorCall2 =
            transformer().typeAnalyze(dtorCall);
        dtorCall.setTypeDescriptor(var.typeDescriptor());

        CS_DestructorStateChangeExpression &sce =
                ef().createDestructorStateChangeExpression(loc,
                    CS_DestructorStateChangeExpression::IsDestruction,
                    &dste,
                    &dtorCall);
        sce.setTypeDescriptor(voidType());
        CS_Expression &expr = ef().createCommaExpression(loc, *dsce, sce);
        expr.setTypeDescriptor(dsce->typeDescriptor());
```

```
        replaceExpression(*parent, expr, *dsce);
    }
}

CS_Expression&
ReferenceCounterTransformerImplementation::decrementReferenceCounterExpression(
    CS_Expression &expr)
{
    CS_TokenLocation loc =
        expr.sourceLocation().sourceRegion()->tokenLocation();

    CS_FunctionCallExpression &decCall =
        ef().createFunctionCallExpression(loc,
            ef().createNameExpression(loc, findOrCreateDecrement()));
    decCall.appendArgument(
        ef().createExpressionInList(
            ic().cloneExpression(expr)));
    return decCall;
}

CS_Expression&
ReferenceCounterTransformerImplementation::incrementReferenceCounterExpression(
    CS_Expression &expr)
{
    CS_TokenLocation loc =
        expr.sourceLocation().sourceRegion()->tokenLocation();

    CS_FunctionCallExpression &incCall =
        ef().createFunctionCallExpression(loc,
            ef().createNameExpression(loc, findOrCreateIncrement()));
    incCall.appendArgument(
        ef().createExpressionInList(
            ic().cloneExpression(expr)));
    return incCall;
}

const CS_Atom &ReferenceCounterTransformerImplementation::referenceCountMember()
{
    return cs().atomStore().findOrCreate("rc");
}

CS_ClassDeclaration* ReferenceCounterTransformerImplementation::
    findClassDeclaration()
{
    const CS_Atom &name =
        codeStore().atomStore().findOrCreate("ReferenceCounter");
    CS_Declaration *decl =
        codeStore().globalDeclarationStore().findFirstDeclaration(name);
    if (! decl)
        return NULL;
    return decl->asClassDeclaration();
}
```