1. Introduction. LossLess is a programming language and environment similar to scheme. This document describes the implementation of a LossLess runtime written in C and LossLess itself will be described elsewhere. In unambiguous cases LossLess may be used to refer specifically to the implementation.

This code started off its life as s9fes by Nils M. Holm[1]. After a few iterations including being briefly ported to perl this rather different code is the result, although at its core it follows the same design.

All of the functions, variables, etc. used by LossLess are exported via `lossless.h`, even those which are nominally internal. Although this is not best practice for a library it makes this document less repetetive and facilitates easier testing.

⟨ lossless.h 1 ⟩ ≡
#ifndef LOSSLESS_H
#define LOSSLESS_H
⟨ System headers 4 ⟩
⟨ Preprocessor definitions ⟩
⟨ Complex definitions & macros 140 ⟩
⟨ Type definitions 5 ⟩
⟨ Function declarations 8 ⟩
⟨ Externalised global variables 7 ⟩
#endif

2. The structure is of a virtual machine with a single accumulator register and a stack. There is a single entry point to the VM—*interpret*—called after parsed source code has been put into the accumulator, where the result will also be left.

⟨ System headers 4 ⟩
⟨ Preprocessor definitions ⟩
⟨ Complex definitions & macros 140 ⟩
⟨ Type definitions 5 ⟩
⟨ Function declarations 8 ⟩
⟨ Global variables 6 ⟩

3. ⟨ Global initialisation 3 ⟩ ≡ /* This is located here to name it in full for CWEB's benefit */

See also sections 33, 69, 101, 112, and 186.
This code is cited in section 95.
This code is used in section 96.

[1] `http://t3x.org/s9fes/`

4. `LossLess` has few external dependencies, primarily *stdio* and *stdlib*, plus some obvious memory mangling functions from the C library there's no point in duplicating.

`LL_ALLOCATE` allows us to define a wrapper around *reallocarray* which is used to make it artificially fail during testing.

⟨ System headers 4 ⟩ ≡
```
#include <ctype.h>
#include <limits.h>
#include <setjmp.h>
#include <stdarg.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>    /* for memset */
#include <sys/types.h>
#ifndef LL_ALLOCATE
#define LL_ALLOCATE reallocarray
#endif
```
This code is used in sections 1, 2, and 213.

5. The *boolean* and *predicate* C types are used to distinguish between *boolean*-returning functions reporting C truth (0 or 1) or *predicate*-returning functions reporting `LossLess` truth (`FALSE` or `TRUE`). Otherwise-untyped C macros always report C truth.

```
#define bfalse 0
#define btrue 1
```

⟨ Type definitions 5 ⟩ ≡
```
typedef int32_t cell;
typedef int boolean;
typedef cell predicate;
```
See also section 88.

This code is used in sections 1 and 2.

6. Error Handling. Everything needs to be able to report errors and so even though the details will make little sense without a more complete understanding of `LossLess` the code and data to handle them come first in full.

When the VM begins it establishes two jump buffers. To understand jump buffers it's necessary to understand how C's stack works and we have enough stacks already.

The main thing to know is that whenever C code calls a function it grows its own stack with the caller's return address. When *setjmp* is called the position in this stack is saved. When jumping back to that position with *longjmp*, anything which has been added to C*'s* stack since the corresponding call to *setjmp* is discarded which has the effect of returning to exactly the point in the program where the corresponding *setjmp* was called, this time with a non-zero return value (the value that was given as an argument to *longjmp*; this facility is not used by `LossLess` for anything and it always sends 1).

The other thing that you don't need to know is that sometimes C compilers can make the previous paragraph a tissue of lies.

#define `ERR_UNIMPLEMENTED "unimplemented"`
#define **error**(x, d) *handle_error*$((x), \texttt{NIL}, (d))$
#define *ex_id* *car*
#define *ex_detail* *cdr*

⟨ Global variables 6 ⟩ ≡
```
volatile boolean Error_Handler = bfalse;
jmp_buf Goto_Begin;
jmp_buf Goto_Error;
```
See also sections 12, 19, 25, 30, 47, 56, 66, 89, 91, 99, 110, 159, 185, 211, 214, and 227.

This code is used in section 2.

7. ⟨ Externalised global variables 7 ⟩ ≡
```
extern volatile boolean Error_Handler;
extern jmp_buf Goto_Begin;
extern jmp_buf Goto_Error;
```
See also sections 13, 20, 26, 31, 42, 48, 57, 67, 90, 92, 100, 111, 160, 212, 215, and 228.

This code is used in section 1.

8. ⟨ Function declarations 8 ⟩ ≡
```
void handle_error(char *, cell, cell) __dead;
void warn(char *, cell);
```
See also sections 14, 22, 27, 36, 43, 51, 59, 70, 73, 79, 86, 95, 102, 107, 113, 133, 161, 197, 218, 226, 349, and 379.

This code is used in sections 1 and 2.

9. Raised errors may either be a C-'string'[1] when raised by an internal process or a *symbol* when raised at run-time.

If an error handler has been established then the *id* and *detail* are promoted to an *exception* object and the handler entered.

```
void handle_error(char *message, cell id, cell detail)
{
  int len;
  if (¬null_p(id)) {
    message = symbol_store(id);
    len = symbol_length(id);
  }
  else len = strlen(message);
  if (Error_Handler) {    /* TODO: Save Acc or rely on id being it? */
    vms_push(detail);
    if (null_p(id)) id = sym(message);
    Acc = atom(id, detail, FORMAT_EXCEPTION);
    vms_clear();
    longjmp(Goto_Error, 1);
  }
  printf("UNHANDLED␣ERROR:␣");
  for ( ; len−−; message++) putchar(*message);
  putchar(':');
  putchar('␣');
  write_form(detail, 0);
  printf("\n");
  longjmp(Goto_Begin, 1);
}
```

10. Run-time errors are raised by the OP_ERROR opcode which passes control to *handle_error* (and never returns). The code which compiles **error** to emit this opcode comes later after the compiler has been defined.

```
⟨ Opcode implementations 10 ⟩ ≡
case OP_ERROR:
  handle_error(Λ, Acc, rts_pop(1));
  break;    /* superfluous */
```

See also sections 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 154, 156, 157, and 158.

This code is used in section 108.

11. We additionally define *warn* here because where else is it going to go?

```
void warn(char *message, cell detail)
{
  printf("WARNING:␣%s:␣", message);
  write_form(detail, 0);
  printf("\n");
}
```

[1] C does not have strings, it has pointers to memory buffers that probably contain ASCII and might also happen to have a Λ in them somewhere.

12. Memory Management. The most commonly used data type in lisp-like languages is the *pair*, also called a "cons cell" for histerical raisins, which is a datum consisting of two equally-sized halves. For reasons that don't bear thinking about they are called the *car* for the "first" half and the *cdr* for the "second" half. In this code & document, **cell** refers to each half of a *pair*. **cell** is not used to refer to a whole cons cell in order to avoid confusion.

A *pair* in `LossLess` is stored in 2 equally-sized areas of memory. On 64-bit x86 implementations, which are all I'm considering at the moment, each half is 32 bits wide. Each pair additionally has an 8 bit tag (1 byte) associated with it, stored in a third array.

Internally a *pair* is represented by an offset into these memory areas. Negative numbers are therefore available for a few global constants.

The *pair*'s tag is treated as a bitfield. The garbage collector uses two bits (`TAG_MARK` and `TAG_STATE`). The other 6 bits are used to identify what data is stored in the **cell**s.

```
#define NIL −1      /* Not Λ, but not nil_p/nil? either */
#define FALSE −2    /* Yes, */
#define TRUE −3     /* really. */
#define END_OF_FILE −4    /* stdio has EOF */
#define VOID −5
#define UNDEFINED −6

#define TAG_NONE #00
#define TAG_MARK #80     /* GC mark bit */
#define TAG_STATE #40    /* GC state bit */
#define TAG_ACARP #20    /* CAR is a pair */
#define TAG_ACDRP #10    /* CDR is a pair */
#define TAG_FORMAT #3f   /* Mask lower 6 bits */

#define HEAP_SEGMENT #8000
```

⟨ Global variables 6 ⟩ +≡

```
cell *CAR = Λ;
cell *CDR = Λ;
char *TAG = Λ;
cell Cells_Free = NIL;
int Cells_Poolsize = 0;
int Cells_Segment = HEAP_SEGMENT;
```

13. ⟨ Externalised global variables 7 ⟩ +≡

```
extern cell *CAR, *CDR, Cells_Free;
extern char *TAG;
extern int Cells_Poolsize, Cells_Segment;
```

14. ⟨ Function declarations 8 ⟩ +≡

```
void new_cells_segment(void);
```

15. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ ≡

```
free(CAR);
free(CDR);
free(TAG);
CAR = CDR = Λ;
TAG = Λ;
Cells_Free = NIL;
Cells_Poolsize = 0;
Cells_Segment = HEAP_SEGMENT;
```

See also sections 23, 28, 34, 50, 60, 68, 94, 163, and 217.

This code is used in section 96.

16. The pool is spread across CAR, CDR and TAG and starts off with a size of zero **cell**s, growing by *Cells_Segment* **cell**s each time it's enlarged. When the heap is enlarged newly allocated memory is set to zero and the segment size set to half of the total pool size.

```
#define ERR_OOM "out-of-memory"
#define ERR_OOM_P(p) do { if ((p) ≡ Λ) error (ERR_OOM, NIL); } while (0)
#define ERR_DOOM_P(p, d) do { if ((p) ≡ Λ) error (ERR_OOM, (d)); } while (0)
#define enlarge_pool(p, m, t) do
        {
          void *n;
          n = LL_ALLOCATE((p), (m), sizeof (t));
          ERR_OOM_P(n);
          (p) = n;
        }
        while (0)

void new_cells_segment(void)
{
  enlarge_pool(CAR, Cells_Poolsize + Cells_Segment, cell);
  enlarge_pool(CDR, Cells_Poolsize + Cells_Segment, cell);
  enlarge_pool(TAG, Cells_Poolsize + Cells_Segment, char);
  bzero(CAR + Cells_Poolsize, Cells_Segment * sizeof(cell));
  bzero(CDR + Cells_Poolsize, Cells_Segment * sizeof(cell));
  bzero(TAG + Cells_Poolsize, Cells_Segment * sizeof(char));
  Cells_Poolsize += Cells_Segment;
  Cells_Segment = Cells_Poolsize/2;
}
```

17. Preprocessor directives provide precidates to interrogate a *pair*'s tag and find out what it is.

Although not all of these *cXr* macros are used they are all defined here for completeness (and it's easier than working out which ones really are needed).

#define *special_p*(p) ((p) < 0)
#define *boolean_p*(p) ((p) ≡ FALSE ∨ (p) ≡ TRUE)
#define *eof_p*(p) ((p) ≡ END_OF_FILE)
#define *false_p*(p) ((p) ≡ FALSE)
#define *null_p*(p) ((p) ≡ NIL)
#define *true_p*(p) ((p) ≡ TRUE)
#define *void_p*(p) ((p) ≡ VOID)
#define *undefined_p*(p) ((p) ≡ UNDEFINED)

#define *mark_p*(p) (¬*special_p*(p) ∧ (TAG[(p)] & TAG_MARK))
#define *state_p*(p) (¬*special_p*(p) ∧ (TAG[(p)] & TAG_STATE))
#define *acar_p*(p) (¬*special_p*(p) ∧ (TAG[(p)] & TAG_ACARP))
#define *acdr_p*(p) (¬*special_p*(p) ∧ (TAG[(p)] & TAG_ACDRP))
#define *mark_clear*(p) (TAG[(p)] &= ∼TAG_MARK)
#define *mark_set*(p) (TAG[(p)] |= TAG_MARK)
#define *state_clear*(p) (TAG[(p)] &= ∼TAG_STATE)
#define *state_set*(p) (TAG[(p)] |= TAG_STATE)
#define *format*(p) (TAG[(p)] & TAG_FORMAT)

#define *tag*(p) (TAG[(p)])

#define *car*(p) (CAR[(p)])
#define *cdr*(p) (CDR[(p)])
#define *caar*(p) (CAR[CAR[(p)]])
#define *cadr*(p) (CAR[CDR[(p)]])
#define *cdar*(p) (CDR[CAR[(p)]])
#define *cddr*(p) (CDR[CDR[(p)]])
#define *caaar*(p) (CAR[CAR[CAR[(p)]]])
#define *caadr*(p) (CAR[CAR[CDR[(p)]]])
#define *cadar*(p) (CAR[CDR[CAR[(p)]]])
#define *caddr*(p) (CAR[CDR[CDR[(p)]]])
#define *cdaar*(p) (CDR[CAR[CAR[(p)]]])
#define *cdadr*(p) (CDR[CAR[CDR[(p)]]])
#define *cddar*(p) (CDR[CDR[CAR[(p)]]])
#define *cdddr*(p) (CDR[CDR[CDR[(p)]]])
#define *caaaar*(p) (CAR[CAR[CAR[CAR[(p)]]]])
#define *caaadr*(p) (CAR[CAR[CAR[CDR[(p)]]]])
#define *caadar*(p) (CAR[CAR[CDR[CAR[(p)]]]])
#define *caaddr*(p) (CAR[CAR[CDR[CDR[(p)]]]])
#define *cadaar*(p) (CAR[CDR[CAR[CAR[(p)]]]])
#define *cadadr*(p) (CAR[CDR[CAR[CDR[(p)]]]])
#define *caddar*(p) (CAR[CDR[CDR[CAR[(p)]]]])
#define *cadddr*(p) (CAR[CDR[CDR[CDR[(p)]]]])
#define *cdaaar*(p) (CDR[CAR[CAR[CAR[(p)]]]])
#define *cdaadr*(p) (CDR[CAR[CAR[CDR[(p)]]]])
#define *cdadar*(p) (CDR[CAR[CDR[CAR[(p)]]]])
#define *cdaddr*(p) (CDR[CAR[CDR[CDR[(p)]]]])
#define *cddaar*(p) (CDR[CDR[CAR[CAR[(p)]]]])
#define *cddadr*(p) (CDR[CDR[CAR[CDR[(p)]]]])
#define *cdddar*(p) (CDR[CDR[CDR[CAR[(p)]]]])
#define *cddddr*(p) (CDR[CDR[CDR[CDR[(p)]]]])

18. Both *atoms* and cons cells are stored in *pair*s. The lower 6 bits of the tag define the format of data stored in that *pair*. The *atoms* are grouped into three types depending on whether both **cell**s point to another *pair*, whether only the *cdr* does, or whether both **cell**s are opaque. From this we obtain the core data types.

```
#define FORMAT_CONS (TAG_ACARP | TAG_ACDRP | #00)
#define FORMAT_APPLICATIVE (TAG_ACARP | TAG_ACDRP | #01)
#define FORMAT_OPERATIVE (TAG_ACARP | TAG_ACDRP | #02)
#define FORMAT_SYNTAX (TAG_ACARP | TAG_ACDRP | #03)
#define FORMAT_ENVIRONMENT (TAG_ACARP | TAG_ACDRP | #04)
#define FORMAT_EXCEPTION (TAG_ACARP | TAG_ACDRP | #05)
#define FORMAT_INTEGER (TAG_ACDRP | #00)  /* value : next/NIL */
#define FORMAT_SYMBOL (TAG_NONE | #00)  /* length : offset */
#define FORMAT_VECTOR (TAG_NONE | #01)  /* gc-index : offset */
#define FORMAT_COMPILER (TAG_NONE | #02)  /* offset : NIL */
#define atom_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≠ (TAG_ACARP | TAG_ACDRP)))
#define pair_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ (TAG_ACARP | TAG_ACDRP)))
#define applicative_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_APPLICATIVE))
#define compiler_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_COMPILER))
#define environment_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_ENVIRONMENT))
#define integer_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_INTEGER))
#define operative_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_OPERATIVE))
#define symbol_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_SYMBOL))
#define syntax_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_SYNTAX))
#define vector_p(p) (¬special_p(p) ∧ ((tag(p) & TAG_FORMAT) ≡ FORMAT_VECTOR))
```

19. Allocating a new *pair* may require garbage collection to be performed. If the data being put into either half of the new *pair* is itself a *pair* it may be discarded by the collector. To avoid this happening the data are saved into preallocated temporary storage while a new *pair* is being located.

```
⟨ Global variables 6 ⟩ +≡
  cell Tmp_CAR = NIL;
  cell Tmp_CDR = NIL;
```

20. ⟨ Externalised global variables 7 ⟩ +≡

```
  extern cell Tmp_CAR, Tmp_CDR;
```

21. ⟨ Protected Globals 21 ⟩ ≡

```
  &Tmp_CAR, &Tmp_CDR ,
```

See also sections 32, 49, 58, 93, 162, and 216.
This code is used in section 41.

22. ⟨ Function declarations 8 ⟩ +≡

```
  cell atom(cell, cell, char);
```

23. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡

```
  Tmp_CAR = Tmp_CDR = NIL;
```

24. **#define** *cons*(a,d) *atom*$((a),(d),$`FORMAT_CONS`$)$

```
cell atom(cell ncar, cell ncdr, char ntag)
{
  cell r;
  if (null_p(Cells_Free)) {
    if (ntag & TAG_ACARP) Tmp_CAR = ncar;
    if (ntag & TAG_ACDRP) Tmp_CDR = ncdr;
    if (gc() ≤ (Cells_Poolsize/2)) {
      new_cells_segment();
      gc();
    }
    Tmp_CAR = Tmp_CDR = NIL;
  }
  r = Cells_Free;
  Cells_Free = cdr(Cells_Free);
  car(r) = ncar;
  cdr(r) = ncdr;
  tag(r) = ntag;
  return r;
}
```

25. Vectors. A *vector* stores a contiguous sequence of **cell**s, each referring to a *pair* on the heap. Unlike *pair*s *vector*s are compacted during garbage collection to avoid fragmentation.

Storage is largely the same as **cell**s except for how the free pointer is maintained: an index into the next unused **cell** in VECTOR.

⟨ Global variables 6 ⟩ +≡
```
cell *VECTOR = Λ;
int Vectors_Free = 0;
int Vectors_Poolsize = 0;
int Vectors_Segment = HEAP_SEGMENT;
```

26. ⟨ Externalised global variables 7 ⟩ +≡
```
extern cell *VECTOR;
extern int Vectors_Free, Vectors_Poolsize, Vectors_Segment;
```

27. ⟨ Function declarations 8 ⟩ +≡
```
void new_vector_segment(void);
```

28. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡
```
free(VECTOR);
VECTOR = Λ;
Vectors_Free = Vectors_Poolsize = 0;
Vectors_Segment = HEAP_SEGMENT;
```

29.
```
void new_vector_segment(void)
{
    cell *new_vector;
    new_vector = LL_ALLOCATE(VECTOR, Vectors_Poolsize + Vectors_Segment, sizeof(cell));
    ERR_OOM_P(new_vector);
    bzero(new_vector + Vectors_Poolsize, Vectors_Segment * sizeof(cell));
    VECTOR = new_vector;
    Vectors_Poolsize += Vectors_Segment;
    Vectors_Segment = Vectors_Poolsize/2;
}
```

30. When a *pair* holds a *vector* its tag is FORMAT_VECTOR, the *car* is used by the garbage collecter and the *cdr* is an index into VECTOR.

Each *vector* contains 2 additional pieces of metadata (which are **above** the index), the length of the *vector* and a reference back to the *pair* holding the *vector*.

A *vector* of length 0 is treated as a global constant akin to NIL but it must be stored in a variable and created during initialisation.

```
#define VECTOR_CELL 0
#define VECTOR_SIZE 1
#define VECTOR_HEAD 2
#define vector_realsize(s) ((s) + VECTOR_HEAD)
#define vector_cell(v) (VECTOR[vector_offset(v) - (VECTOR_HEAD - VECTOR_CELL)])
#define vector_index(v) (car(v))
#define vector_length(v) (VECTOR[vector_offset(v) - (VECTOR_HEAD - VECTOR_SIZE)])
#define vector_offset(v) (cdr(v))
#define vector_ref(v, o) (VECTOR[vector_offset(v) + (o)])
```

⟨ Global variables 6 ⟩ +≡
```
cell Zero_Vector = NIL;
```

31. ⟨ Externalised global variables 7 ⟩ +≡
extern cell *Zero_Vector*;

32. ⟨ Protected Globals 21 ⟩ +≡
&*Zero_Vector* ,

33. ⟨ Global initialisation 3 ⟩ +≡
Zero_Vector = *vector_new_imp*(0, 0, 0);

34. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡
Zero_Vector = NIL;

35. Separate storage means separate garbage collection and a different allocator. *vector_new_imp*, again, is broadly similar to *atom* without the need for preallocated storage.

36. ⟨ Function declarations 8 ⟩ +≡

```
cell vector_new(int, cell);
cell vector_new_imp(int, boolean, cell);
cell vector_new_list(cell, int);
cell vector_sub(cell, int, int, int, int, cell);
```

37.

```
cell vector_new_imp(int size, boolean fill_p, cell fill)
{
  int wsize, off, i;
  cell r;
  wsize = vector_realsize(size);
  if (Vectors_Free + wsize ≥ Vectors_Poolsize) {
    gc_vectors();
    while (Vectors_Free + wsize ≥ (Vectors_Poolsize − (Vectors_Poolsize/2))) new_vector_segment();
  }
  r = atom(NIL, NIL, FORMAT_VECTOR);
  off = Vectors_Free;
  Vectors_Free += wsize;
  vector_offset(r) = off + VECTOR_HEAD;   /* must be first */
  vector_length(r) = size;
  vector_cell(r) = r;
  vector_index(r) = 0;
  if (fill_p)
    for (i = VECTOR_HEAD; i ≤ size + (VECTOR_HEAD − 1); i++)
      vector_ref(r, i) = fill;
  return r;
}
```

38.

```
cell vector_new(int size, cell fill)
{
  if (size ≡ 0) return Zero_Vector;
  return vector_new_imp(size, btrue, fill);
}
```

39. *vector_new_list* turns a *list* of *pair*s into a *vector*.

```
cell vector_new_list(cell list, int len)
{
  cell r;
  int i;
  r = vector_new(len, 0);
  for (i = 0; i < len; i++) {
    vector_ref(r, i) = car(list);
    list = cdr(list);
  }
  return r;
}
```

40. Although a little early in the narrative *vector_sub* is defined here because it's the only other function substantially dealing with *vector* data.

```
cell vector_sub(cell src, int srcfrom, int srcto, int dstfrom, int dstto, cell fill)
{
  cell dst;
  int copy, i;
  copy = srcto - srcfrom;
  if (dstto < 0) dstto = dstfrom + copy;
  dst = vector_new_imp(dstto, 0, 0);
  for (i = 0; i < dstfrom; i++) vector_ref(dst, i) = fill;
  for (i = srcfrom; i < srcto; i++)
    vector_ref(dst, (dstfrom - srcfrom) + i) = vector_ref(src, i);
  for (i = dstfrom + copy; i < dstto; i++) vector_ref(dst, i) = fill;
  return dst;
}
```

41. Garbage Collection. The garbage collector is a straightforward mark and sweep collector. *mark* is called for every entry in ROOTS to recursively set the mark bit on every reachable *pair*, then the whole pool is scanned and any *pair*s which aren't marked are added to the free list.

ROOTS is a Λ-terminated C array of objects to protect from collection. I can't think of any better way of declaring it but hard-coding it right here.

cell ∗ROOTS[ ] = {⟨ Protected Globals 21 ⟩, Λ};

42. ⟨ Externalised global variables 7 ⟩ +≡
extern cell ∗ROOTS;

43. ⟨ Function declarations 8 ⟩ +≡
int *gc*(**void**);
int *gc_vectors*(**void**);
void *mark*(**cell**);
int *sweep*(**void**);

44.
```
void mark(cell next)
{
  cell parent, prev;
  int i;

  parent = prev = NIL;
  while (1) {
    if (¬(special_p(next) ∨ mark_p(next))) {
      if (vector_p(next)) {      /* S0 → S.1 */
        mark_set(next);
        vector_cell(next) = next;
        if (vector_length(next) > 0) {
          state_set(next);
          vector_index(next) = 0;
          prev = vector_ref(next, 0);
          vector_ref(next, 0) = parent;
          parent = next;
          next = prev;
        }
      }
      else if (¬acar_p(next) ∧ acdr_p(next)) {      /* S0 → S2 */
        prev = cdr(next);
        cdr(next) = parent;
        parent = next;
        next = prev;
        mark_set(parent);
      }
      else if (acar_p(next)) {      /* S0 → S1 */
        prev = car(next);
        car(next) = parent;
        mark_set(next);
        parent = next;
        next = prev;
        state_set(parent);
      }
      else {      /* S0 → S1 */
        mark_set(next);
      }
    }
    else {
      if (null_p(parent)) break;
      if (vector_p(parent)) {      /* S.1 → S.1/done */
        i = vector_index(parent);
        if ((i + 1) < vector_length(parent)) {
          prev = vector_ref(parent, i + 1);
          vector_ref(parent, i + 1) = vector_ref(parent, i);
          vector_ref(parent, i) = next;
          next = prev;
          vector_index(parent) = i + 1;
        }
        else {      /* S.1 → done */
          state_clear(parent);
          prev = parent;
```

```
          parent = vector_ref(prev, i);
          vector_ref(prev, i) = next;
          next = prev;
        }
      }
      else if (state_p(parent)) {      /* S1 → S2 */
        prev = cdr(parent);
        cdr(parent) = car(parent);
        car(parent) = next;
        state_clear(parent);
        next = prev;
      }
      else if (acdr_p(parent)) {      /* S2 → done */
        prev = parent;
        parent = cdr(prev);
        cdr(prev) = next;
        next = prev;
      }
      else {
        error(ERR_UNIMPLEMENTED, NIL);
      }
    }
  }
}
int sweep(void)
{
  int count, i;
  Cells_Free = NIL;
  count = 0;
  for (i = 0; i < Cells_Poolsize; i++) {
    if (¬mark_p(i)) {
      tag(i) = TAG_NONE;
      cdr(i) = Cells_Free;
      Cells_Free = i;
      count++;
    }
    else {
      mark_clear(i);
    }
  }
  return count;
}
int gc(void)
{
  int sk, i;
  if (¬null_p(RTS)) {
    sk = vector_length(RTS);
    vector_length(RTS) = RTSp + 1;
  }
  for (i = 0; ROOTS[i]; i++) mark(*ROOTS[i]);
  for (i = SCHAR_MIN; i ≤ SCHAR_MAX; i++) mark(Small_Int[(unsigned char) i]);
```

```
    if (¬null_p(RTS)) vector_length(RTS) = sk;
    return sweep();
  }
```

45. *vector* garbage collection works by using the *pair*s garbage collector to scan ROOTS and determine which vectors are really in use then removes any which aren't from VECTORS, decrementing *Vectors_Free* if it can.

```
  int gc_vectors(void)
  {
    int to, from, d, i, r;
    ⟨ Unmark all vectors 46 ⟩
    gc();
    from = to = 0;
    while (from < Vectors_Free) {
      d = vector_realsize(VECTOR[from + VECTOR_SIZE]);
      if (¬null_p(VECTOR[from + VECTOR_CELL])) {
        if (to ≠ from) {
          for (i = 0; i < d; i++) VECTOR[to + i] = VECTOR[from + i];
          vector_offset(VECTOR[to + VECTOR_CELL]) = to + VECTOR_HEAD;
        }
        to += d;
      }
      from += d;
    }
    r = Vectors_Free − to;
    Vectors_Free = to;
    return r;
  }
```

46. To "unmark" a *vector*, all the links in VECTOR back to the cell which refers to it (*vector_cell*) are set to NIL. *gc* will re-set the link in any vectors that it can reach.

```
⟨ Unmark all vectors 46 ⟩ ≡
  i = 0;
  while (i < Vectors_Free) {
    VECTOR[i + VECTOR_CELL] = NIL;
    i += vector_realsize(VECTOR[i + VECTOR_SIZE]);
  }
```

This code is used in section 45.

47. Objects. Although not `objects` per se, the first `objects` which will be defined are three stacks. We could define the run-time stack later because it's not used until the virtual machine is implemented but the implementations mirror each other and the internal VM stack is required before real objects can be defined. Also the runtime stack uses the VM stack in its implementation.

The compiler stack is included here because it's identical to the VM stack.

The VM stack is a pointer to the head of a *list*. This means that accessing the top few elements of the stack—especially pushing and popping a single `object`—is effectively free but accessing an arbitrary part of the stack requires an expensive walk over each item in turn.

On the other hand the run-time stack is stored in a *vector* with a pointer *RTSp* to the current head of the stack, which is -1 if the stack is empty.

This has the obvious disadvantage that its storage space is finite and occasionally the whole stack will need to be copied into a new, larger *vector* (and conversely it may waste space or require occasional trimming). On the other hand random access to any part of the stack has the same (negligable) cost.

When it's not ambiguous "stack" in this document refers to the run-time stack; the VM stack is an implementation detail. In fact the run-time stack is also an implementation detail but the VM stack is an implementation detail of that implementation detail; do you like recursion yet?.

The main interface to each stack is its *push*/*pop*/*ref*/*clear* functions. There are some additional handlers for the run-time stack.

```
#define ERR_UNDERFLOW "underflow"
#define ERR_OVERFLOW "overflow"
#define CHECK_UNDERFLOW(s) if (null_p(s)) error (ERR_UNDERFLOW, VOID)
#define RTS_UNDERFLOW(p) if ((p) < −1) error (ERR_UNDERFLOW, RTS)
#define RTS_OVERFLOW(p) if ((p) > RTSp) error (ERR_OVERFLOW, RTS)
```

⟨ Global variables 6 ⟩ +≡

```
cell CTS = NIL;
cell RTS = NIL;
cell VMS = NIL;
int RTS_Size = 0;
int RTSp = −1;
```

48. ⟨ Externalised global variables 7 ⟩ +≡

```
extern cell CTS, RTS, VMS;
extern int RTS_Size, RTSp;
```

49. ⟨ Protected Globals 21 ⟩ +≡

```
&CTS, &RTS, &VMS ,
```

50. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡

```
CTS = RTS = VMS = NIL;
RTS_Size = 0;
RTSp = −1;
```

51. ⟨ Function declarations 8 ⟩ +≡

```
cell cts_pop(void);
void cts_push(cell);
cell cts_ref(void);
void cts_set(cell);
cell rts_pop(int);
void rts_prepare(int);
void rts_push(cell);
cell rts_ref(int);
cell rts_ref_abs(int);
void rts_set(int, cell);
void rts_set_abs(int, cell);
cell vms_pop(void);
void vms_push(cell);
cell vms_ref(void);
void vms_set(cell);
```

52. The VM and compiler stacks VMS and CTS are built on *lists*.

```
#define vms_clear() ((void) vms_pop())
  cell vms_pop(void)
  {
    cell r;
    CHECK_UNDERFLOW(VMS);
    r = car(VMS);
    VMS = cdr(VMS);
    return r;
  }
  void vms_push(cell item)
  { VMS = cons(item, VMS); }
  cell vms_ref(void)
  {
    CHECK_UNDERFLOW(VMS);
    return car(VMS);
  }
  void vms_set(cell item)
  {
    CHECK_UNDERFLOW(VMS);
    car(VMS) = item;
  }
```

53. CTS is treated identically to VMS. Using the C preprocessor for this would be unnecessarily inelegant so instead here is a delicious bowl of pasta.

```
#define cts_clear() ((void) cts_pop())
#define cts_reset() CTS = NIL

  cell cts_pop()
  {
    cell r;

    CHECK_UNDERFLOW(CTS);
    r = car(CTS);
    CTS = cdr(CTS);
    return r;
  }

  void cts_push(cell item)
  { CTS = cons(item, CTS); }

  cell cts_ref(void)
  {
    CHECK_UNDERFLOW(CTS);
    return car(CTS);
  }

  void cts_set(cell item)
  {
    CHECK_UNDERFLOW(CTS);
    car(CTS) = item;
  }
```

54. Being built on a *vector* the run-time stack needs to increase its size when it's full. Functions can call *rts_prepare* to ensure that the stack is big enough for their needs.

```
#define RTS_SEGMENT #1000

  void rts_prepare(int need)
  {
    int b, s;

    if (RTSp + need ≥ RTS_Size) {
      b = RTS_SEGMENT * ((need + RTS_SEGMENT)/RTS_SEGMENT);
      s = RTS_Size + b;
      RTS = vector_sub(RTS, 0, RTS_Size, 0, s, UNDEFINED);
      RTS_Size = s;
    }
  }
```

55. Otherwise, the run-time stack has the same interface but a different implementation.

```
#define rts_clear(c) ((void) rts_pop(c))
#define rts_reset() Fp = RTSp = -1;
  cell rts_pop(int count)
  {
    RTS_UNDERFLOW(RTSp - count);
    RTSp -= count;
    return vector_ref(RTS, RTSp + 1);
  }
  void rts_push(cell o)
  {
    vms_push(o);
    rts_prepare(1);
    vector_ref(RTS, ++RTSp) = vms_pop();
  }
  cell rts_ref(int d)
  {
    RTS_UNDERFLOW(RTSp - d);
    RTS_OVERFLOW(RTSp - d);
    return vector_ref(RTS, RTSp - d);
  }
  cell rts_ref_abs(int d)
  {
    RTS_UNDERFLOW(d);
    RTS_OVERFLOW(d);
    return vector_ref(RTS, d);
  }
  void rts_set(int d, cell v)
  {
    RTS_UNDERFLOW(RTSp - d);
    RTS_OVERFLOW(RTSp - d);
    vector_ref(RTS, RTSp - d) = v;
  }
  void rts_set_abs(int d, cell v)
  {
    RTS_UNDERFLOW(d);
    RTS_OVERFLOW(d);
    vector_ref(RTS, d) = v;
  }
```

56. Symbols. With the basics in place, the first thing to define is *symbol*s; they're not needed yet but everything becomes easier with them extant and they depend on nothing but themselves since they are themselves.

*symbol*s are never garbage collected. This was not a conscious decision it just doesn't seem like it matters. Instead, every *symbol* once created is immediately added to the *Symbol_Table list*. When a reference to a *symbol* is requested, the object in this *list* is returned.

Eventually this should implement a hash table but I'm not making one of those this morning.

Owing to the nasty c-to-perl-to-c route that I've taken, combined with plans for vector/byte storage, the storage backing symbols is going to be hairy without explanation (for now it's a mini duplicate of vector storage).

```
#define sym(s) symbol((s), 1)
#define symbol_length car
#define symbol_offset cdr
#define symbol_store(s) (SYMBOL + symbol_offset(s))
```

⟨ Global variables 6 ⟩ +≡

```
cell Symbol_Table = NIL;
char *SYMBOL = Λ;
int Symbol_Free = 0;
int Symbol_Poolsize = 0;
```

57. ⟨ Externalised global variables 7 ⟩ +≡

```
extern cell Symbol_Table;
extern char *SYMBOL;
extern int Symbol_Free, Symbol_Poolsize;
```

58. ⟨ Protected Globals 21 ⟩ +≡

```
&Symbol_Table ,
```

59. ⟨ Function declarations 8 ⟩ +≡

```
cell symbol(char *, boolean);
void symbol_expand(void);
void symbol_reify(cell);
boolean symbol_same_p(cell, cell);
cell symbol_steal(char *);
```

60. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡

```
free(SYMBOL);
SYMBOL = Λ;
Symbol_Poolsize = Symbol_Free = 0;
Symbol_Table = NIL;
```

61.

```
void symbol_expand(void)
{
  char *new;
  new = realloc(SYMBOL, Symbol_Poolsize + HEAP_SEGMENT);
  ERR_OOM_P(new);
  Symbol_Poolsize += HEAP_SEGMENT;
  SYMBOL = new;
}
```

62. A *symbol* can "steal" storage from SYMBOL which results in an object which can be mostly treated like a normal *symbol*, used to compare a potentially new *symbol* with those currently stored in *Symbol_Table*. This is the closest that *symbol*s get to being garbage collected.

```
cell symbol_steal(char *cstr)
{
  cell r;
  int len;
  len = strlen(cstr);
  while (Symbol_Free + len > Symbol_Poolsize) symbol_expand();
  r = atom(len, Symbol_Free, FORMAT_SYMBOL);
  memcpy(SYMBOL + Symbol_Free, cstr, len);     /* Symbol_Free is not incremented here */
  return r;
}
```

63. Temporary *symbol*s compare byte-by-byte with existing *symbol*s. This is not efficient at all.

```
boolean symbol_same_p(cell maybe, cell match)
{
  char *pmaybe, *pmatch;
  int i, len;
  len = symbol_length(match);
  if (symbol_length(maybe) ≠ len) return bfalse;
  pmaybe = symbol_store(maybe);
  pmatch = symbol_store(match);
  if (maybe ≡ match)     /* This shouldn't happen */
    return btrue;
  for (i = 0; i < len; i++) {
    if (pmaybe[i] ≠ pmatch[i]) return bfalse;
  }
  return btrue;
}
```

64.
```
void symbol_reify(cell s)
{
  Symbol_Free += symbol_length(s);
  Symbol_Table = cons(s, Symbol_Table);
}
```

65.
```
cell symbol(char *cstr, boolean permanent_p)
{
  cell st, s;
  s = symbol_steal(cstr);
  st = Symbol_Table;
  while (¬null_p(st)) {
    if (symbol_same_p(s, car(st))) return car(st);
    st = cdr(st);
  }
  if (permanent_p) symbol_reify(s);
  return s;
}
```

66. Numbers. The only numbers supported by this early implementation of LossLess are signed integers that fit in a single **cell** (ie. 32-bit integers).

The 256 numbers closest to 0 (ie. from $-$#80 to $+$#7f) are preallocated during initialisation. If you live in a parallel universe where the char type isn't 8 bits then adjust those numbers accordingly.

```
#define fixint_p(p) (integer_p(p) ∧ null_p(int_next(p)))
#define smallint_p(p) (fixint_p(p) ∧ int_value(p) ≥ SCHAR_MIN ∧ int_value(p) ≤ SCHAR_MAX)
#define int_value(p) ((int)(car(p)))
#define int_next cdr
```

⟨ Global variables 6 ⟩ +≡

```
cell Small_Int[UCHAR_MAX + 1];
```

67. ⟨ Externalised global variables 7 ⟩ +≡

```
extern cell *Small_Int;
```

68. Even though the *Small_Int* objects are about to be created, in order to create objects garbage collection will happen and assume that *Small_Int* has already been initialised and attempt to protect data which don't exist from collection. This is a silly solution but I'm leaving it alone until I have a better memory model.

⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡

```
for (i = 0; i < 256; i++) Small_Int[i] = NIL;
```

69. ⟨ Global initialisation 3 ⟩ +≡

```
for (i = SCHAR_MIN; i ≤ SCHAR_MAX; i++)
  Small_Int[(unsigned char) i] = int_new_imp(i, NIL);
```

70. As with *vectors*, *int_new* checks whether it should return an *object* from *Small_Int* or build a new one.

⟨ Function declarations 8 ⟩ +≡

```
cell int_new_imp(int, cell);
cell int_new(int);
```

71.

```
cell int_new_imp(int value, cell next)
{
  if (¬null_p(next)) error (ERR_UNIMPLEMENTED, NIL);
  return atom((cell) value, next, FORMAT_INTEGER);
}
```

72.

```
cell int_new(int value)
{
  if (value ≥ SCHAR_MIN ∧ value ≤ SCHAR_MAX)
    return Small_Int[(unsigned char) value];
  return int_new_imp(value, NIL);
}
```

73. Pairs & Lists. Of course *pair*s—and so by definition *list*s—have already been implemented but so far only enough to implement core features. Here we define handlers for operations specifically on *list* `objects`.

First to count it's length a *list* is simply walked from head to tail. It is not considered an error if the *list* is improper (or not a *list* at all). To indicate this case the returned length is negated.

```
⟨ Function declarations 8 ⟩ +≡
  int list_length(cell);
  predicate list_p(cell, predicate, cell ∗);
  cell list_reverse_m(cell, boolean);
```

74.
```
int list_length(cell l)
{
  int c = 0;
  if (null_p(l)) return 0;
  for ( ; pair_p(l); l = cdr(l)) c++;
  if (¬null_p(l)) c = −(c + 1);
  return c;
}
```

75. A *list* is either NIL or a pair with one restriction, that its *cdr* must itself be a *list*. The size of the *list* is also counted to avoid walking it twice but nothing uses that (yet?).

```
predicate list_p(cell o, predicate improper_p, cell ∗sum)
{
  int c = 0;
  if (null_p(o)) {
    if (sum ≠ Λ) ∗sum = int_new(0);
    return TRUE;
  }
  while (pair_p(o)) {
    o = cdr(o);
    c++;
  }
  if (sum ≠ Λ) ∗sum = int_new(c);
  if (null_p(o)) return TRUE;
  if (sum ≠ Λ) ∗sum = int_new(−(c + 1));
  return improper_p;
}
```

76. A proper *list* can be reversed simply into a new *list*.

```
#define ERR_IMPROPER_LIST "improper-list"
  cell list_reverse(cell l, cell *improper, cell *sum)
  {
    cell saved, r;
    int c;
    saved = l;
    c = 0;
    vms_push(NIL);
    while (¬null_p(l)) {
      if (¬pair_p(l)) {
        r = vms_pop();
        if (improper ≠ Λ) {
          *improper = l;
          if (sum ≠ Λ) *sum = c;
          return r;
        }
        else error (ERR_IMPROPER_LIST, saved);
      }
      vms_set(cons(car(l), vms_ref()));
      l = cdr(l);
      c++;
    }
    if (sum ≠ Λ) *sum = int_new(c);
    return vms_pop();
  }
```

77. Reversing a *list* in-place means maintaining a link to the previous *pair* (or NIL) and replacing each *pair*'s *cdr*. The new head *pair* is returned, or FALSE if the *list* turned out to be improper.

```
  cell list_reverse_m(cell l, boolean error_p)
  {
    cell m, t, saved;
    saved = l;
    m = NIL;
    while (¬null_p(l)) {
      if (¬pair_p(l)) {
        if (¬error_p)    /* TODO: repair? */
          return FALSE;
        error (ERR_IMPROPER_LIST, saved);
      }
      t = cdr(l);
      cdr(l) = m;
      m = l;
      l = t;
    }
    return m;
  }
```

78. Environments. In order to associate a value with a *symbol* (a variable) they are paired together in an *environment*.

Like an onion or an ogre[1], an *environment* has layers. The top layer is both the current layer and the current *environment*. The bottom layer is the root *environment* *Root*.

An *environment* is stored in an *atom* with the *car* pointing to the previous layer (or `NIL` in the root *environment*).

The *cdr* is a *list* of association *pair*s representing the variables in that layer. An association *pair* is a proper *list* with two items: an identifier, in this case a *symbol*, and a value.

environment-handling functions and macros are generally named "env".

```
#define ERR_BOUND "already-bound"
#define ERR_UNBOUND "unbound"

#define env_empty() atom(NIL, NIL, FORMAT_ENVIRONMENT)
#define env_extend(e) atom((e), NIL, FORMAT_ENVIRONMENT)
#define env_layer cdr
#define env_parent car
#define env_empty_p(e) (environment_p(e) ∧ null_p(car(e)) ∧ null_p(cdr(e)))
#define env_root_p(e) (environment_p(e) ∧ null_p(car(e)))
```

79. Searching through an *environment* starts at its top layer and walks along each *pair*. If it encounters a *pair* who's *symbol* matches, the value is returned. If not then the search repeats layer by layer until the *environment* is exhausted and `UNDEFINED` is returned.

env_search does not raise an error if a *symbol* isn't found. This means that `UNDEFINED` is the only value which cannot be stored in a variable as there is no way to distinguish its return from this function.

⟨ Function declarations 8 ⟩ +≡

```
cell env_here(cell, cell);
cell env_lift_stack(cell, int, cell);
cell env_search(cell, cell);
void env_set(cell, cell, cell, boolean);
```

80.

```
cell env_search(cell haystack, cell needle)
{
  cell n;

  for ( ; ¬null_p(haystack); haystack = env_parent(haystack))
    for (n = env_layer(haystack); ¬null_p(n); n = cdr(n))
      if (caar(n) ≡ needle) return cadar(n);
  return UNDEFINED;
}
```

81.

```
cell env_here(cell haystack, cell needle)
{
  cell n;

  for (n = env_layer(haystack); ¬null_p(n); n = cdr(n))
    if (caar(n) ≡ needle) return cadar(n);
  return UNDEFINED;
}
```

[1] Or a cake.

82. To set a variable's value the *environment*'s top layer is first searched to see if the *symbol* is already bound. An **error** is raised if the symbol is bound (when running on behalf of *define!*) or not bound (when running on behalf of *set!*).

```
void env_set(cell e, cell name, cell value, boolean new_p)
{
  cell ass, t;
  ass = cons(name, cons(value, NIL));
  if (new_p) {⟨Mutate if unbound 84⟩}
  else {⟨Mutate if bound 83⟩}
}
```

83. Updating an already-bound variable means removing the existing binding from the *environment* and inserting the new binding. During the walk over the layer t is one pair ahead of the pair being considered so that when *name* is found t's *cdr* can be changed, snipping the old binding out, so the first pair is checked specially.

```
⟨Mutate if bound 83⟩ ≡
  if (null_p(env_layer(e))) error (ERR_UNBOUND, name);
  if (caar(env_layer(e)) ≡ name) {
    env_layer(e) = cons(ass, cdr(env_layer(e)));
    return;
  }
  for (t = env_layer(e); ¬null_p(cdr(t)); t = cdr(t)) {
    if (caadr(t) ≡ name) {
      cdr(t) = cddr(t);
      env_layer(e) = cons(ass, env_layer(e));
      return;
    }
  }
  error (ERR_UNBOUND, name);
```

This code is used in section 82.

84. The case is simpler if the *name* must **not** be bound already as the new binding can be prepended to the layer after searching with no need for special cases.

```
⟨Mutate if unbound 84⟩ ≡
  for (t = env_layer(e); ¬null_p(t); t = cdr(t))
    if (caar(t) ≡ name) error (ERR_BOUND, name);
  env_layer(e) = cons(ass, env_layer(e));
```

This code is used in section 82.

85. Values are passed to functions on the stack. *env_lift_stack* moves these values from the stack into an *environment*.

```
cell env_lift_stack(cell e, int nargs, cell formals)
{
  cell p, name, value, ass;
  vms_push(env_extend(e));
  p = NIL;    /* prepare a new layer */
  vms_push(p);
  while (nargs--) {
    if (pair_p(formals)) {
      name = car(formals);
      formals = cdr(formals);
    }
    else { name = formals; }
    value = rts_pop(1);
    if (¬null_p(name)) {
      ass = cons(name, cons(value, NIL));
      vms_set((p = cons(ass, p)));
    }
  }
  vms_pop();
  cdr(vms_ref()) = p;    /* place the new layer in the extended environment */
  return vms_pop();
}
```

86. Closures & Compilers. Finally we have data structures to save run-time state: *closure*s. The way the compiler and virtual machine work to get *closure* objects built is described below—here is only a description of their backing stores.

LossLess has two types of *closure*, *applicative* and *operative*. They store the same data in identical containers; the difference is in how they're used.

The data required to define a *closure* are a program & the *environment* to run it in. A *closure* in LossLess also contains the formals given in the **lambda** or **vov** expression that was used to define it.

Program code in LossLess is stored as compiled bytecode in a *vector* with an instruction pointer indicating the entry point (0 is not implied). The *closure*s then look like this:

$$(APPLICATIVE\ \langle formals\rangle\ \langle environment\rangle\ \langle code\rangle\ \langle pointer\rangle)$$

$$(OPERATIVE\ \langle formals\rangle\ \langle environment\rangle\ \langle code\rangle\ \langle pointer\rangle)$$

However the *environment*, code and pointer are never referred to directly until the closure is unpicked by `OP_APPLY`/`OP_APPLY_TAIL`. Instead the objects effectively look like this:

$$(A|O\ \langle formals\rangle\ .\ \langle opaque\_closure\rangle)$$

```
#define applicative_closure cdr
#define applicative_formals car
#define applicative_new(f, e, p, i) closure_new_imp(FORMAT_APPLICATIVE, (f), (e), (p), (i))
#define operative_closure cdr
#define operative_formals car
#define operative_new(f, e, p, i) closure_new_imp(FORMAT_OPERATIVE, (f), (e), (p), (i))
```

⟨ Function declarations 8 ⟩ +≡

```
cell closure_new_imp(char, cell, cell, cell, cell);
```

87.

```
cell closure_new_imp(char ntag, cell formals, cell env, cell prog, cell ip)
{
  cell r;
  r = cons(int_new(ip), NIL);
  r = cons(prog, r);
  r = cons(env, r);
  return atom(formals, r, ntag);
}
```

88. Other than closures, and required in order to make them, the evaluator uses "*compiler*" `objects` that compile LossLess source code to VM bytecode. Each *compiler* is described in the structure *primitive*, containing the native function pointer to it.

```
#define compiler_cname(c) COMPILER[car(c)].name
#define compiler_fn(c) COMPILER[car(c)].fn
```

⟨ Type definitions 5 ⟩ +≡

```
typedef void (*native)(cell, cell, boolean);
typedef struct {
  char *name;
  native fn;
} primitive;
```

89. The contents of `COMPILER` are populated by the C compiler of this source. During initialisation *Root* then becomes the root environment filled with an association *pair* for each one.

$\langle$ Global variables 6 $\rangle$ +≡
primitive `COMPILER`[ ] = { $\langle$ List of opcode primitives 377 $\rangle$, $\{\Lambda, \Lambda\}$ } ;

90. $\langle$ Externalised global variables 7 $\rangle$ +≡
extern primitive ∗`COMPILER`;

91. Virtual Machine. This implementation of `LossLess` compiles user source code to an internal bytecode representation which is then executed sequentially by a virtual machine (VM).

Additionally to the myriad stacks already mentioned, the VM maintains (global!) state primarily in 6 registers. Two of these are simple flags (booleans) which indicate whether interpretation should continue.

1. *Running* is a flag raised (1) when the VM begins and lowered by user code to indicate that it should halt cleanly. This flag is checked on the beginning of each iteration of the VM's main loop.

2. *Interrupt* is normally lowered (0) and is raised in response to external events such as a unix signal. Long-running operations—especially those which could potentially run unbounded—check frequently for the state of this flag and abort and return immediately when it's raised.

The other four registers represent the computation.

3. *Acc* is the accumulator. Opcodes generally read and/or write this register to do their work. This is where the final result of computation will be found.

4. *Env* holds the current *environment*. Changing this is the key to implementing *closure*s.

5. *Prog* is the compiled bytecode of the currently running computation, a *vector* of VM opcodes with their in-line arguments.

6. *Ip* is the instruction pointer. This is an **int**, not a **cell** and must be boxed to be used outside of the VM.

Root and *Prog_Main* are also defined here which hold, respectively, the root *environment* and the virtual machine's starting program.

⟨ Global variables 6 ⟩ +≡
```
boolean Interrupt = 0;
boolean Running = 0;
cell Acc = NIL;
cell Env = NIL;
cell Prog = NIL;
cell Prog_Main = NIL;
cell Root = NIL;
int Ip = 0;
```

92. ⟨ Externalised global variables 7 ⟩ +≡
```
extern boolean Interrupt, Running;
extern cell Acc, Env, Prog, Prog_Main, Root;
extern int Ip;
```

93. ⟨ Protected Globals 21 ⟩ +≡
```
&Acc, &Env, &Prog, &Prog_Main, &Root ,
```

94. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡
```
Acc = Env = Prog = Prog_Main = Root = NIL;
Interrupt = Running = Ip = 0;
```

95. The `LossLess` virtual machine is initialised by calling the code snippets built into the ⟨ Global initialisation 3 ⟩ section then constructing the the root *environment* in *Root*.

Initialisation is divided into two phases. The first in *vm_init* sets up emergency jump points (which should never be reached) for errors which occur during initialisation or before the second phase.

The second phase establishes a jump buffer in *Goto_Begin* to support run-time errors that were not handled. It resets VM state which will not have had a chance to recover normally due to the computation aborting early.

The error handler's jump buffer *Goto_Error* on the other hand is established by *interpret* and does *not* reset any VM state, but does return to the previous jump buffer if the handler fails.

```
#define vm_init() do
        {
          if (setjmp(Goto_Begin)) {
            Acc = sym("ABORT");
            return EXIT_FAILURE;
          }
          if (setjmp(Goto_Error)) {
            Acc = sym("ABORT");
            return EXIT_FAILURE;
          }
          vm_init_imp();
        }
        while (0)
#define vm_prepare() do
        {
          setjmp(Goto_Begin);
          vm_prepare_imp();
        }
        while (0)
#define vm_runtime() do
        {
          if (setjmp(Goto_Error)) {
            Ip = −1;     /* TODO: call the handler */
            if (Ip < 0) longjmp(Goto_Begin, 1);
          }
        }
        while (0)
```

⟨ Function declarations 8 ⟩ +≡

```
  void vm_init_imp(void);
  void vm_prepare_imp(void);
  void vm_reset(void);
```

96.

```
void vm_init_imp(void)
{
  cell t;
  int i;
  primitive *n;
  ⟨ Pre-initialise Small_Int & other gc-sensitive buffers 15 ⟩
  ⟨ Global initialisation 3 ⟩
  Prog_Main = compile_main();
  i = 0;
  Root = atom(NIL, NIL, FORMAT_ENVIRONMENT);
  for (n = COMPILER + i; n→fn ≠ Λ; n = COMPILER + (++i)) {
    t = atom(i, NIL, FORMAT_COMPILER);
    t = cons(t, NIL);
    t = cons(sym(n→name), t);
    env_layer(Root) = cons(t, env_layer(Root));
  }
  Env = Root;
}
```

97.

```
void vm_prepare_imp(void)
{
  Acc = Prog = NIL;
  Env = Root;
  rts_reset();
}
```

98.

```
void vm_reset(void)
{
  Prog = Prog_Main;
  Running = Interrupt = Ip = 0;
}
```

99. Frames. The VM enters a *closure*—aka. calls a function—by appending a *frame* header to the stack. A *frame* consists of any work-in-progress items on the stack followed by a fixed-size header. A *frame*'s header captures the state of computation at the time it's created which is what lets another subroutine run and then return. The *frame* header contains 4 objects: ≪*Ip Prog Env Fp*≫.

Fp is a quasi-register which points into the stack to the current *frame*'s header. It's saved when entering a *frame* and its value set to that of the stack pointer *RTSp*. *RTSp* is restored to the saved value when returning from a *frame*.

```
#define FRAME_HEAD 4
#define frame_ip(f) rts_ref_abs((f) + 1)
#define frame_prog(f) rts_ref_abs((f) + 2)
#define frame_env(f) rts_ref_abs((f) + 3)
#define frame_fp(f) rts_ref_abs((f) + 4)
#define frame_set_ip(f, v) rts_set_abs((f) + 1, (v));
#define frame_set_prog(f, v) rts_set_abs((f) + 2, (v));
#define frame_set_env(f, v) rts_set_abs((f) + 3, (v));
#define frame_set_fp(f, v) rts_set_abs((f) + 4, (v));
```

⟨ Global variables 6 ⟩ +≡

```
int Fp = −1;
```

100. ⟨ Externalised global variables 7 ⟩ +≡

```
extern int Fp;
```

101. ⟨ Global initialisation 3 ⟩ +≡

```
Fp = −1;
```

102. Creating a *frame* is pushing the header items onto the stack. Entering it is changing the VM's registers that are now safe. This is done in two stages for some reason.

⟨ Function declarations 8 ⟩ +≡

```
void frame_consume(void);
void frame_enter(cell, cell, cell);
void frame_leave(void);
void frame_push(int);
```

103.

```
void frame_push(int ipdelta)
{
  rts_push(int_new(Ip + ipdelta));
  rts_push(Prog);
  rts_push(Env);
  rts_push(int_new(Fp));
}
```

104.

```
void frame_enter(cell e, cell p, cell i)
{
  Env = e;
  Prog = p;
  Ip = i;
  Fp = RTSp − FRAME_HEAD;
}
```

105. Leaving a *frame* means restoring the registers that were saved in it by *frame_push* and then returning *RTSp* and *Fp* to their previous values; *Fp* from the header and *RTSp* as the current *Fp* minus the *frame* header in case there were previously any in-progress items on top of the stack.

```
void frame_leave(void)
{
  int prev;
  Ip = int_value(frame_ip(Fp));
  Prog = frame_prog(Fp);
  Env = frame_env(Fp);
  prev = int_value(frame_fp(Fp));
  rts_clear(FRAME_HEAD);
  Fp = prev;
}
```

106. Tail Recursion. TODO

This is a straight copy of what I wrote in perl which hasn't been used there. Looks about right. Might work.

```
void frame_consume(void)
{
  int src, dst, i;
  src = Fp;
  dst = int_value(frame_fp(src));    /* Copy the parts of the old frame header that are needed */
  frame_set_prog(src, frame_prog(dst));
  frame_set_ip(src, frame_ip(dst));
  frame_set_fp(src, frame_fp(dst));    /* Move the active frame over the top of the previous one */
  for (i = 1; i <= FRAME_HEAD; i++)
    rts_set_abs(dst + i, rts_ref_abs(src + i));
  rts_clear(src - dst);
  Fp -= src - dst;
}
```

107. Interpreter. The workhorse of the virtual machine is *interpret*. After being reset with *vm_reset*, parsed (but not compiled) source code is put into *Acc* and the VM can be started by calling *interpret*.

```
⟨ Function declarations 8 ⟩ +≡
  void interpret(void);
```

108.

```
#define ERR_INTERRUPTED "interrupted"
  void interpret(void)
  {
    int ins;
    cell tmp;     /* not saved in ROOTS */
    vm_runtime();
    Running = 1;
    while (Running ∧ ¬Interrupt) {
      ins = int_value(vector_ref(Prog, Ip));
      switch (ins) {
        ⟨ Opcode implementations 10 ⟩
#ifdef LL_TEST
        ⟨ Testing implementations 220 ⟩
#endif
      }
    }
    if (Interrupt) error(ERR_INTERRUPTED, NIL);
  }
```

109. I/O. Before embarking on the meat of the interpreter a final detour to describe routines to parse a string (or stream) of source code into s-expressions, and because it's useful to see what's being done routines to write them back again.

These routines use C's *stdio* for now to get a simple implementation finished.

110. Reader (or Parser). The s-expression reader is an ad-hoc LALR parser; a single byte is read to determine which type of form to parse. Bytes are then read one at a time to validate the syntax and create the appropriate `object`.

The reading routines call into themselves recursively (for which it cheats and relies on C's stack). To prevent it running out of control *Read_Level* records the recursion depth and *read_form* aborts if it exceeds `READER_MAX_DEPTH`.

The compiler's rather than the VM's stack is used for temporary storage so that error handling doesn't need to clean it up. This is safe provided the reader and compiler are never used simultaneously.

The parser often needs the byte that was used to determine which kind of form to parse (the one that was "looked ahead" at). *Putback* is a small buffer to contain this byte. In fact this buffer can hold *two* bytes to accomodate lisp's *unquote-splicing* operator ⟪,@⟫.

In order to perform tests of this primitive implementation the reader can be directed to "read" from a C-strings if *Read_Pointer* is set to a value other than Λ.

```
#define ERR_RECURSION "recursion"
#define ERR_UNEXPECTED "unexpected"
#define WARN_AMBIGUOUS_SYMBOL "ambiguous"
#define READER_MAX_DEPTH 1024 /* gotta pick something */
#define READ_SPECIAL −10
#define READ_DOT −10 /* ⟪.⟫ */
#define READ_CLOSE_BRACKET −11 /* ⟪]⟫ */
#define READ_CLOSE_PAREN −12 /* ⟪)⟫ */
#define SYNTAX_DOTTED "dotted" /* ⟪.⟫ */
#define SYNTAX_QUOTE "quote" /* ⟪'⟫ */
#define SYNTAX_QUASI "quasiquote" /* ⟪`⟫ */
#define SYNTAX_UNQUOTE "unquote" /* ⟪,⟫ */
#define SYNTAX_UNSPLICE "unquote-splicing" /* ⟪,@⟫ */
```

```
⟨ Global variables 6 ⟩ +≡
  char Putback[2] = {'\0', '\0'};
  int Read_Level = 0;
  char *Read_Pointer = Λ;
  cell Sym_ERR_UNEXPECTED = NIL;
  cell Sym_SYNTAX_DOTTED = NIL;
  cell Sym_SYNTAX_QUASI = NIL;
  cell Sym_SYNTAX_QUOTE = NIL;
  cell Sym_SYNTAX_UNQUOTE = NIL;
  cell Sym_SYNTAX_UNSPLICE = NIL;
```

111. ⟨ Externalised global variables 7 ⟩ +≡

```
  extern char Putback[2], *Read_Pointer;
  extern int Read_Level;
  extern cell Sym_ERR_UNEXPECTED, Sym_SYNTAX_DOTTED, Sym_SYNTAX_QUASI;
  extern cell Sym_SYNTAX_QUOTE, Sym_SYNTAX_UNQUOTE, Sym_SYNTAX_UNSPLICE;
```

112. ⟨ Global initialisation 3 ⟩ +≡

```
  Sym_ERR_UNEXPECTED = sym(ERR_UNEXPECTED);
  Sym_SYNTAX_DOTTED = sym(SYNTAX_DOTTED);
  Sym_SYNTAX_QUASI = sym(SYNTAX_QUASI);
  Sym_SYNTAX_QUOTE = sym(SYNTAX_QUOTE);
  Sym_SYNTAX_UNQUOTE = sym(SYNTAX_UNQUOTE);
  Sym_SYNTAX_UNSPLICE = sym(SYNTAX_UNSPLICE);
```

113. ⟨ Function declarations 8 ⟩ +≡

```
int read_byte(void);
cell read_cstring(char *);
cell read_form(void);
cell read_list(cell);
cell read_number(void);
cell read_sexp(void);
cell read_symbol(void);
cell read_symbol(void);
void unread_byte(char);
int useful_byte(void);
```

114.

```
int read_byte(void)
{
  int r;
  if ((r = Putback[0]) ≠ '\0') {
    Putback[0] = Putback[1];
    Putback[1] = '\0';
    return r;
  }
  if (Read_Pointer ≠ Λ) {
    r = *Read_Pointer;
    if (r ≡ '\0') r = EOF;
    Read_Pointer++;
    return r;
  }
  return getchar();
}
void unread_byte(char c)
{
  Putback[1] = Putback[0];
  Putback[0] = c;
}
```

115. The internal test suite defined below needs to be able to evaluate code it supplies from hard-coded C-strings. The mechanism defined here to make this work is extremely brittle and not meant to be used by user code. Or for very long until it can be replaced by something less quonky.

```
cell read_cstring(char *src)
{
  cell r;
  Read_Pointer = src;
  r = read_form();
  Read_Pointer = Λ;
  return r;
}
```

116. Even this primitive parser should support primitive comments.

```
int useful_byte(void)
{
  int c;
  while (¬Interrupt) {
    c = read_byte();
    switch (c) {
    case '␣': case '\n': case '\r': case '\t': continue;
    case ';': c = read_byte();
      while (c ≠ '\n' ∧ ¬Interrupt) {    /* read up to but not beyond the next newline */
        c = read_byte();
        if (c ≡ EOF) return c;
      }
      break;    /* go around again */
    default: return c;    /* includes EOF (which ≠ END_OF_FILE) */
    }
  }
  return EOF;
}
```

117. The public entry point to the reader is *read_sexp*. This simply resets the reader's global state and calls *read_form*.

```
cell read_sexp(void)
{
  cts_clear();
  Read_Level = 0;
  Putback[0] = Putback[1] = '\0';
  return read_form();
}
```

118. *read_form* reads a single (in most cases) byte which it uses to determine which parser function to dispatch to. The parser function will then return a complete s-expression (or raise an error).

```
cell read_form(void)
{
  cell r;
  int c, n;
  if (Interrupt) return VOID;
  if (Read_Level > READER_MAX_DEPTH) error (ERR_RECURSION, NIL);
  c = useful_byte();
  switch (c) {⟨Reader forms 119⟩}
  error (ERR_UNEXPECTED, NIL);
}
```

119. Here are the different bytes which *read_form* can understand, starting with the non-byte value EOF which is an error if the reader is part-way through parsing an expression.

```
⟨Reader forms 119⟩ ≡
case EOF:
  if (¬Read_Level) return END_OF_FILE;
  else error (ERR_ARITY_SYNTAX, NIL);
```

See also sections 121, 122, 123, 124, and 125.

This code is used in section 118.

120. Lists and vectors are read in exactly the same way, differentiating by being told to expect the appropriate delimiter.

121. ⟨ Reader forms 119 ⟩ +≡

```
case '(': return read_list(READ_CLOSE_PAREN);
case '[': return read_list(READ_CLOSE_BRACKET);
case ')': case ']':
    /* If Read_Level > 0 then read_form was called by read_list, otherwise read_sexp */
  if (¬Read_Level) error (ERR_ARITY_SYNTAX, NIL);
  else return c ≡ ')' ? READ_CLOSE_PAREN : READ_CLOSE_BRACKET;
```

122. A lone dot can only appear in a *list* and only before precisely one more expression. This is verified later by *read_list*.

⟨ Reader forms 119 ⟩ +≡

```
case '.':
  if (¬Read_Level) error (ERR_ARITY_SYNTAX, NIL);
  c = useful_byte();
  if (c ≡ EOF) error (ERR_ARITY_SYNTAX, NIL);
  unread_byte(c);
  return READ_DOT;
```

123. Special forms and strings aren't supported yet.

⟨ Reader forms 119 ⟩ +≡

```
case '"': case '#': case '|': error (ERR_UNIMPLEMENTED, NIL);
```

124. In addition to the main syntactic characters, three other characters commonly have special meaning in lisps: ⟪`'`⟫, ⟪`` ` ``⟫ and ⟪`,`⟫. ⟪`,`⟫ can also appear as ⟪`,@`⟫. Primarily these are for working with the macro expander.

In `LossLess` this syntax is unnecessary thanks to its first-class operatives but it's helpful so it's been retained. To differentiate between having parsed the syntactic form of these operators (eg. ⟪`'foo`⟫ or ⟪`'(bar baz)`⟫) and their symbolic form (eg. ⟪`(quote . foo)`⟫ or ⟪`(quote bar baz)`⟫) an otherwise ordinary *pair* with the operative's *symbol* in the *car* is created with the tag `FORMAT_SYNTAX`. These *syntax* `objects` are treated specially by the compiler and the writer.

```
⟨ Reader forms 119 ⟩ +≡
case '\'': case '`': n = useful_byte();
  if (n ≡ EOF) error (ERR_ARITY_SYNTAX, NIL);
  unread_byte(n);
  if (n ≡ ')' ∨ n ≡ ']') error (ERR_ARITY_SYNTAX, NIL);
  r = sym(c ≡ '`' ? SYNTAX_QUASI : SYNTAX_QUOTE);
  return atom(r, read_form(), FORMAT_SYNTAX);
case ',': c = read_byte();
  if (c ≡ EOF) error (ERR_ARITY_SYNTAX, NIL);
  if (c ≡ ')' ∨ c ≡ ']') error (ERR_ARITY_SYNTAX, NIL);
  if (c ≡ '@') {
    r = sym(SYNTAX_UNSPLICE);
    return atom(r, read_form(), FORMAT_SYNTAX);
  }
  else {
    unread_byte(c);
    r = sym(SYNTAX_UNQUOTE);
    return atom(r, read_form(), FORMAT_SYNTAX);
  }
```

125. Anything else is a number or a symbol (and this byte is part of it) provided it's ASCII.

```
⟨ Reader forms 119 ⟩ +≡
default:
  if (¬isprint(c)) error (ERR_ARITY_SYNTAX, NIL);
  unread_byte(c);
  if (isdigit(c)) return read_number();
  else return read_symbol();
```

126. *read_list* sequentially reads complete forms until it encounters the closing delimiter ⟪)⟫ or ⟪]⟫.

A pointer to the head of the *list* is saved and another pointer to its tail, *write*, is updated and used to insert the next object after it's been read, avoiding the need to reverse the *list* at the end.

```
cell read_list(cell delimiter)
{
  cell write, next, r;
  int count = 0;

  Read_Level++;
  write = cons(NIL, NIL);
  cts_push(write);
  while (1) {
    if (Interrupt) {
      cts_pop();
      Read_Level--;
      return VOID;
    }
    next = read_form();
    if (special_p(next)) {      /* These must return or terminate unless n is a 'real' special */
      ⟨Handle terminable 'forms' during list construction 127⟩
    }
    count++;
    cdr(write) = cons(NIL, NIL);
    write = cdr(write);
    car(write) = next;
  }
  Read_Level--;
  r = cdr(cts_pop());
  if (delimiter ≡ READ_CLOSE_BRACKET)
    return vector_new_list(r, count);
  return count ? r : NIL;
}
```

127. *read_form* is expected to return an s-expression or raise an error if the input is invalid. In order to recognise when a closing parenthesis/bracket is read 3 'special' special forms are defined, READ_CLOSE_PAREN, READ_CLOSE_BRACKET and READ_DOT. Although these look and act like the other global constants they don't exist outside of the parser.

```
⟨Handle terminable 'forms' during list construction 127⟩ ≡
  if (eof_p(next)) error (ERR_ARITY_SYNTAX, NIL);
  else if (next ≡ READ_CLOSE_BRACKET ∨ next ≡ READ_CLOSE_PAREN) {
    if (next ≠ delimiter) error (ERR_ARITY_SYNTAX, NIL);
    break;
  }
  else if (next ≡ READ_DOT) {⟨Read dotted pair 128⟩}
```

This code is used in section 126.

128. Encountering a ⟪.⟫ requires more special care than it deserves, made worse because if a *list* is dotted, a *syntax* `object` is created instead of a normal s-expression so that the style in which it's written out will be in the same that was read in.

⟨Read dotted pair 128⟩ ≡

```
if (count < 1 ∨ delimiter ≠ READ_CLOSE_PAREN)
      /* There must be at least one item already and we must be parsing a list. */
   error (ERR_ARITY_SYNTAX, NIL);
next = read_form();
if (special_p(next) ∧ next ≤ READ_SPECIAL)
      /* Check that the next 'form' isn't one of ⟪.⟫, ⟪)⟫ or ⟪]⟫ */
   error (ERR_ARITY_SYNTAX, NIL);
cdr(write) = atom(sym(SYNTAX_DOTTED), next, FORMAT_SYNTAX);
next = read_form();
if (next ≠ delimiter)
      /* Check that the next 'form' is really the closing delimiter */
   error (ERR_ARITY_SYNTAX, NIL);
break;
```

This code is used in section 127.

129. If it's not a *list* or a *vector* (or a *string* (⟪"⟫), *special form* (⟪#⟫), *raw symbol* (⟪|⟫) or *comment*) then the form being read is an *atom*. If the atom starts with a numeric digit then control proceeds directly to *read_number* otherwise *read_symbol* reads enough to determine whether the atom is a number beginning with ± or a valid or invalid symbol.

```
#define CHAR_TERMINATE "()[]\";␣\t\r\n"
#define terminable_p(c) strchr(CHAR_TERMINATE, (c))

cell read_number(void)
{
   char buf[12] = {0};      /* 2^32 is 10 digits, also ± and Λ */
   int c, i;
   long r;

   i = 0;
   while (1) {
      c = read_byte();
      if (c ≡ EOF)      /* TODO: If Read_Level is 0 is this an error? */
         error (ERR_ARITY_SYNTAX, NIL);
      if (i ≡ 0 ∧ (c ≡ '-' ∨ c ≡ '+')) buf[i++] = c;
      else if (isdigit(c)) buf[i++] = c;
      else if (¬terminable_p(c)) error (ERR_ARITY_SYNTAX, NIL);
      else {
         unread_byte(c);
         break;
      }
      if (i > 11) error (ERR_UNIMPLEMENTED, NIL);
   }
   r = atol(buf);
   if (r > INT_MAX ∨ r < INT_MIN) error (ERR_UNIMPLEMENTED, NIL);
   return int_new(r);
}
```

130. Although LossLess specifices (read: would specify) that there are no restrictions on the value of a *symbol*'s label, memory permitting, an artificial limit is being placed on the length of *symbol*s of 16KB[1].

That said, there are no restrictions on the value of a *symbol*'s label, memory permitting. There are limits on what can be *parsed* as a *symbol* in source code. The limits on plain *symbol*s are primarily to avoid things that look vaguely like numbers to the human eye being parsed as *symbol*s when the programmer thinks they should be parsed as a number. This helps to avoid mistakes like '3..14159' and harder to spot human errors being silently ignored.

• A *symbol* must not begin with a numeric digit or a syntactic character (comments (⟪;⟫), whitespace and everything recognised by *read_form*).

• The syntactic characters ⟪(⟫, ⟪)⟫, ⟪[⟫, ⟪]⟫, ⟪;⟫ and ⟪"⟫ cannot appear anywhere in the *symbol*.
nb. This means that the following otherwise syntactic characters *are* permitted in a symbol provided they do not occupy the first byte: ⟪.⟫, ⟪,⟫, ⟪'⟫, ⟪`⟫, ⟪#⟫ and ⟪|⟫. You probably shouldn't do that lightly though.

• If the first character of a *symbol* is ⟪-⟫ or ⟪+⟫ then it cannot be followed a numeric digit. ++⟨*digit*⟩ is valid.

• A ⟪-⟫ character or ⟪+⟫ followed by a ⟪.⟫ is a valid if strange *symbol* but a warning should probably be emitted by the parser if it finds that.

```
#define CHUNK_SIZE #80
#define READSYM_EOF_P if (c ≡ EOF) error (ERR_ARITY_SYNTAX, NIL)
      /* TODO: If Read_Level is 0 is this an error? */
  cell read_symbol(void)
  {
    cell r;
    char *buf, *nbuf;
    int c, i, s;
    c = read_byte();
    READSYM_EOF_P;
    ERR_OOM_P(buf = malloc(CHUNK_SIZE));
    s = CHUNK_SIZE;
    ⟨Read the first two bytes to check for a number 131⟩
    while (1) {⟨Read bytes until an invalid or terminating character 132⟩}
    buf[i] = '\0';     /* Λ-terminate the C-'string' */
    r = sym(buf);
    free(buf);
    return r;
  }
```

[1] 640KB was deemed to be far more than enough for anyone's needs.

131. Reading the first two bytes of a symbol is done specially to detect numbers beginning with ±. The first byte—which has already been read to check for `EOF`—is put into *buf* then if it matches ± the next byte is also read and also put into *buf*.

If that second byte is a digit then we're actually reading a number so put the bytes that were read so far into *Putback* and go to *read_number*, which will read them again. If the second byte is ⟪.⟫ then the *symbol* is valid but possibly a typo, so emit a warning and carry on.

```
⟨ Read the first two bytes to check for a number 131 ⟩ ≡
  buf[0] = c;
  i = 1;
  if (c ≡ '-' ∨ c ≡ '+') {
    c = read_byte();
    READSYM_EOF_P;
    buf[1] = c;
    i++;
    if (isdigit(buf[1])) {     /* This is a number! */
      unread_byte(buf[1]);
      unread_byte(buf[0]);
      free(buf);
      return read_number();
    }
    else if (buf[1] ≡ '.') warn(WARN_AMBIGUOUS_SYMBOL, NIL);
    else if (¬isprint(c)) error (ERR_ARITY_SYNTAX, NIL);
  }
```

This code is used in section 130.

132. After the first two bytes we're definitely reading a *symbol* so anything goes except non-printable characters (which are an error) or syntactic terminators which indicate the end of the *symbol*.

```
⟨ Read bytes until an invalid or terminating character 132 ⟩ ≡
  c = read_byte();
  READSYM_EOF_P;
  if (terminable_p(c)) {
    unread_byte(c);
    break;
  }
  if (¬isprint(c)) error (ERR_ARITY_SYNTAX, NIL);
  buf[i++] = c;
  if (i ≡ s) {     /* Enlarge buf if it's now full (this will also allow the Λ-terminator to fit) */
    nbuf = realloc(buf, s *= 2);
    if (nbuf ≡ Λ) {
      free(buf);
      error (ERR_OOM, NIL);
    }
    buf = nbuf;
  }
```

This code is used in section 130.

133. Writer. Although not an essential part of the language itself, the ability to display an s-expression to the user/programmer is obviously invaluable.

It is expected that this will (very!) shortly be changed to return a *string* representing the s-expression which can be passed on to an output routine but for the time being `LossLess` has no support for *string*s or output routines so the expression is written directly to *stdout*.

```
#define WRITER_MAX_DEPTH 1024   /* gotta pick something */
```

⟨ Function declarations 8 ⟩ +≡

```
boolean write_applicative(cell, int);
boolean write_compiler(cell, int);
boolean write_environment(cell, int);
boolean write_integer(cell, int);
boolean write_list(cell, int);
boolean write_operative(cell, int);
boolean write_symbol(cell, int);
boolean write_syntax(cell, int);
boolean write_vector(cell, int);
void write_form(cell, int);
```

134. Opaque Objects. *applicative*s, *compiler*s and *operative*s don't have much to say.

```
boolean write_applicative(cell sexp, int depth__unused)
{
  if (¬applicative_p(sexp)) return bfalse;
  printf("#<applicative␣...>");
  return btrue;
}
boolean write_compiler(cell sexp, int depth__unused)
{
  if (¬compiler_p(sexp)) return bfalse;
  printf("#<compiler-%s>", compiler_cname(sexp));
  return btrue;
}
boolean write_operative(cell sexp, int depth__unused)
{
  if (¬operative_p(sexp)) return bfalse;
  printf("#<operative␣...>");
  return btrue;
}
```

135. As-Is Objects. *integer*s and *symbol*s print themselves.

```
boolean write_integer(cell sexp, int depth __unused)
{
  if (¬integer_p(sexp)) return bfalse;
  printf("%d", int_value(sexp));
  return btrue;
}
boolean write_symbol(cell sexp, int depth __unused)
{
  int i;
  if (¬symbol_p(sexp)) return bfalse;
  for (i = 0; i < symbol_length(sexp); i++) putchar(symbol_store(sexp)[i]);
  return btrue;
}
```

136. Secret Objects. The hidden *syntax* object prints its syntactic form and then itself.

```
boolean write_syntax(cell sexp, int depth)
{
  if (¬syntax_p(sexp)) return bfalse;
  else if (car(sexp) ≡ sym(SYNTAX_DOTTED)) printf(".␣");
  else if (car(sexp) ≡ sym(SYNTAX_QUASI)) printf("‘");
  else if (car(sexp) ≡ sym(SYNTAX_QUOTE)) printf("’");
  else if (car(sexp) ≡ sym(SYNTAX_UNQUOTE)) printf(",");
  else if (car(sexp) ≡ sym(SYNTAX_UNSPLICE)) printf(",@");
  write_form(cdr(sexp), depth + 1);
  return btrue;
}
```

137. Environment Objects. An *environment* prints its own layer and then the layers above it.

```
boolean write_environment(cell sexp, int depth)
{
  if (¬environment_p(sexp)) return bfalse;
  printf("#<environment␣");
  write_form(env_layer(sexp), depth + 1);
  if (¬null_p(env_parent(sexp))) {
    printf("␣ON␣");
    write_form(env_parent(sexp), depth + 1);
    printf(">");
  }
  else printf("␣ROOT>");
  return btrue;
}
```

138. Sequential Objects. The routines for a *list* and *vector* are more or less the same – write each item in turn with whitespace after each form but the last, with the appropriate delimiters. *list*s also need to deal with being improper.

```
boolean write_list(cell sexp, int depth)
{
  if (¬pair_p(sexp)) return bfalse;
  printf("(");
  while (pair_p(sexp)) {
    write_form(car(sexp), depth + 1);
    if (pair_p(cdr(sexp)) ∨ syntax_p(cdr(sexp))) printf("␣");
    else if (¬null_p(cdr(sexp)) ∧ ¬pair_p(cdr(sexp)) ∧ ¬syntax_p(cdr(sexp))) printf("␣.␣");
    sexp = cdr(sexp);
  }
  if (¬null_p(sexp)) write_form(sexp, depth + 1);
  printf(")");
  return btrue;
}

boolean write_vector(cell sexp, int depth)
{
  int i;

  if (¬vector_p(sexp)) return bfalse;
  printf("[");
  for (i = 0; i < vector_length(sexp); i++) {
    write_form(vector_ref(sexp, i), depth + 1);
    if (i + 1 < vector_length(sexp)) printf("␣");
  }
  printf("]");
  return btrue;
}
```

139. *write_form* simply calls each writer in turn, stopping after the first one returning (C's) true.

```
void write_form(cell sexp, int depth)
{
  if (Interrupt) {
    if (¬depth) printf("...␣\n");
    return;
  }
  if (depth > WRITER_MAX_DEPTH) error (ERR_RECURSION, NIL);
  if (undefined_p(sexp)) printf("#><");      /* nothing should ever print this */
  else if (eof_p(sexp)) printf("#<eof>");
  else if (false_p(sexp)) printf("#f");
  else if (null_p(sexp)) printf("()");
  else if (true_p(sexp)) printf("#t");
  else if (void_p(sexp)) printf("#<>");
  else if (write_applicative(sexp, depth))      /* NOP */ ;
  else if (write_compiler(sexp, depth))      /* NOP */ ;
  else if (write_environment(sexp, depth))      /* NOP */ ;
  else if (write_integer(sexp, depth))      /* NOP */ ;
  else if (write_list(sexp, depth))      /* NOP */ ;
  else if (write_operative(sexp, depth))      /* NOP */ ;
  else if (write_symbol(sexp, depth))      /* NOP */ ;
  else if (write_syntax(sexp, depth))      /* NOP */ ;
  else if (write_vector(sexp, depth))      /* NOP */ ;
  else printf("#<wtf?>");      /* impossibru! */
}
```

140. Opcodes. With the core infrastucture out of the way we can finally turn our attention to the virtual machine implementation, or the implementation of the opcodes that the compiler will turn LossLess code into.

The opcodes that the virtual machine can perform must be declared before anything can be said about them. They take the form of an **enum**, this one unnamed. This list is sorted alphabetically for want of anything else.

Also defined here are *fetch* and *skip* which *opcode* implementations will use to obtain their argument(s) from *Prog* or advance *Ip*, respectively.

```
#define skip(d) Ip += (d)
#define fetch(d) vector_ref(Prog, Ip + (d))
```

⟨ Complex definitions & macros 140 ⟩ ≡

```
enum { OP_APPLY, OP_APPLY_TAIL, OP_CAR, OP_CDR,   /* 3 */
OP_COMPILE, OP_CONS, OP_CYCLE, OP_ENVIRONMENT_P,   /* 7 */
OP_ENV_MUTATE_M, OP_ENV_QUOTE, OP_ENV_ROOT, OP_ENV_SET_ROOT_M,   /* 11 */
OP_ERROR, OP_HALT, OP_JUMP, OP_JUMP_FALSE,   /* 15 */
OP_JUMP_TRUE, OP_LAMBDA, OP_LIST_P, OP_LIST_REVERSE,   /* 19 */
OP_LIST_REVERSE_M, OP_LOOKUP, OP_NIL, OP_NOOP,   /* 23 */
OP_NULL_P, OP_PAIR_P, OP_PEEK, OP_POP,   /* 27 */
OP_PUSH, OP_QUOTE, OP_RETURN, OP_RUN,   /* 31 */
OP_RUN_THERE, OP_SET_CAR_M, OP_SET_CDR_M, OP_SNOC,   /* 35 */
OP_SWAP, OP_SYNTAX, OP_VOV ,
#ifdef LL_TEST
⟨ Testing opcodes 219 ⟩
#endif
OPCODE_MAX } ;
```

See also sections 141 and 231.

This code is used in sections 1 and 2.

141. ⟨ Complex definitions & macros 140 ⟩ +≡

```
#ifndef LL_TEST
enum {   /* Ensure testing opcodes translate into undefined behaviour */
OP_TEST_UNDEFINED_BEHAVIOUR = #f00f , ⟨ Testing opcodes 219 ⟩
OPTEST_MAX } ;
#endif
```

142. Basic Flow Control. The most basic opcodes that the virtual machine needs are those which control whether to operate and where.

143. ⟨ Opcode implementations 10 ⟩ +≡

```
case OP_HALT:
  Running = 0;
  break;
case OP_JUMP:
  Ip = int_value(fetch(1));
  break;
case OP_JUMP_FALSE:
  if (void_p(Acc)) error (ERR_UNEXPECTED, VOID);
  else if (false_p(Acc)) Ip = int_value(fetch(1));
  else skip(2);
  break;
case OP_JUMP_TRUE:
  if (void_p(Acc)) error (ERR_UNEXPECTED, VOID);
  else if (true_p(Acc)) Ip = int_value(fetch(1));
  else skip(2);
  break;
case OP_NOOP:
  skip(1);
  break;
```

144. `OP_QUOTE` isn't really flow control but I don't know where else to put it.

⟨ Opcode implementations 10 ⟩ +≡

```
case OP_QUOTE:
  Acc = fetch(1);
  skip(2);
  break;
```

145. Pairs & Lists. `OP_CAR`, `OP_CDR`, `OP_NULL_P` and `OP_PAIR_P` are self explanatory.

⟨ Opcode implementations 10 ⟩ +≡

```
case OP_CAR:
  Acc = car(Acc);
  skip(1);
  break;
case OP_CDR:
  Acc = cdr(Acc);
  skip(1);
  break;
case OP_NULL_P:
  Acc = null_p(Acc) ? TRUE : FALSE;
  skip(1);
  break;
case OP_PAIR_P:
  Acc = pair_p(Acc) ? TRUE : FALSE;
  skip(1);
  break;
```

146. `OP_CONS` consumes one stack item (for the *cdr*) and puts the new pair in *Acc*. `OP_SNOC` does the opposite, pushing *Acc*'s *cdr* to the stack and leaving its *car* in *Acc*.

⟨ Opcode implementations 10 ⟩ +≡

```
case OP_CONS:
  Acc = cons(Acc, rts_pop(1));
  skip(1);
  break;
case OP_SNOC:
  rts_push(cdr(Acc));
  Acc = car(Acc);
  skip(1);
  break;
```

147. Cons cell mutators clear take an item from the stack and clear *Acc*.

⟨ Opcode implementations 10 ⟩ +≡

```
case OP_SET_CAR_M:
  car(rts_pop(1)) = Acc;
  Acc = VOID;
  skip(1);
  break;
case OP_SET_CDR_M:
  cdr(rts_pop(1)) = Acc;
  Acc = VOID;
  skip(1);
  break;
```

148. Other Objects. There is not much to say about these.

⟨ Opcode implementations 10 ⟩ +≡

case `OP_LIST_P`:
  if $(\neg false\_p(fetch(2)))$ **error** (`ERR_UNIMPLEMENTED`, `NIL`);
  $Acc = list\_p(Acc, fetch(1), \Lambda)$;
  $skip(3)$;
  break;
case `OP_LIST_REVERSE`:
  if $(\neg true\_p(fetch(1)) \lor \neg false\_p(fetch(2)))$ **error** (`ERR_UNIMPLEMENTED`, `NIL`);
  $Acc = list\_reverse(Acc, \Lambda, \Lambda)$;
  $skip(3)$;
  break;
case `OP_LIST_REVERSE_M`:
  $Acc = list\_reverse\_m(Acc, btrue)$;
  $skip(1)$;
  break;
case `OP_SYNTAX`:
  $Acc = atom(fetch(1), Acc,$ `FORMAT_SYNTAX`$)$;
  $skip(2)$;
  break;

149. Stack. `OP_PUSH` and `OP_POP` push the `object` in *Acc* onto the stack, or remove the top stack `object` into *Acc*, respectively. `OP_PEEK` is `OP_POP` without removing the item from the stack.

`OP_SWAP` swaps the `object` in *Acc* with the `object` on top of the stack.

`OP_CYCLE` swaps the top two stack items with each other.

`OP_NIL` pushes a `NIL` straight onto the stack without the need to quote it first.

⟨ Opcode implementations 10 ⟩ +≡

```
case OP_CYCLE:
  tmp = rts_ref(0);
  rts_set(0, rts_ref(1));
  rts_set(1, tmp);
  skip(1);
  break;
case OP_PEEK:
  Acc = rts_ref(0);
  skip(1);
  break;
case OP_POP:
  Acc = rts_pop(1);
  skip(1);
  break;
case OP_PUSH:
  rts_push(Acc);
  skip(1);
  break;
case OP_SWAP:
  tmp = Acc;
  Acc = rts_ref(0);
  rts_set(0, tmp);
  skip(1);
  break;
case OP_NIL:
  rts_push(NIL);
  skip(1);
  break;
```

150. Environments. Get or mutate *environment* objects. `OP_ENV_SET_ROOT_M` isn't used yet.

⟨ Opcode implementations 10 ⟩ +≡

```
case OP_ENVIRONMENT_P:
  Acc = environment_p(Acc) ? TRUE : FALSE;
  skip(1);
  break;
case OP_ENV_MUTATE_M:
  env_set(rts_pop(1), fetch(1), Acc, true_p(fetch(2)));
  Acc = VOID;
  skip(3);
  break;
case OP_ENV_QUOTE:
  Acc = Env;
  skip(1);
  break;
case OP_ENV_ROOT:
  Acc = Root;
  skip(1);
  break;
case OP_ENV_SET_ROOT_M:
  Root = Acc;     /* Root is 'lost'! */
  skip(1);
  break;
```

151. To look up the value of a variable in an *environment* we use `OP_LOOKUP` which calls the (recursive) *env_search*, interpreting the `UNDEFINED` it might return.

⟨ Opcode implementations 10 ⟩ +≡

```
case OP_LOOKUP:
  vms_push(Acc);
  Acc = env_search(Env, vms_ref());
  if (undefined_p(Acc)) {
    Acc = vms_pop();
    error(ERR_UNBOUND, Acc);
  }
  vms_pop();
  skip(1);
  break;
```

152. Closures. A *closure* is the combination of code to interpret and an *environment* to interpret it in. Usually a closure has arguments—making it useful—although in some cases a closure may work with global state or be idempotent.

In order to apply the arguments (if any) to the *closure* it must be entered by one of the opcodes `OP_APPLY` or `OP_APPLY_TAIL`. `OP_APPLY_TAIL` works identically to `OP_APPLY` and then consumes the stack frame which `OP_APPLY` created, allowing for *proper tail recursion* with further support from the compiler.

```
⟨ Opcode implementations 10 ⟩ +≡
case OP_APPLY:
  ⟨ Enter a closure 153 ⟩
  break;
case OP_APPLY_TAIL:
  ⟨ Enter a closure 153 ⟩
  frame_consume();
  break;
case OP_RETURN:
  frame_leave();
  break;
```

153. Whether in tail position or not, entering a *closure* is the same.

```
⟨ Enter a closure 153 ⟩ ≡
  {
    cell e, i, p;
    tmp = fetch(2);
    e = env_lift_stack(cadr(tmp), int_value(fetch(1)), car(tmp));
    p = caddr(tmp);
    i = int_value(cadddr(tmp));
    frame_push(3);
    frame_enter(e, p, i);
  }
```

This code is used in section 152.

154. Creating a closure in the first place follows an identical procedure whether it's an applicative or an operative but creates a different type of `object` in each case.

```
⟨ Opcode implementations 10 ⟩ +≡
case OP_LAMBDA:    /* The applicative */
  Acc = applicative_new(rts_pop(1), Env, Prog, int_value(fetch(1)));
  skip(2);
  break;
case OP_VOV:    /* The operative */
  Acc = operative_new(rts_pop(1), Env, Prog, int_value(fetch(1)));
  skip(2);
  break;
```

155. Compiler. The compiler needs to instruct the interpreter to compile more code and then run it, so these *opcode*s do that. `OP_COMPILE` compiles an s-expression into `LossLess` bytecode.

156. ⟨ Opcode implementations 10 ⟩ +≡

case `OP_COMPILE`:
  Acc = *compile*(*Acc*);
  skip(1);
  break;

157. `OP_RUN` interprets the bytecode in *Acc* in the current *environment*; the VM's live state is saved into a new stack *frame* then that *frame* is entered by executing the bytecode in *Acc*, starting at instruction 0.

⟨ Opcode implementations 10 ⟩ +≡

case `OP_RUN`:
  frame_push(1);
  frame_enter(*Env*, *Acc*, 0);
  break;

158. `OP_RUN_THERE` is like `OP_RUN` except that the *environment* to interpret the bytecode in is taken from the stack rather than staying in the active *environment*.

⟨ Opcode implementations 10 ⟩ +≡

case `OP_RUN_THERE`:
  vms_push(*rts_pop*(1));
  frame_push(1);
  frame_enter(*vms_pop*( ), *Acc*, 0);
  break;

159. Compiler. Speaking of the compiler, we can now turn our attention to writing it. The compiler is not advanced in any way but it is a little unusual. Due to the nature of first-class operatives, how to compile any expression can't be known until the combinator has been evaluated (read: compiled and then interpreted) in order to distinguish an applicative from an operative so that it knows whether to evaluate the arguments in the expression. I don't know if this qualifies it for a *Just-In-Time* compiler; I think *Finally-Able-To* is more suitable.

The compiler uses a small set of C macros which grow and fill *Compilation*—a *vector* holding the compilation in-progress.

```
#define ERR_COMPILE_DIRTY "compiler"
#define ERR_UNCOMBINABLE "uncombinable"
#define COMPILATION_SEGMENT #80
#define emitop(o) emit(int_new(o))
#define emitq(o) do { emitop(OP_QUOTE); emit(o); }
        while (0)    /* C... */
#define patch(i, v) (vector_ref(Compilation, (i)) = (v))
#define undot(p) ((syntax_p(p) ∧ car(p) ≡ Sym_SYNTAX_DOTTED) ? cdr(p) : (p))
```

⟨ Global variables 6 ⟩ +≡

```
int Here = 0;
cell Compilation = NIL;
```

160. ⟨ Externalised global variables 7 ⟩ +≡

```
extern int Here;
extern cell Compilation;
```

161. ⟨ Function declarations 8 ⟩ +≡

```
cell arity(cell, cell, int, boolean);
cell arity_next(cell, cell, cell, boolean, boolean);
int comefrom(void);
cell compile(cell);
void compile_car(cell, cell, boolean);
void compile_cdr(cell, cell, boolean);
void compile_conditional(cell, cell, boolean);
void compile_cons(cell, cell, boolean);
void compile_define_m(cell, cell, boolean);
void compile_env_current(cell, cell, boolean);
void compile_env_root(cell, cell, boolean);
void compile_error(cell, cell, boolean);
void compile_eval(cell, cell, boolean);
void compile_expression(cell, boolean);
void compile_lambda(cell, cell, boolean);
void compile_list(cell, cell, boolean);
cell compile_main(void);
void compile_null_p(cell, cell, boolean);
void compile_pair_p(cell, cell, boolean);
void compile_quasicompiler(cell, cell, cell, int, boolean);
void compile_quasiquote(cell, cell, boolean);
void compile_quote(cell, cell, boolean);
void compile_set_car_m(cell, cell, boolean);
void compile_set_cdr_m(cell, cell, boolean);
void compile_set_m(cell, cell, boolean);
void compile_symbol_p(cell, cell, boolean);
void compile_vov(cell, cell, boolean);
void emit(cell);
```

162. ⟨ Protected Globals 21 ⟩ +≡

```
&Compilation,
```

163. ⟨ Pre-initialise *Small_Int* & other gc-sensitive buffers 15 ⟩ +≡

```
Compilation = NIL;
```

164.

```
void emit(cell bc)
{
  int l;
  l = vector_length(Compilation);
  if (Here ≥ l)
    Compilation = vector_sub(Compilation, 0, l,
        0, l + COMPILATION_SEGMENT,
        OP_HALT);
  vector_ref(Compilation, Here++) = bc;
}
```

165. While compiling it frequently occurs that the value to emit isn't known at the time it's being emitted. The most common and obvious example of this is a forward jump who's address must immediately follow the opcode but the address won't be known until more compilation has been performed.

To make this work *comefrom* emits a `NIL` as a placeholder and returns its offset, which can later be passed in the first argument of *patch* to replace the `NIL` with the desired address etc.

```
int comefrom(void)
{
  emit(NIL);
  return Here − 1;
}
```

166. Compilation begins by preparing *Compilation* and `CTS` then recursively walks the tree in *source* dispatching to individual compilation routines to emit the appropriate bytecode.

```
cell compile(cell source)
{
  cell r;
  vms_push(source);
  Compilation = vector_new(COMPILATION_SEGMENT, int_new(OP_HALT));
  Here = 0;
  cts_reset( );
  compile_expression(source, 1);
  emitop(OP_RETURN);
  r = vector_sub(Compilation, 0, Here, 0, Here, VOID);
  Compilation = NIL;
  vms_clear( );
  if (¬null_p(CTS)) error (ERR_COMPILE_DIRTY, source);
  return r;
}
```

167. *compile_main* is used during initialisation to build the bytecode ≪`OP_COMPILE OP_RUN OP_HALT`≫ which is the program installed initially into the virtual machine.

```
cell compile_main(void)
{
  cell r;
  r = vector_new_imp(3, 0, 0);
  vector_ref(r, 0) = int_new(OP_COMPILE);
  vector_ref(r, 1) = int_new(OP_RUN);
  vector_ref(r, 2) = int_new(OP_HALT);
  return r;
}
```

168. The first job of the compiler is to figure out what type of expression it's compiling, chiefly whether it's a *list* to combine or an *atom* which is itself.

```
void compile_expression(cell sexp, int tail_p)
{
  if (¬pair_p(sexp) ∧ ¬syntax_p(sexp)) {⟨ Compile an atom 169 ⟩}
  else {⟨ Compile a combiner 170 ⟩}
}
```

169. The only *atom* which doesn't evaluate to itself is a *symbol*. A *symbol* being evaluated references a variable which must be looked up in the active environment.

⟨ Compile an atom 169 ⟩ ≡
```
if (symbol_p(sexp)) {
    emitq(sexp);
    emitop(OP_LOOKUP);
}
else { emitq(sexp); }
```
This code is used in section 168.

170. Combining a *list* requires more work. This is also where operatives obtain the property of being first-class objects by delaying compilation of all but the first expression in the *list* until after that compiled bytecode has been interpreted.

⟨ Compile a combiner 170 ⟩ ≡
```
cell args, combiner;
combiner = car(sexp);
args = undot(cdr(sexp));
⟨ Search Root for syntactic combiners 171 ⟩
if (compiler_p(combiner)) {⟨ Compile native combiner 172 ⟩}
else if (applicative_p(combiner)) {⟨ Compile applicative combiner 180 ⟩}
else if (operative_p(combiner)) {⟨ Compile operative combiner 189 ⟩}
else if (symbol_p(combiner) ∨ pair_p(combiner)) {⟨ Compile unknown combiner 173 ⟩}
else { error (ERR_UNCOMBINABLE, combiner); }
```
This code is used in section 168.

171. If the combiner (*sexp*'s *car*) is a *syntax* object then it represents the result of parsing (for example) ⟪`'(expression)`⟫ into ⟪`(quote expression)`⟫ and it must always mean the *real quote* operator, so *syntax* combiners are always looked for directly (and only) in *Root*.

⟨ Search *Root* for syntactic combiners 171 ⟩ ≡
```
if (syntax_p(sexp)) {
    cell c;
    c = env_search(Root, combiner);
    if (undefined_p(c)) error (ERR_UNBOUND, combiner);    /* should never happen */
    combiner = c;
}
```
This code is used in section 170.

172. A native compiler is simple; look up its address in `COMPILER` and go there. The individual native compilers are defined below.

⟨ Compile native combiner 172 ⟩ ≡
```
compiler_fn(combiner)(combiner, args, tail_p);
```
This code is used in section 170.

173. If the compiler doesn't know whether *combiner* is applicative or operative then that must be determined before *args* can be considered.

⟨ Compile unknown combiner 173 ⟩ ≡

```
emitq(args);
emitop(OP_PUSH);   /* save args onto the stack */
compile_expression(combiner, 0);   /* evaluate the combiner, leaving it in Acc */
emitop(OP_CONS);   /* rebuild sexp with the evaluated combiner */
emitop(OP_COMPILE);   /* continue compiling sexp */
emitop(OP_RUN);   /* run that code in the same environment */
```

This code is used in section 170.

174. Function Bodies. Nearly everything has arguments to process and it's nearly always done in the same way. *arity* and *arity_next* work in concert to help the compiler implementations check how many arguments there are (but not their value or type) and raise any errors encountered.

arity pushes the minimum required arguments onto the compiler stack (in reverse) and returns a pointer to the rest of the argument list.

```
#define ERR_ARITY_EXTRA "extra"
#define ERR_ARITY_MISSING "missing"
#define ERR_ARITY_SYNTAX "syntax"
#define arity_error(e, c, a) error ((e), cons((c), (a)))
  cell arity(cell op, cell args, int min, boolean more_p)
  {
    cell a = args;
    int i = 0;
    for ( ; i < min; i++) {
      if (null_p(a)) {
        if (compiler_p(op) ∨ operative_p(op)) arity_error(ERR_ARITY_SYNTAX, op, args);
        else arity_error(ERR_ARITY_MISSING, op, args);
      }
      if (¬pair_p(a)) arity_error(ERR_ARITY_SYNTAX, op, args);
      cts_push(car(a));
      a = cdr(a);
    }
    if (min ∧ ¬more_p ∧ ¬null_p(a)) {
      if (pair_p(a)) arity_error(ERR_ARITY_EXTRA, op, args);
      else arity_error(ERR_ARITY_SYNTAX, op, args);
    }
    return a;
  }
```

175. *arity_next*, given the remainder of the arguments that were returned from *arity*, checks whether another one is present and whether it's allowed to be, then returns a value suitable for another call to *arity_next*.

```
  cell arity_next(cell op, cell args, cell more, boolean required_p, boolean last_p)
  {
    if (null_p(more)) {
      if (required_p) arity_error(ERR_ARITY_MISSING, op, args);
      else {
        cts_push(UNDEFINED);
        return NIL;
      }
    }
    else if (¬pair_p(more))
      arity_error(ERR_ARITY_SYNTAX, op, args);
    else if (last_p ∧ ¬null_p(cdr(more))) {
      if (operative_p(op) ∧ pair_p(cdr(more))) arity_error(ERR_ARITY_EXTRA, op, args);
      else arity_error(ERR_ARITY_SYNTAX, op, args);
    }
    cts_push(car(more));
    return cdr(more);
  }
```

176. *closure* bodies, and the contents of a *begin* expression, are compiled by simply walking the list and recursing into *compile_expression* for everything on it. When compiling the last item in the list the *tail_p* flag is raised so that the expression can use `OP_APPLY_TAIL` if appropriate, making tail recursion proper.

```
void compile_list(cell op, cell sexp, boolean tail_p)
{
  boolean t;
  cell body, next, this;

  body = undot(sexp);
  t = null_p(body);
  if (t) {
    emitq(VOID);
    return;
  }
  while (¬t) {
    if (¬pair_p(body)) arity_error(ERR_ARITY_SYNTAX, op, sexp);
    this = car(body);
    next = undot(cdr(body));
    t = null_p(next);
    compile_expression(this, t ∧ tail_p);
    body = next;
  }
}
```

177. Closures (Applicatives & Operatives). The first thing to understand is that at their core *applicative*s and *operative*s work in largely the same way and have the same internal representation:

- The static *environment* which will expand into a local *environment* when entering the *closure*. This is where the variables that were "closed over" are stored.
- The program which the *closure* will perform, as compiled bytecode and a starting instruction pointer.
- A list of formals naming any arguments which will be passed to the *closure*, so that they can be put into the newly-extended *environment*.

Entering a *closure* means extracting these saved values and restoring them to the virtual machine's registers, *Env*, *Prog* & *Ip*.

A *closure* can (usually does) have arguments and it's how they're handled that differentiates an *applicative* from an *operative*.

178. The main type of *closure* everyone is familiar with already even if they don't know it is a function[1] or *applicative*.

An *applicative* is created in response to evaluating a **lambda** expression. The bytecode which does this evaluating is created by *compile_lambda*.

```
void compile_lambda(cell op, cell args, boolean tail_p)
{
  cell body, in, formals, f;
  int begin_address, comefrom_end;
  body = arity(op, args, 1, 1);
  body = undot(body);
  formals = cts_pop();
  formals = undot(formals);
  if (¬symbol_p(formals)) {⟨Process lambda formals 179⟩}
  emitq(formals);    /* push formals onto the stack */
  emitop(OP_PUSH);
  emitop(OP_LAMBDA);    /* create the applicative */
  begin_address = comefrom();    /* start address; argument to OP_LAMBDA */
  emitop(OP_JUMP);    /* jump over the compiled closure body */
  comefrom_end = comefrom();
  patch(begin_address, int_new(Here));
  compile_list(op, body, tail_p);    /* compile the code that entering the closure will interpret */
  emitop(OP_RETURN);    /* returns from the closure at run-time */
  patch(comefrom_end, int_new(Here));
}
```

[1] The word "function" is horribly misused everywhere and this trend will continue without my getting in its way.

179. If the *formals* given in the **lambda** expression are not in fact a single *symbol* then it must be a list of *symbol*s which is verified here. At the same time if the list is a dotted pair then the *syntax* wrapper is removed.

⟨ Process lambda formals 179 ⟩ ≡
```
cts_push(f = cons(NIL, NIL));
in = formals;
while (pair_p(in)) {
  if (¬symbol_p(car(in)) ∧ ¬null_p(car(in))) arity_error(ERR_ARITY_SYNTAX, op, args);
  cdr(f) = cons(car(in), NIL);
  f = cdr(f);
  in = undot(cdr(in));
}
if (¬null_p(in)) {
  if (¬symbol_p(in) ∧ ¬null_p(in)) arity_error(ERR_ARITY_SYNTAX, op, args);
  cdr(f) = in;
}
formals = cdr(cts_pop());
```
This code is used in section 178.

180. To enter this *closure* at run-time—aka. to call the function returned by **lambda**—the arguments it's called with must be evaluated (after being arity checked) then OP_APPLY or OP_APPLY_TAIL enters the *closure*, consuming a stack *frame* in the latter case.

The arguments and the formals saved in the *applicative* are walked together and saved in *direct*. If the formals list ends in a dotted *pair* then the remainder of the arguments are saved in *collect*.

When *collect* and *direct* have been prepared, being a copy of the unevaluated arguments in reverse order, they are walked again to emit the opcodes which will evaluate each argument and put the results onto the stack.

⟨ Compile applicative combiner 180 ⟩ ≡
```
cell collect, direct, formals, a;
int nargs = 0;

formals = applicative_formals(combiner);
cts_push(direct = NIL);
a = undot(args);
⟨ Look for required arguments 181 ⟩
⟨ Look for optional arguments 182 ⟩
if (pair_p(a)) arity_error(ERR_ARITY_EXTRA, combiner, args);
else if (¬null_p(a)) arity_error(ERR_ARITY_SYNTAX, combiner, args);
⟨ Evaluate optional arguments into a list 184 ⟩
⟨ Evaluate required arguments onto the stack 183 ⟩
cts_clear();
emitop(tail_p ? OP_APPLY_TAIL : OP_APPLY);
emit(int_new(nargs));
emit(combiner);
```
This code is used in section 170.

181. It's a syntax error if the arguments are not a proper list, otherwise there is nothing much to say about this.

⟨ Look for required arguments 181 ⟩ ≡
```
while (pair_p(formals)) {
  if (¬pair_p(a)) {
    if (null_p(a)) arity_error(ERR_ARITY_SYNTAX, combiner, args);
    else arity_error(ERR_ARITY_SYNTAX, combiner, args);
  }
  direct = cons(car(a), direct);
  cts_set(direct);
  a = undot(cdr(a));
  formals = cdr(formals);
  nargs++;
}
```
This code is used in section 180.

182. If the *applicative* formals indicate that it can be called with a varying number of arguments then that counts as one more argument which will be a list of whatever arguments remain.

⟨ Look for optional arguments 182 ⟩ ≡
```
if (symbol_p(formals)) {
  nargs++;
  cts_push(collect = NIL);
  while (pair_p(a)) {
    collect = cons(car(a), collect);
    cts_set(collect);
    a = undot(cdr(a));
  }
}
```
This code is used in section 180.

183. To perform the evaluation, each argument in the (now reversed) list *direct* is compiled followed by an OP_PUSH to save the result on the stack.

⟨ Evaluate required arguments onto the stack 183 ⟩ ≡
```
while (¬null_p(direct)) {
  compile_expression(car(direct), 0);
  emitop(OP_PUSH);
  direct = cdr(direct);
}
```
This code is used in section 180.

184. If the *applicative* expects a varying number of arguments then the (also reversed) list in *collect* is compiled in the same way but before `OP_PUSH`, `OP_CONS` removes the growing list from the stack and prepends the new result to it and it's this *list* which is pushed.

⟨ Evaluate optional arguments into a *list* 184 ⟩ ≡

```
if (symbol_p(formals)) {
    emitop(OP_NIL);
    while (¬null_p(collect)) {
        compile_expression(car(collect), 0);
        emitop(OP_CONS);
        emitop(OP_PUSH);
        collect = cdr(collect);
    }
    cts_clear();
}
```

This code is used in section 180.

185. Analogous to *compile_lambda* for *applicatives* is *compile_vov* for *operatives*. An *operative closure* is a simpler than an *applicative* because the arguments are not evaluated. Instead *compile_vov* needs to handle **vov**'s very different way of specifying its formals.

Resembling *let* rather than **lambda**, **vov**'s formals specify what run-time detail the *operative* needs: The unevaluated arguments, the active environment and/or (unimplemented) a continuation delimiter. To do this each entry in the formals list is an association pair with the *symbol*ic name for that detail associated with another symbol specifying what: *vov/arguments*, *vov/environment* or *vov/continuation*. Because no-one wants RSI these have the abbreviations *vov/args*, *vov/env* and *vov/cont*.

⟨ Global variables 6 ⟩ +≡

```
cell Sym_vov_args = UNDEFINED;
cell Sym_vov_args_long = UNDEFINED;
cell Sym_vov_cont = UNDEFINED;
cell Sym_vov_cont_long = UNDEFINED;
cell Sym_vov_env = UNDEFINED;
cell Sym_vov_env_long = UNDEFINED;
```

186. ⟨ Global initialisation 3 ⟩ +≡

```
Sym_vov_args = sym("vov/args");
Sym_vov_args_long = sym("vov/arguments");
Sym_vov_cont = sym("vov/cont");
Sym_vov_cont_long = sym("vov/continuation");
Sym_vov_env = sym("vov/env");
Sym_vov_env_long = sym("vov/environment");
```

187. **void** *compile_vov*(**cell** *op*, **cell** *args*, **boolean** *tail_p*)

```
{
  cell body, formals;
  int begin_address, comefrom_end;
  cell a = NIL;
  cell c = NIL;
  cell e = NIL;
  body = arity(op, args, 1, 1);
  body = undot(body);
  formals = cts_pop();
  formals = undot(formals);
  ⟨Scan operative informals 188⟩
  emitop(OP_NIL);  /* push formals onto the stack */
  emitq(c); emitop(OP_CONS); emitop(OP_PUSH);
  emitq(e); emitop(OP_CONS); emitop(OP_PUSH);
  emitq(a); emitop(OP_CONS); emitop(OP_PUSH);
  emitop(OP_VOV);  /* create the operative */
    /* The rest of compile_vov is identical to compile_lambda: */
  begin_address = comefrom();  /* start address; argument to opcode */
  emitop(OP_JUMP);  /* jump over the compiled closure body */
  comefrom_end = comefrom();
  patch(begin_address, int_new(Here));
  compile_list(op, body, tail_p);  /* compile the code that entering the closure will interpret */
  emitop(OP_RETURN);  /* return from the run-time closure */
  patch(comefrom_end, int_new(Here));  /* finish building the closure */
}
```

188. To scan the "informals" three variables, *a*, *c* and *e* are prepared with NIL representing the *symbol* for the arguments, *continuation* and *environment* respectively. Each "informal" is checked in turn using *arity* and the appropriate placeholder's NIL replaced with the *symbol*.

⟨Scan operative informals 188⟩ ≡

```
  cell r, s;
  if (¬pair_p(formals)) arity_error(ERR_ARITY_SYNTAX, op, args);
#define CHECK_AND_ASSIGN(v)
  {
    if (¬null_p(v)) arity_error(ERR_ARITY_SYNTAX, op, args);
    (v) = s;
  }
  while (pair_p(formals)) {
    arity(op, car(formals), 2, 0);
    r = cts_pop();
    s = cts_pop();
    if (¬symbol_p(s)) arity_error(ERR_ARITY_SYNTAX, op, args);
    else if (r ≡ Sym_vov_args ∨ r ≡ Sym_vov_args_long) CHECK_AND_ASSIGN(a)
    else if (r ≡ Sym_vov_env ∨ r ≡ Sym_vov_env_long) CHECK_AND_ASSIGN(e)
    else if (r ≡ Sym_vov_cont ∨ r ≡ Sym_vov_cont_long) CHECK_AND_ASSIGN(c)
    formals = cdr(formals);
  }
  if (¬null_p(formals)) arity_error(ERR_ARITY_SYNTAX, op, args);
```

This code is used in section 187.

189. Entering an *operative* involves pushing the 3 desired run-time properties, or `NIL`, onto the stack as though arguments to an *applicative closure* (remember that the unevaluated run-time arguments of the *closure* are potentially one of those run-time properties).

⟨ Compile operative combiner 189 ⟩ ≡

```
cell a, c, e, f;
f = operative_formals(combiner);
a = ¬null_p(car(f)); f = cdr(f);
e = ¬null_p(car(f)); f = cdr(f);
c = ¬null_p(car(f)); f = cdr(f);
if (c) error (ERR_UNIMPLEMENTED, NIL);
else emitop(OP_NIL);
if (e) {
  emitop(OP_ENV_QUOTE);
  emitop(OP_PUSH);
}
else emitop(OP_NIL);
if (a) {
  emitq(args);
  emitop(OP_PUSH);
}
else emitop(OP_NIL);
emitop(tail_p ? OP_APPLY_TAIL : OP_APPLY);
emit(int_new(3));
emit(combiner);
```

This code is used in section 170.

190. Conditionals (if). Although you could define a whole language with just **lambda** and **vov**[1] that way lies Church Numerals and other madness, so we will define the basic conditional, **if**.

```
void compile_conditional(cell op, cell args, boolean tail_p)
{
  cell alternate, condition, consequent, more;
  int jump_false, jump_true;
  more = arity(op, args, 2, 1);
  arity_next(op, args, more, 0, 1);
  alternate = cts_pop();
  consequent = cts_pop();
  condition = cts_pop();
  compile_expression(condition, 0);
  emitop(OP_JUMP_FALSE);
  jump_false = comefrom();
  compile_expression(consequent, tail_p);
  emitop(OP_JUMP);
  jump_true = comefrom();
  patch(jump_false, int_new(Here));
  if (undefined_p(alternate)) emitq(VOID);
  else compile_expression(alternate, tail_p);
  patch(jump_true, int_new(Here));
}
```

[1] In fact I think conditionals can be achieved in both somehow, so you only need one.

191. Run-time Evaluation (eval). **eval** must evaluate its 1 or 2 arguments in the current environment, and then enter the environment described by the second to execute the program in the first.

```
void compile_eval(cell op, cell args, boolean tail_p__unused)
{
  cell more, sexp, eenv;
  int goto_env_p;
  more = arity(op, args, 1, 1);
  sexp = cts_pop();
  arity_next(op, args, more, 0, 1);
  eenv = cts_pop();
  if (undefined_p(eenv)) {
    emitop(OP_ENV_QUOTE);
    emitop(OP_PUSH);
  }
  else {
    compile_expression(eenv, 0);
    emitop(OP_PUSH);
    emitop(OP_ENVIRONMENT_P);
    emitop(OP_JUMP_TRUE);
    goto_env_p = comefrom();
    emitq(Sym_ERR_UNEXPECTED);
    emitop(OP_ERROR);
    patch(goto_env_p, int_new(Here));
  }
  compile_expression(sexp, 0);
  emitop(OP_COMPILE);
  emitop(OP_RUN_THERE);
}
```

192. Run-time Errors. **error** expects a symbol an the first position and an optional form to evaluate in the second.

```
void compile_error(cell op, cell args, boolean tail_p __unused)
{
  cell id, more, value;
  more = arity(op, args, 1, 1);
  arity_next(op, args, more, 0, 1);
  value = cts_pop();
  id = cts_pop();
  if (¬symbol_p(id)) arity_error(ERR_ARITY_SYNTAX, op, args);
  if (undefined_p(value)) emitq(NIL);
  else compile_expression(value, 0);
  emitop(OP_PUSH);
  emitq(id);
  emitop(OP_ERROR);
}
```

193. Cons Cells. These operators have been written out directly despite the obvious potential for refactoring into reusable pieces. This is short-lived until more compiler routines have been written and the similarity patterns between them become apparent.

Cons cells are defined by the *cons*, *car*, *cdr*, *null?* and *pair?* symbols with *set-car!* and *set-cdr!* providing for mutation.

```
void compile_cons(cell op, cell args, boolean tail_p__unused)
{    /* pattern 0; arity ≡ (O, O) */
  cell ncar, ncdr;

  arity(op, args, 2, 0);
  ncdr = cts_pop();
  ncar = cts_pop();
  compile_expression(ncdr, 0);
  emitop(OP_PUSH);
  compile_expression(ncar, 0);
  emitop(OP_CONS);
}

void compile_car(cell op, cell args, boolean tail_p__unused)
{    /* pattern 1; arity = (OP_PAIR_P) */
  int comefrom_pair_p;

  arity(op, args, 1, 0);
  compile_expression(cts_pop(), 0);
  emitop(OP_PUSH);
  emitop(OP_PAIR_P);
  emitop(OP_JUMP_TRUE);
  comefrom_pair_p = Here;
  emit(NIL);
  emitq(sym(ERR_UNEXPECTED));    /* TODO */
  emitop(OP_ERROR);
  patch(comefrom_pair_p, int_new(Here));
  emitop(OP_POP);
  emitop(OP_CAR);
}

void compile_cdr(cell op, cell args, boolean tail_p__unused)
{
  int comefrom_pair_p;

  arity(op, args, 1, 0);
  compile_expression(cts_pop(), 0);
  emitop(OP_PUSH);
  emitop(OP_PAIR_P);
  emitop(OP_JUMP_TRUE);
  comefrom_pair_p = Here;
  emit(NIL);
  emitq(sym(ERR_UNEXPECTED));    /* TODO */
  emitop(OP_ERROR);
  patch(comefrom_pair_p, int_new(Here));
  emitop(OP_POP);
  emitop(OP_CDR);    /* this is the only difference from the above */
}

void compile_null_p(cell op, cell args, boolean tail_p__unused)
{    /* pattern 2 = predicate */
```

```
  arity(op, args, 1, 0);
  compile_expression(cts_pop(), 0);
  emitop(OP_NULL_P);
}
void compile_pair_p(cell op, cell args, boolean tail_p __unused)
{
  arity(op, args, 1, 0);
  compile_expression(cts_pop(), 0);
  emitop(OP_PAIR_P);
}
void compile_set_car_m(cell op, cell args, boolean tail_p __unused)
{    /* pattern 3 = arity = (OP_PAIR_P, O) */
  cell value, object;
  int goto_pair_p;
  arity(op, args, 2, 0);
  value = cts_pop();
  object = cts_pop();
  compile_expression(object, bfalse);
  emitop(OP_PUSH);
  emitop(OP_PAIR_P);
  emitop(OP_JUMP_TRUE);
  goto_pair_p = comefrom();
  emitq(Sym_ERR_UNEXPECTED);
  emitop(OP_ERROR);
  patch(goto_pair_p, int_new(Here));
  compile_expression(value, bfalse);
  emitop(OP_SET_CAR_M);
}
void compile_set_cdr_m(cell op, cell args, boolean tail_p __unused)
{
  cell value, object;
  int goto_pair_p;
  arity(op, args, 2, 0);
  value = cts_pop();
  object = cts_pop();
  compile_expression(object, bfalse);
  emitop(OP_PUSH);
  emitop(OP_PAIR_P);
  emitop(OP_JUMP_TRUE);
  goto_pair_p = comefrom();
  emitq(Sym_ERR_UNEXPECTED);
  emitop(OP_ERROR);
  patch(goto_pair_p, int_new(Here));
  compile_expression(value, bfalse);
  emitop(OP_SET_CDR_M);
}
```

194. Environment. The *environment* mutators are the same except for the flag given to the final opcode.

```
void compile_set_m(cell op, cell args, boolean tail_p __unused)
{     /* pattern 4, arity = ( OP_ENV_P #<> symbol ? ) */
  cell env, name, value;
  int goto_env_p;

  arity(op, args, 3, bfalse);
  value = cts_pop();
  name = cts_pop();
  env = cts_pop();
  if (¬symbol_p(name)) error (ERR_ARITY_SYNTAX, NIL);
  compile_expression(env, bfalse);
  emitop(OP_PUSH);
  emitop(OP_ENVIRONMENT_P);
  emitop(OP_JUMP_TRUE);
  goto_env_p = comefrom();
  emitq(Sym_ERR_UNEXPECTED);
  emitop(OP_ERROR);
  patch(goto_env_p, int_new(Here));
  compile_expression(value, bfalse);
  emitop(OP_ENV_MUTATE_M);
  emit(name);
  emit(FALSE);
}

void compile_define_m(cell op, cell args, boolean tail_p __unused)
{
  cell env, name, value;
  int goto_env_p;

  arity(op, args, 3, bfalse);
  value = cts_pop();
  name = cts_pop();
  env = cts_pop();
  if (¬symbol_p(name)) error (ERR_ARITY_SYNTAX, NIL);
  compile_expression(env, bfalse);
  emitop(OP_PUSH);
  emitop(OP_ENVIRONMENT_P);
  emitop(OP_JUMP_TRUE);
  goto_env_p = comefrom();
  emitq(Sym_ERR_UNEXPECTED);
  emitop(OP_ERROR);
  patch(goto_env_p, int_new(Here));
  compile_expression(value, bfalse);
  emitop(OP_ENV_MUTATE_M);
  emit(name);
  emit(TRUE);
}

void compile_env_root(cell op, cell args, boolean tail_p __unused)
{     /* pattern 5 = no args */
  arity(op, args, 0, bfalse);
  emitop(OP_ENV_ROOT);
}
```

```
void compile_env_current(cell op, cell args, boolean tail_p __unused)
{
   arity(op, args, 0, bfalse);
   emitop(OP_ENV_QUOTE);
}
```

195. Quotation & Quasiquotation. A quoted object is one which is not evaluated and we have an opcode to do just that, used by many of the implementations above.

```
void compile_quote(cell op__unused, cell args, boolean tail_p__unused)
{ emitq(args); }
```

196. Quasiquoting an object is almost, but not quite, entirely different. The end result is the same however—a run-time object which (almost) exactly matches the unevaluated source code that it was created from.

A quasiquoted object is converted into its final form by changing any *unquote* (and *unquote-splicing*) within it to the result of evaulating them. This is complicated enough because we're now writing a compiler within our compiler[1] but additionally the quasiquoted object may contain quasiquoted objects, changing the nature of the inner-*unquote* operators.

197. The compiler for compiling quasiquoted code only calls directly into the recursive quasicompiler engine (let's call it the quasicompiler).

```
⟨ Function declarations 8 ⟩ +≡
void compile_quasicompiler(cell, cell, cell, int, boolean);
```

198.

```
void compile_quasiquote(cell op, cell args, boolean tail_p__unused)
{    /* pattern Q */
  compile_quasicompiler(op, args, args, 0, bfalse);
}
```

199. As with any compiler, the first task is to figure out what sort of expression is being quasicompiled. Atoms are themselves. Otherwise lists and vectors must be recursively compiled item-by-item, and the syntactic operators must operate when encountered.

Quasiquoting *vector*s is not supported but I'm not anticipating it being difficult, just not useful yet.

```
void compile_quasicompiler(cell op, cell oargs, cell arg, int depth, boolean in_list_p)
{
  if (pair_p(arg)) {⟨ Quasiquote a pair/list 200 ⟩}
  else if (vector_p(arg)) { error (ERR_UNIMPLEMENTED, NIL); }
  else if (syntax_p(arg)) {⟨ Quasiquote syntax 201 ⟩}
  else {
    emitq(arg);
    if (in_list_p) emitop(OP_CONS);
  }
}
```

[1] Yo!

200. Dealing first with the simple case of a list the quasicompiler reverses the list to find its tail, which may or may not be `NIL`, and recursively calling *compile_quasicompiler* for every item.

After each item has been quasicompiled it will be combined with the transformed list being grown on top of the stack.

When quasicompiling the list's tail there is no partial list to prepend it to so the quasicompiler is entered in atomic mode. *compile_quasicompiler* can be relied on to handle the tail of a proper or improper list.

⟨ Quasiquote a pair/list 200 ⟩ ≡

```
cell todo, tail;
tail = NIL;
todo = list_reverse(arg, &tail, Λ);
compile_quasicompiler(op, oargs, tail, depth, bfalse);
for ( ; ¬null_p(todo); todo = cdr(todo)) {
  emitop(OP_PUSH);     /* Push the list so far */
  compile_quasicompiler(op, oargs, car(todo), depth, btrue);
}
if (in_list_p) emitop(OP_CONS);
```

This code is used in section 199.

201. The quote & unquote syntax is where the quasicompiler starts to get interesting. *quotes* and *quasiquotes* (and a *dotted* tail) recurse back into the quasicompiler to emit the transformation of the quoted object, then re-apply the syntax operator.

depth is increased when recursing into a *quasiquote* so that the compiler knows whether to evaluate an unquote operator.

⟨ Quasiquote syntax 201 ⟩ ≡

```
int d;
if (car(arg) ≡ Sym_SYNTAX_DOTTED
      ∨ car(arg) ≡ Sym_SYNTAX_QUOTE
      ∨ car(arg) ≡ Sym_SYNTAX_QUASI) {
  d = (car(arg) ≡ Sym_SYNTAX_QUASI) ? 1 : 0;
  compile_quasicompiler(op, oargs, cdr(arg), depth + d, bfalse);
  emitop(OP_SYNTAX);
  emit(car(arg));
  if (in_list_p) emitop(OP_CONS);
}
```

See also sections 202 and 203.

This code is used in section 199.

202. *unquote* evaluates the unquoted object. If quasiquote is quasicompiling an inner quasiquote then the unquoted object isn't evaluated but compiled at a decreased *depth*. This enables the correct unquoting-or-not of quasiquoting quasiquoted quotes.

⟨ Quasiquote syntax 201 ⟩ +≡

```
else
  if (car(arg) ≡ Sym_SYNTAX_UNQUOTE) {
    if (depth > 0) {
      compile_quasicompiler(op, oargs, cdr(arg), depth − 1, bfalse);
      emitop(OP_SYNTAX);
      emit(Sym_SYNTAX_UNQUOTE);
    }
    else compile_expression(cdr(arg), bfalse);
    if (in_list_p) emitop(OP_CONS);
  }
```

203. Similarly to *unquote*, *unquote-splicing* recurses back into the quasicompiler at a lower depth when unquoting an inner quasiquote.

⟨ Quasiquote syntax 201 ⟩ +≡

```
else
  if (car(arg) ≡ Sym_SYNTAX_UNSPLICE) {
    if (depth > 0) {
      compile_quasicompiler(op, oargs, cdr(arg), depth − 1, bfalse);
      emitop(OP_SYNTAX);
      emit(Sym_SYNTAX_UNSPLICE);
      if (in_list_p) emitop(OP_CONS);
    }
    else {⟨ Compile unquote-splicing 204 ⟩}
  }
  else error (ERR_UNIMPLEMENTED, NIL);
```

204. Splicing Lists. If not recursing back into the quasicompiler at a lower depth then we are quasi-compiling at the lowest depth and need to do the work.

When splicing into the tail position of a list we can replace its `NIL` with the evaluation with minimal further processing. Unfortunately we don't know until runtime whether we are splicing into the tail position – consider constructs like `` `(,@foo ,@bar) `` where `bar` evaluates to `NIL`.

⟨ Compile unquote-splicing 204 ⟩ ≡
```
int goto_inject_iterate, goto_inject_start, goto_finish;
int goto_list_p, goto_null_p, goto_nnull_p;
if (¬in_list_p) error (ERR_UNEXPECTED, arg);
emitop(OP_PEEK);
emitop(OP_NULL_P);
emitop(OP_JUMP_TRUE); goto_null_p = comefrom();
emitop(OP_PUSH);    /* save FALSE */
emitop(OP_JUMP); goto_nnull_p = comefrom();
patch(goto_null_p, int_new(Here));
emitop(OP_SWAP);    /* become the tail, save TRUE */
patch(goto_nnull_p, int_new(Here));
```
See also sections 205, 206, and 207.
This code is used in section 203.

205. `FALSE` or `TRUE` is now atop the stack indicating whether a new list is being built otherwise the remainder of the list is left on the stack. Now we can evaluate and validate the expression.

⟨ Compile unquote-splicing 204 ⟩ +≡
```
compile_expression(cdr(arg), 0);
emitop(OP_PUSH);
emitop(OP_LIST_P); emit(TRUE); emit(FALSE);
emitop(OP_JUMP_TRUE); goto_list_p = comefrom();
emitq(Sym_ERR_UNEXPECTED);
emitop(OP_ERROR);
```

206. If we have a list we can leave it as-is if we were originally in the tail position.

⟨ Compile unquote-splicing 204 ⟩ +≡
```
patch(goto_list_p, int_new(Here));
emitop(OP_POP);
emitop(OP_SWAP);
emitop(OP_JUMP_TRUE); goto_finish = comefrom();
```

207. Splicing a list into the middle of another list is done item-by-item in reverse. A small efficiency could be gained here by not walking the list a second time (the first to validate it above) at the cost of more complex bytecode.

By now the evaluated list to splice in is first on the stack followed by the partial result.

⟨ Compile unquote-splicing 204 ⟩ +≡
```
emitop(OP_POP);
emitop(OP_LIST_REVERSE); emit(TRUE); emit(FALSE);
⟨ Walk through the splicing list 208 ⟩
```

208. ⟨ Walk through the splicing list 208 ⟩ ≡

```
emitop(OP_JUMP); goto_inject_start = comefrom( );
goto_inject_iterate = Here;
emitop(OP_POP);
emitop(OP_SNOC);
emitop(OP_CYCLE);
emitop(OP_CONS);
emitop(OP_SWAP);
```

See also section 209.

This code is used in section 207.

209. If this was the last item (the first of the evaluated list's) or the evaluation was NIL then we're done otherwise we go around again. This is also where the loop starts to handle the case of evaluating an empty list.

⟨ Walk through the splicing list 208 ⟩ +≡

```
patch(goto_inject_start, int_new(Here));
emitop(OP_PUSH);
emitop(OP_NULL_P);
emitop(OP_JUMP_FALSE); emit(int_new(goto_inject_iterate));
emitop(OP_POP);
patch(goto_finish, int_new(Here));
emitop(OP_POP);
```

210. Testing. A comprehensive test suite is planned for `LossLess` but a testing tool would be no good if it wasn't itself reliable, which these primarily unit tests work towards. In addition to the main library `lossless.o` two libraries with extra functionality needed by the tests are created: `t/lltest.o` and `t/llalloc.o` which additionally to extra operators wraps `reallocarray` to test memory allocation.

```
⟨ t/lltest.c 210 ⟩ ≡
#define LL_TEST
#include "../lossless.c"    /* C source */
```

211.
```
⟨ Global variables 6 ⟩ +≡
#ifdef LL_TEST
  int Allocate_Success = −1;
#endif
```

212.
```
⟨ Externalised global variables 7 ⟩ +≡
#ifdef LL_TEST
  extern int Allocate_Success;
#endif
```

213.
```
⟨ t/llalloc.c 213 ⟩ ≡
#define LL_ALLOCATE fallible_reallocarray
  ⟨ System headers 4 ⟩
  void *fallible_reallocarray(void *, size_t, size_t);
#define LL_TEST
#include "../lossless.c"    /* C source */
  void *fallible_reallocarray(void *ptr, size_t nmemb, size_t size)
  {
    return Allocate_Success−− ? reallocarray(ptr, nmemb, size) : Λ;
  }
```

214. Tests need to be able to save data from the maw of the garbage collector.

```
⟨ Global variables 6 ⟩ +≡
  cell Tmp_Test = NIL;
```

215.
```
⟨ Externalised global variables 7 ⟩ +≡
  extern cell Tmp_Test;
```

216.
```
⟨ Protected Globals 21 ⟩ +≡
#ifdef LL_TEST
  &Tmp_Test ,
#endif
```

217.
```
⟨ Pre-initialise Small_Int & other gc-sensitive buffers 15 ⟩ +≡
#ifdef LL_TEST
  Tmp_Test = NIL;
#endif
```

218. Some tests need to examine a snapshot of the interpreter's run-time state which they do by calling *test!probe*.

```
#define object_copy(o, d, p) object_copy_imp((o), (d), (p), 0)
#define object_copyref(o, d) object_copyref_imp((o), (d), 0)
```

⟨ Function declarations 8 ⟩ +≡

```
void compile_testing_probe(cell, cell, boolean);
void compile_testing_probe_app(cell, cell, boolean);
boolean object_compare(char *, size_t, cell, boolean);
int object_copy_imp(cell, char *, boolean, int);
int object_copyref_imp(cell, cell *, int);
size_t object_sizeof(cell);
size_t object_sizeofref(cell);
cell testing_build_probe(cell);
```

219. ⟨ Testing opcodes 219 ⟩ ≡

```
OP_TEST_PROBE ,
```

This code is used in sections 140 and 141.

220. ⟨ Testing implementations 220 ⟩ ≡

```
case OP_TEST_PROBE:
  Acc = testing_build_probe(rts_pop(1));
  skip(1);
  break;
```

This code is used in section 108.

221. ⟨ Testing primitives 221 ⟩ ≡

```
{"test!probe", compile_testing_probe},
{"test!probe-applying", compile_testing_probe_app} ,
```

This code is used in section 377.

222.

```
void compile_testing_probe(cell op __unused, cell args, boolean tail_p __unused)
{
  emitop(OP_PUSH);
  emitq(args);
  emitop(OP_TEST_PROBE);
}
```

223. This variant evaluates its run-time arguments first.

```
void compile_testing_probe_app(cell op __unused, cell args, boolean tail_p __unused)
{
  emitop(OP_PUSH);
  cts_push(args = list_reverse(args, Λ, Λ));
  emitq(NIL);
  for ( ; pair_p(args); args = cdr(args)) {
    emitop(OP_PUSH);
    compile_expression(car(args), bfalse);
    emitop(OP_CONS);
  }
  cts_pop();
  emitop(OP_TEST_PROBE);
}
```

224. TODO: This should make a deep copy of the objects not merely reference them.

```
size_t object_sizeof(cell o)
{
  size_t s;
  int i;
  if (special_p(o)) return sizeof(char);
  s = sizeof(char) + 2 * sizeof(cell);
  if (acar_p(o)) s += object_sizeof(car(o));
  if (acdr_p(o)) s += object_sizeof(cdr(o));
  if (vector_p(o)) {
    s += (vector_length(o) + VECTOR_HEAD) * sizeof(cell);
    for (i = 0; i < vector_length(o); i++) s += object_sizeof(vector_ref(o, i));
  }
  return s;
}
int object_copy_imp(cell o, char *dst, boolean offset_p, int p)
{
  int i;
  if (special_p(o)) dst[p++] = (char) o;
  else {
    bcopy(&tag(o), dst + p, sizeof(char));
    p++;
    if (¬vector_p(o))    /* car is gc's index */
      bcopy(&car(o), dst + p, sizeof(cell));
    else bzero(dst + p, sizeof(cell));
    p += sizeof(cell);
    if (¬vector_p(o) ∨ offset_p) bcopy(&cdr(o), dst + p, sizeof(cell));
    else bzero(dst + p, sizeof(cell));
    p += sizeof(cell);
    if (acar_p(o)) p = object_copy_imp(car(o), dst, offset_p, p);
    if (acdr_p(o)) p = object_copy_imp(cdr(o), dst, offset_p, p);
    if (vector_p(o)) {
      bcopy(&(vector_ref(o, 0)) − VECTOR_HEAD, dst + p, sizeof(cell) * (vector_length(o) + VECTOR_HEAD));
      p += sizeof(cell) * (vector_length(o) + VECTOR_HEAD);
      for (i = 0; i < vector_length(o); i++) p = object_copy_imp(vector_ref(o, i), dst, offset_p, p);
    }
  }
  return p;
}
boolean object_compare(char *buf1, size_t len, cell o2, boolean offset_p)
{
  char *buf2;
  boolean r;
  if (object_sizeof(o2) ≠ len) return bfalse;
  ERR_OOM_P(buf2 = malloc(len));
  object_copy(o2, buf2, offset_p);
  r = (bcmp(buf1, buf2, len) ≡ 0) ? btrue : bfalse;
  free(buf2);
  return r;
}
```

```
size_t object_sizeofref(cell o)
{
  int i;
  size_t s = 1;
  if (special_p(o)) return s;
  if (acar_p(o)) s += object_sizeofref(car(o));
  if (acdr_p(o)) s += object_sizeofref(cdr(o));
  if (vector_p(o))
    for (i = 0; i < vector_length(o); i++) s += object_sizeofref(vector_ref(o, i));
  return s;
}
int object_copyref_imp(cell o, cell *dst, int p)
{
  int i;
  dst[p++] = o;
  if (special_p(o)) return p;
  if (acar_p(o)) p = object_copyref_imp(car(o), dst, p);
  if (acdr_p(o)) p = object_copyref_imp(cdr(o), dst, p);
  if (vector_p(o))
    for (i = 0; i < vector_length(o); i++) p = object_copyref_imp(vector_ref(o, i), dst, p);
  return p;
}
#define probe_push(n, o) do {
    vms_push(cons((o), NIL));
    vms_set(cons(sym(n), vms_ref()));
    t = vms_pop();
    vms_set(cons(t, vms_ref()));
  } while (0)
cell testing_build_probe(cell was_Acc)
{
  cell t;
  vms_push(NIL);
  probe_push("Acc", was_Acc);
  probe_push("Args", Acc);
  probe_push("Env", Env);
  return vms_pop();
}
#undef probe_push
```

225.

```
#define test_copy_env() Env
#define test_compare_env(o) ((o) ≡ Env)
#define test_is_env(o, e) ((o) ≡ (e))
```

⟨ Old test executable wrapper 225 ⟩ ≡

```
#define LL_TEST 1
#include "lossless.h"
  void test_main(void);
  int main(int argc __unused, char **argv __unused)
  {
    volatile boolean first = btrue;
    vm_init();
    if (argc > 1) error(ERR_ARITY_EXTRA, NIL);
    vm_prepare();
    if (¬first) {
      printf("Bail␣out!␣Unhandled␣exception␣in␣test\n");
      return EXIT_FAILURE;
    }
    first = bfalse;
    test_main();
    tap_plan(0);
    return EXIT_SUCCESS;
  }
```

This code is used in sections 332, 344, 353, 362, 369, and 376.

226. The Perl ecosystem has a well-deserved reputation for its thorough testing regime and the quality (if not necessarily the quality) of the results so LossLess is deliberately aping the interfaces that were developed there.

The LossLess internal tests are a collection of test “script”s each of which massages some LossLess function or other and then reports what happened in a series of binary pass/fail “test”s. A test in this sense isn’t the performance of any activity but comparing the result of having *already performed* some activity with the expected outcome. Any one action normally requires a lot of individual tests to confirm the validity of its result. Occasionally “test” refers to a collection of these tests which are performed together, which is a bad habit.

This design is modelled on the Test Anything Protocol and the test scripts call an API that looks suspiciously like a tiny version of *Test::Simple*.

tap_plan is optionally called before the test script starts if the total number of tests is known in advance and then again at the end of testing with an argument of 0 to emit exactly one test plan.

```
#define tap_fail(m) tap_ok(bfalse, (m))
#define tap_pass(m) tap_ok(btrue, (m))
#define tap_again(t, r, m) tap_ok(((t) = ((t) ∧ (r))), (m))    /* intentional assignment */
#define tap_more(t, r, m) (t) &= tap_ok((r), (m))
#define tap_or(p, m) if (¬tap_ok((p), (m)))
```

⟨ Function declarations 8 ⟩ +≡

```
#ifdef LL_TEST
  void tap_plan(int);
  boolean tap_ok(boolean, char *);
  char *test_msgf(char *, const char *, char *, ...);
  void test_vm_state(char *, int);
#endif
```

227. ⟨ Global variables 6 ⟩ +≡

```
boolean Test_Passing = btrue;
int Test_Plan = −1;
int Next_Test = 1;    /* not 0 */
```

228. ⟨ Externalised global variables 7 ⟩ +≡

```
extern int Test_Plan, Next_Test;
```

229.

```
void tap_plan(int plan)
{
  if (plan ≡ 0) {
    if (Test_Plan < 0) printf("1..%d\n", Next_Test − 1);
    else if (Next_Test − 1 ≠ Test_Plan) {
      printf("#␣Planned␣%3$d␣%1$s␣but␣ran␣%2$s%4$d!\n", (Test_Plan ≡ 1 ? "test" : "tests"),
          (Next_Test ≤ Test_Plan ? "only␣" : ""), Test_Plan, Next_Test − 1);
      Test_Passing = bfalse;
    }
    return;
  }
  if (Test_Plan > 0) error ("plan-exists", int_new(Test_Plan));
  if (plan < 0) error (ERR_UNEXPECTED, cons(sym("test-plan"), int_new(plan)));
  Test_Plan = plan;
  printf("1..%d\n", plan);
}
```

230.

```
boolean tap_ok(boolean result, char *message)
{
  printf("%s␣%d␣%s\n", (result ? "ok" : "not␣ok"),
      Next_Test++,
      (message ∧ *message) ? message : "?");
  if (result) return btrue;
  return Test_Passing = bfalse;
}
```

231. LossLess is a programming language and so a lot of its tests involve code. *test_vmsgf* formats messages describing tests which involve code (or any other s-expression) in a consistent way. The caller is expected to maintain its own buffer of TEST_BUFSIZE bytes a pointer to which goes in and out so that the function can be used in-line.

tmsgf hardcodes the names of the variables a function passes into *test_vmsgf* for brevity.

```
#define TEST_BUFSIZE 1024
```

⟨ Complex definitions & macros 140 ⟩ +≡

```
#define tmsgf(...) test_msgf (msg, prefix, __VA_ARGS__)
```

232. **char** **test_msgf*(**char** **tmsg*, **const char** **tsrc*, **char** **fmt*, ...)

```
{
  char ttmp[TEST_BUFSIZE] = {0};
  int ret;
  va_list ap;
  va_start(ap, fmt);
  ret = vsnprintf(ttmp, TEST_BUFSIZE, fmt, ap);
  va_end(ap);
  snprintf(tmsg, TEST_BUFSIZE, "%s: %s", tsrc, ttmp);
  return tmsg;
}
```

233. The majority of tests validate some parts of the VM state, which parts is controlled by the *flags* parameter.

```
#define TEST_VMSTATE_RUNNING #01
#define TEST_VMSTATE_NOT_RUNNING #00
#define TEST_VMSTATE_INTERRUPTED #02
#define TEST_VMSTATE_NOT_INTERRUPTED #00
#define TEST_VMSTATE_VMS #04
#define TEST_VMSTATE_CTS #08
#define TEST_VMSTATE_RTS #10
#define TEST_VMSTATE_STACKS (TEST_VMSTATE_VMS | TEST_VMSTATE_CTS | TEST_VMSTATE_RTS)
#define TEST_VMSTATE_ENV_ROOT #20
#define TEST_VMSTATE_PROG_MAIN #40
#define test_vm_state_full(p)
        test_vm_state((p), TEST_VMSTATE_NOT_RUNNING | TEST_VMSTATE_NOT_INTERRUPTED |
            TEST_VMSTATE_ENV_ROOT | TEST_VMSTATE_PROG_MAIN | TEST_VMSTATE_STACKS)
#define test_vm_state_normal(p)
        test_vm_state((p), TEST_VMSTATE_NOT_RUNNING | TEST_VMSTATE_NOT_INTERRUPTED |
            TEST_VMSTATE_PROG_MAIN | TEST_VMSTATE_STACKS)    /* ¬TEST_VMSTATE_ENV_ROOT */
  void test_vm_state(char *prefix, int flags)
  {
    char msg[TEST_BUFSIZE] = {0};
    if (flags & TEST_VMSTATE_RUNNING) tap_ok(Running, tmsgf("(== Running 1)"));
    else tap_ok(¬Running, tmsgf("(== Running 0)"));
    if (flags & TEST_VMSTATE_INTERRUPTED) tap_ok(Interrupt, tmsgf("(== Interrupt 1)"));
    else tap_ok(¬Interrupt, tmsgf("(== Interrupt 0)"));
    if (flags & TEST_VMSTATE_VMS) tap_ok(null_p(VMS), tmsgf("(null? VMS)"));
    if (flags & TEST_VMSTATE_CTS) tap_ok(null_p(CTS), tmsgf("(null? CTS)"));
    if (flags & TEST_VMSTATE_RTS) tap_ok(RTSp ≡ −1, tmsgf("(== RTSp -1)"));
    if (flags & TEST_VMSTATE_ENV_ROOT) tap_ok(Env ≡ Root, tmsgf("(== Env Root)"));
    if (flags & TEST_VMSTATE_PROG_MAIN) {
      tap_ok(Prog ≡ Prog_Main, tmsgf("Prog_Main is returned to"));
      tap_ok(Ip ≡ vector_length(Prog_Main) − 1, tmsgf("Prog_Main is completed"));
    }
        /* TODO? Others: root unchanged; */
  }
```

234. Sanity Test. This seemingly pointless test achieves two goals: the test harness can run it first and can abort the entire test suite if it fails, and it provides a simple demonstration of how individual test scripts interact with the harness, without obscuring it with the more complicated unit test framework below.

⟨ t/sanity.c 234 ⟩ ≡

```
#define LL_TEST
#include "lossless.h"
  int main()
  {
    tap_plan(1);
    vm_init();
    vm_prepare();
    vm_reset();
    interpret();
    tap_pass("LossLess compiles and runs");
  }
```

235. Unit Tests. This is the very boring process of laboriously checking that each function or otherwise segregable unit of code does what it says on the tin. For want of a better model to follow I've taken inspiration from Mike Bland's article "Goto Fail, Heartbleed, and Unit Testing Culture" describing how he created unit tests for the major OpenSSL vulnerabilities known as "goto fail" and "Heartbleed". The article itself is behind some sort of Google wall but Martin Fowler has reproduced it at https://martinfowler.com/articles/testing-culture.html.

```
⟨ t/llt.h 235 ⟩ ≡
#ifndef LLT_H
#define LLT_H
  ⟨ Unit test fixture header 236 ⟩
  typedef struct llt_Fixture llt_Fixture;     /* user-defined */
  typedef void (*llt_thunk)(llt_Fixture *);
  typedef boolean (*llt_unit)(llt_Fixture *);
  typedef llt_Fixture *(*llt_fixture)(void);
  extern llt_fixture Test_Fixtures[];     /* user-defined */
#define fmsgf(...)test_msgf (buf, fix.name, __VA_ARGS__)
#define fpmsgf(...)test_msgf (buf, fix→name, __VA_ARGS__)
  llt_Fixture *llt_alloc(size_t);
  boolean llt_main(llt_Fixture *);
  llt_Fixture *llt_prepare(void);
#endif     /* LLT_H */
```

236. Unit test fixtures are defiend in a **llt_Fixture** structure which is only declared in this header; it is up to each unit test to implement its own **llt_Fixture** with this common header.

```
⟨ Unit test fixture header 236 ⟩ ≡
#define LLT_FIXTURE_HEADER
        const char *name;
        const char *suffix;
        int id;
        int max;
        llt_thunk prepare;
        llt_thunk act;
        llt_unit test;
        llt_thunk destroy     /* no semicolon */
```

This code is used in section 235.

237. The vast majority (all, so far) of unit tests follow the same simple structure. There are plans for more interactive tests but they aren't necessary yet.

```
⟨ Unit test header 237 ⟩ ≡
#define LL_TEST
#include "lossless.h"
#include "llt.h"
```

This code is used in sections 243, 262, 275, 308, and 318.

238. ⟨ Unit test body 238 ⟩ ≡

```
int main(int argc __unused, char **argv __unused)
{
   llt_Fixture *suite;
   if (argc > 1) {
      printf("usage: %s", argv[0]);
      return EXIT_FAILURE;
   }
#ifndef LLT_NOINIT
   vm_init();
#endif
   suite = llt_prepare();
   llt_main(suite);
   free(suite);
   tap_plan(0);
}
```

See also sections 239, 240, and 241.
This code is used in sections 243, 262, 275, 308, and 318.

239. ⟨ Unit test body 238 ⟩ +≡

```
llt_Fixture *llt_alloc(size_t n)
{
   llt_Fixture *f;
   size_t i;
   ERR_OOM_P(f = calloc(n, sizeof (llt_Fixture)));
   for (i = 0; i < n; i++) f[i].max = n;
   return f;
}
```

240. ⟨ Unit test body 238 ⟩ +≡

```
boolean llt_main(llt_Fixture *suite)
{
   int i;
   boolean all, ok;
   char buf[TEST_BUFSIZE] = {0};
   ok = btrue;
   for (i = 0; i < suite->max; i++) {
      if (suite[i].prepare) suite[i].prepare((suite + i));
      suite[i].act((suite + i));
      if (suite[i].suffix) snprintf(buf, TEST_BUFSIZE, "%s (%s)", suite[i].name, suite[i].suffix);
      else snprintf(buf, TEST_BUFSIZE, "%s", suite[i].name);
      suite[i].name = (char *) buf;
      ok = suite[i].test((suite + i));
      tap_ok(ok, buf);
      all = all && ok;
      if (suite[i].destroy) suite[i].destroy((suite + i));
   }
   return all;
}
```

241. ⟨ Unit test body 238 ⟩ +≡

```
llt_Fixture *llt_prepare(void)
{
  llt_fixture *s = Test_Fixtures, *t;
  llt_Fixture *r = Λ, *q, *p;
  int c = 0, i;
  int f = sizeof(llt_Fixture);
  for (t = s; *t; t++) {
    q = (*t)();
    p = reallocarray(r, c + q→max, f);
    ERR_OOM_P(p);
    r = p;
    bcopy(q, r + c, f * q→max);
    c += q→max;
    free(q);
  }
  for (i = 0; i < c; i++) {
    r[i].id = i;
    r[i].max = c;
  }
  return r;
}
```

242. Heap Allocation. The first units we test are the memory allocators because I've already found embarrassing bugs there proving that even that "obvious" code needs manual verification. To do that we will need to be able to make *reallocarray* fail without actually exhausting the system's memory. A global counter is decremented each time this variant is called and returns Λ if it reaches zero.

This method of implementing unit tests has us pose 5 questions:

1. *What is the contract fulfilled by the code under test?*

new_cells_segment performs 3, or 5 if each allocation is counted seperately, actions: Enlarge each of `CAR`, `CDR` & `TAG` in turn, checking for out-of-memory for each; zero-out the newly-allocated range of memory; update the global counters *Cells_Poolsize* & *Cells_Segment*.

There is no return value but either the heap will have been enlarged or one of 3 (mostly identical) errors will have been raised.

2. *What preconditions are required, and how are they enforced?*

Cells_Segment describes how much the pool will grow by. If *Cells_Poolsize* is 0 the three pointers must be Λ otherwise they each point to an area of allocated memory *Cells_Poolsize* elements wide. There is no explicit enforcement.

3. *What postconditions are guaranteed?*

IFF there was an allocation error for any of the 3 pools, the pointer under question will not have changed but those reallocated before it may have. *Cells_Poolsize* & *Cells_Segment* will be unchanged. Any newly-allocated memory should not be considered available

Otherwise `CAR`, `CDR` & `TAG` will point to still-valid memory but possibly at the same address.

The newly allocated memory will have been zerod.

Cells_Poolsize & *Cells_Segment* will have been enlarged.

new_cells_segment also guarantees that previously-allocated data will not have changed but it's safe for now to rely on *reallocarray* getting that right.

4. *What example inputs trigger different behaviors?*

Chiefly there are two classes of inputs, whether or not *Cells_Poolsize* is 0, and whether allocation succeeds for each of the 3 attempts.

5. *What set of tests will trigger each behavior and validate each guarantee?*

Eight tests, four starting from no heap and four from a heap with data in it. One for success and one for each potentially failed allocation.

243. This unit test relies on the VM being uninitialised so that it can safely switch out the heap pointers. The *save_CAR*, *save_CDR* & *save_TAG* pointers in the fixture are convenience pointers into *heapcopy*.

```
⟨ t/cell-heap.c 243 ⟩ ≡
#define LLT_NOINIT
  ⟨ Unit test header 237 ⟩
  enum llt_Grow_Pool_result {
    LLT_GROW_POOL_SUCCESS, LLT_GROW_POOL_FAIL_CAR, LLT_GROW_POOL_FAIL_CDR,
        LLT_GROW_POOL_FAIL_TAG
  };
  struct llt_Fixture {
    LLT_FIXTURE_HEADER;
    enum llt_Grow_Pool_result expect;
    int allocations;
    int Poolsize;
    int Segment;
    cell *CAR;
    cell *CDR;
    char *TAG;
    char *heapcopy;
    cell *save_CAR;
    cell *save_CDR;
    char *save_TAG;
  };
  ⟨ Unit test body 238 ⟩
  ⟨ Unit test: grow heap pool 244 ⟩
  llt_fixture Test_Fixtures[] = {
      llt_Grow_Pool__Initial_Success, llt_Grow_Pool__Immediate_Fail, llt_Grow_Pool__Second_Fail,
      llt_Grow_Pool__Third_Fail, llt_Grow_Pool__Full_Success, llt_Grow_Pool__Full_Immediate_Fail,
      llt_Grow_Pool__Full_Second_Fail, llt_Grow_Pool__Full_Third_Fail, Λ
  };
```

244. ⟨ Unit test: grow heap pool 244 ⟩ ≡

```
void llt_Grow_Pool_prepare(llt_Fixture *fix)
{
  if (fix→Poolsize) {
    int cs = fix→Poolsize;
    fix→heapcopy = reallocarray(Λ, cs, 2 * sizeof(cell) + sizeof(char));
    fix→save_CAR = (cell *) fix→heapcopy;
    fix→save_CDR = (cell *)(fix→heapcopy + sizeof(cell) * cs);
    fix→save_TAG = fix→heapcopy + sizeof(cell) * cs * 2;
    bcopy(fix→CAR, fix→save_CAR, sizeof(cell) * cs);
    bcopy(fix→CDR, fix→save_CDR, sizeof(cell) * cs);
    bcopy(fix→TAG, fix→save_TAG, sizeof(char) * cs);
  }
  CAR = fix→CAR;
  CDR = fix→CDR;
  TAG = fix→TAG;
  Cells_Poolsize = fix→Poolsize;
  Cells_Segment = fix→Segment;
}
```

See also sections 245, 246, 247, 252, 253, 254, 255, 256, 257, 258, 259, 260, and 261.
This code is used in section 243.

245. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
void llt_Grow_Pool_destroy(llt_Fixture *fix)
{
  free(CAR);
  free(CDR);
  free(TAG);
  free(fix→heapcopy);
  CAR = CDR = Λ;
  TAG = Λ;
  Cells_Poolsize = 0;
  Cells_Segment = HEAP_SEGMENT;
}
```

246. There is not much for this test to do apart from prepare state and call *new_cells_segment* then validate that the memory was, or was not, correctly reallocated.

⟨ Unit test: grow heap pool 244 ⟩ +≡

```
void llt_Grow_Pool_act(llt_Fixture *fix)
{
  jmp_buf save_jmp;
  Allocate_Success = fix→allocations;
  memcpy(&save_jmp, &Goto_Begin, sizeof(jmp_buf));
  if (¬setjmp(Goto_Begin)) new_cells_segment();
  Allocate_Success = −1;
  memcpy(&Goto_Begin, &save_jmp, sizeof(jmp_buf));
}
```

247. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
boolean llt_Grow_Pool_test(llt_Fixture *fix)
{
  boolean ok;
  char buf[TEST_BUFSIZE] = {0};
  switch (fix→expect) {
  case LLT_GROW_POOL_SUCCESS:
    ⟨ Unit test part: grow heap pool, validate success 248 ⟩
    break;    /* TODO: test for bzero */
  case LLT_GROW_POOL_FAIL_CAR:
    ⟨ Unit test part: grow heap pool, validate car failure 249 ⟩
    break;
  case LLT_GROW_POOL_FAIL_CDR:
    ⟨ Unit test part: grow heap pool, validate cdr failure 250 ⟩
    break;
  case LLT_GROW_POOL_FAIL_TAG:
    ⟨ Unit test part: grow heap pool, validate tag failure 251 ⟩
    break;
  }
  return ok;
}
```

248. ⟨ Unit test part: grow heap pool, validate success 248 ⟩ ≡

```
ok = tap_ok(Cells_Poolsize ≡ (fix→Poolsize + fix→Segment), fpmsgf("Cells_Poolsize is increased"));
tap_more(ok, Cells_Segment ≡ (fix→Poolsize + fix→Segment)/2, fpmsgf("Cells_Segment is increased"));
tap_more(ok, CAR ≠ CDR ∧ CAR ≠ (cell *) TAG, fpmsgf("CAR, CDR & TAG are unique"));
tap_more(ok, CAR ≠ Λ, fpmsgf("CAR is not NULL"));
tap_more(ok, ¬bcmp(CAR, fix→save_CAR, sizeof(cell) * fix→Poolsize), fpmsgf("CAR heap is unchanged"));
tap_more(ok, CDR ≠ Λ, fpmsgf("CDR is not NULL"));
tap_more(ok, ¬memcmp(CDR, fix→save_CDR, sizeof(cell) * fix→Poolsize),
    fpmsgf("CDR heap is unchanged"));
tap_more(ok, TAG ≠ Λ, fpmsgf("TAG is not NULL"));
tap_more(ok, ¬memcmp(TAG, fix→save_TAG, sizeof(char) * fix→Poolsize),
    fpmsgf("TAG heap is unchanged"));
```

This code is used in section 247.

249. ⟨ Unit test part: grow heap pool, validate car failure 249 ⟩ ≡

```
ok = tap_ok(Cells_Poolsize ≡ fix→Poolsize, fpmsgf("Cells_Poolsize is not increased"));
tap_more(ok, Cells_Segment ≡ fix→Segment, fpmsgf("Cells_Segment is not increased"));
tap_more(ok, CAR ≡ fix→CAR, fpmsgf("CAR is unchanged"));
tap_more(ok, CDR ≡ fix→CDR, fpmsgf("CDR is unchanged"));
tap_more(ok, TAG ≡ fix→TAG, fpmsgf("TAG is unchanged"));
```

This code is used in section 247.

250. ⟨ Unit test part: grow heap pool, validate cdr failure 250 ⟩ ≡

```
ok = tap_ok(Cells_Poolsize ≡ fix→Poolsize, fpmsgf("Cells_Poolsize is not increased"));
tap_more(ok, Cells_Segment ≡ fix→Segment, fpmsgf("Cells_Segment is not increased"));
tap_more(ok, ¬memcmp(CAR, fix→save_CAR, sizeof(cell) * fix→Poolsize),
    fpmsgf("CAR heap is unchanged"));
tap_more(ok, CDR ≡ fix→CDR, fpmsgf("CDR is unchanged"));
tap_more(ok, TAG ≡ fix→TAG, fpmsgf("TAG is unchanged"));
```

This code is used in section 247.

251. ⟨ Unit test part: grow heap pool, validate tag failure 251 ⟩ ≡

```
ok = tap_ok(Cells_Poolsize ≡ fix→Poolsize, fpmsgf("Cells_Poolsize␣is␣not␣increased"));
tap_more(ok, Cells_Segment ≡ fix→Segment, fpmsgf("Cells_Segment␣is␣not␣increased"));
tap_more(ok, ¬memcmp(CAR, fix→save_CAR, sizeof(cell) * fix→Poolsize),
    fpmsgf("CAR␣heap␣is␣unchanged"));
tap_more(ok, ¬memcmp(CDR, fix→save_CDR, sizeof(cell) * fix→Poolsize),
    fpmsgf("CDR␣heap␣is␣unchanged"));
tap_more(ok, TAG ≡ fix→TAG, fpmsgf("TAG␣is␣unchanged"));
```

This code is used in section 247.

252. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool_fix(llt_Fixture *fix, const char *name)
{
  fix→name = name;
  fix→prepare = llt_Grow_Pool_prepare;
  fix→destroy = llt_Grow_Pool_destroy;
  fix→act = llt_Grow_Pool_act;
  fix→test = llt_Grow_Pool_test;
  fix→expect = LLT_GROW_POOL_SUCCESS;
  fix→allocations = −1;
  fix→Segment = HEAP_SEGMENT;
  return fix;
}
```

253. This tests that allocation is successful the first time the heap is ever allocated. It is the simplest test in this unit.

⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Initial_Success(void)
{
  return llt_Grow_Pool_fix(llt_alloc(1), __func__);
}
```

254. If the very first call to *reallocarray* fails then everything should remain unchanged.

⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Immediate_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_POOL_FAIL_CAR;
  fix→allocations = 0;
  return fix;
}
```

255. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Second_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_POOL_FAIL_CDR;
  fix→allocations = 1;
  return fix;
}
```

256. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Third_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_POOL_FAIL_TAG;
  fix→allocations = 2;
  return fix;
}
```

257. Data already on the heap must be preserved exactly.

⟨ Unit test: grow heap pool 244 ⟩ +≡

```
void llt_Grow_Pool__fill(llt_Fixture *fix)
{
  size_t i;
  fix→CAR = reallocarray(Λ, fix→Poolsize, sizeof(cell));
  fix→CDR = reallocarray(Λ, fix→Poolsize, sizeof(cell));
  fix→TAG = reallocarray(Λ, fix→Poolsize, sizeof(char));
  for (i = 0; i < (fix→Poolsize * sizeof(cell))/sizeof(int); i++) *(((int *) fix→CAR) + i) = rand();
  for (i = 0; i < (fix→Poolsize * sizeof(cell))/sizeof(int); i++) *(((int *) fix→CDR) + i) = rand();
  for (i = 0; i < (fix→Poolsize * sizeof(char))/sizeof(int); i++) *(((int *) fix→TAG) + i) = rand();
}
```

258. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Full_Success(void)
{
  llt_Fixture *fix = llt_Grow_Pool_fix(llt_alloc(1), __func__);
  fix→Poolsize = HEAP_SEGMENT;
  llt_Grow_Pool__fill(fix);
  return fix;
}
```

259. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Full_Immediate_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_POOL_FAIL_CAR;
  fix→allocations = 0;
  fix→Poolsize = HEAP_SEGMENT;
  llt_Grow_Pool__fill(fix);
  return fix;
}
```

260. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Full_Second_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_POOL_FAIL_CDR;
  fix→allocations = 1;
  fix→Poolsize = HEAP_SEGMENT;
  llt_Grow_Pool__fill(fix);
  return fix;
}
```

261. ⟨ Unit test: grow heap pool 244 ⟩ +≡

```
llt_Fixture *llt_Grow_Pool__Full_Third_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_POOL_FAIL_TAG;
  fix→allocations = 2;
  fix→Poolsize = HEAP_SEGMENT;
  llt_Grow_Pool__fill(fix);
  return fix;
}
```

262. Vector Heap. Testing the vector's heap is the same but simpler because it has 1 not 3 possible error conditions so this section is duplicated from the previous without further explanation.

```
⟨ t/vector-heap.c 262 ⟩ ≡
#define LLT_NOINIT
  ⟨ Unit test header 237 ⟩
  enum llt_Grow_Vector_Pool_result {
    LLT_GROW_VECTOR_POOL_SUCCESS, LLT_GROW_VECTOR_POOL_FAIL
  };
  struct llt_Fixture {
    LLT_FIXTURE_HEADER;
    enum llt_Grow_Vector_Pool_result expect;
    int allocations;
    int Poolsize;
    int Segment;
    cell *VECTOR;
    cell *save_VECTOR;
  };
  ⟨ Unit test body 238 ⟩
  ⟨ Unit test: grow vector pool 263 ⟩
  llt_fixture Test_Fixtures[ ] = {
      llt_Grow_Vector_Pool__Empty_Success, llt_Grow_Vector_Pool__Empty_Fail,
      llt_Grow_Vector_Pool__Full_Success, llt_Grow_Vector_Pool__Full_Fail, Λ
  };
```

263. ⟨ Unit test: grow vector pool 263 ⟩ ≡

```
  void llt_Grow_Vector_Pool_prepare(llt_Fixture *fix)
  {
    if (fix→Poolsize) {
      int cs = fix→Poolsize;
      fix→save_VECTOR = reallocarray(Λ, cs, sizeof(cell));
      bcopy(fix→VECTOR, fix→save_VECTOR, sizeof(cell) * cs);
    }
    VECTOR = fix→VECTOR;
    Vectors_Poolsize = fix→Poolsize;
    Vectors_Segment = fix→Segment;
  }
```

See also sections 264, 265, 266, 269, 270, 271, 272, 273, and 274.

This code is used in section 262.

264. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
  void llt_Grow_Vector_Pool_destroy(llt_Fixture *fix)
  {
    free(VECTOR);
    free(fix→save_VECTOR);
    VECTOR = Λ;
    Vectors_Poolsize = 0;
    Vectors_Segment = HEAP_SEGMENT;
  }
```

265. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
void llt_Grow_Vector_Pool_act(llt_Fixture *fix)
{
  jmp_buf save_jmp;
  Allocate_Success = fix→allocations;
  memcpy(&save_jmp, &Goto_Begin, sizeof(jmp_buf));
  if (¬setjmp(Goto_Begin)) new_vector_segment();
  Allocate_Success = −1;
  memcpy(&Goto_Begin, &save_jmp, sizeof(jmp_buf));
}
```

266. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
boolean llt_Grow_Vector_Pool_test(llt_Fixture *fix)
{
  boolean ok;
  char buf[TEST_BUFSIZE] = {0};
  switch (fix→expect) {
  case LLT_GROW_VECTOR_POOL_SUCCESS:
    ⟨ Unit test part: grow vector pool, validate success 267 ⟩
    break;    /* TODO: test for bzero */
  case LLT_GROW_VECTOR_POOL_FAIL:
    ⟨ Unit test part: grow vector pool, validate failure 268 ⟩
    break;
  }
  return ok;
}
```

267. ⟨ Unit test part: grow vector pool, validate success 267 ⟩ ≡

```
ok = tap_ok(Vectors_Poolsize ≡ (fix→Poolsize + fix→Segment),
    fpmsgf("Vectors_Poolsize␣is␣increased"));
tap_more(ok, Vectors_Segment ≡ (fix→Poolsize + fix→Segment)/2,
    fpmsgf("Vectors_Segment␣is␣increased"));
tap_more(ok, VECTOR ≠ Λ, fpmsgf("VECTOR␣is␣not␣NULL"));
tap_more(ok, ¬bcmp(VECTOR, fix→save_VECTOR, sizeof(cell) * fix→Poolsize),
    fpmsgf("VECTOR␣heap␣is␣unchanged"));
```

This code is used in section 266.

268. ⟨ Unit test part: grow vector pool, validate failure 268 ⟩ ≡

```
ok = tap_ok(Vectors_Poolsize ≡ fix→Poolsize, fpmsgf("Vectors_Poolsize␣is␣not␣increased"));
tap_more(ok, Vectors_Segment ≡ fix→Segment, fpmsgf("Vectors_Segment␣is␣not␣increased"));
tap_more(ok, VECTOR ≡ fix→VECTOR, fpmsgf("VECTOR␣is␣unchanged"));
```

This code is used in section 266.

269. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
llt_Fixture *llt_Grow_Vector_Pool_fix(llt_Fixture *fix, const char *name)
{
  fix→name = name;
  fix→prepare = llt_Grow_Vector_Pool_prepare;
  fix→destroy = llt_Grow_Vector_Pool_destroy;
  fix→act = llt_Grow_Vector_Pool_act;
  fix→test = llt_Grow_Vector_Pool_test;
  fix→expect = LLT_GROW_VECTOR_POOL_SUCCESS;
  fix→allocations = −1;
  fix→Segment = HEAP_SEGMENT;
  return fix;
}
```

270. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
llt_Fixture *llt_Grow_Vector_Pool__Empty_Success(void)
{
  return llt_Grow_Vector_Pool_fix(llt_alloc(1), __func__);
}
```

271. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
llt_Fixture *llt_Grow_Vector_Pool__Empty_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Vector_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_VECTOR_POOL_FAIL;
  fix→allocations = 0;
  return fix;
}
```

272. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
void llt_Grow_Vector_Pool__fill(llt_Fixture *fix)
{
  size_t i;
  fix→VECTOR = reallocarray(Λ, fix→Poolsize, sizeof(cell));
  for (i = 0; i < (fix→Poolsize * sizeof(cell))/sizeof(int); i++) *(((int *) fix→VECTOR) + i) = rand();
}
```

273. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
llt_Fixture *llt_Grow_Vector_Pool__Full_Success(void)
{
  llt_Fixture *fix = llt_Grow_Vector_Pool_fix(llt_alloc(1), __func__);
  fix→Poolsize = HEAP_SEGMENT;
  llt_Grow_Vector_Pool__fill(fix);
  return fix;
}
```

274. ⟨ Unit test: grow vector pool 263 ⟩ +≡

```
llt_Fixture *llt_Grow_Vector_Pool__Full_Fail(void)
{
  llt_Fixture *fix = llt_Grow_Vector_Pool_fix(llt_alloc(1), __func__);
  fix→expect = LLT_GROW_VECTOR_POOL_FAIL;
  fix→allocations = 0;
  fix→Poolsize = HEAP_SEGMENT;
  llt_Grow_Vector_Pool__fill(fix);
  return fix;
}
```

275. Garbage Collector. There are three parts to the garbage collector, each building on the last. The inner-most component is *mark* which searches the heap for any data which are in use.

1. *What is the contract fulfilled by the code under test?*
2. *What preconditions are required, and how are they enforced?*
3. *What postconditions are guaranteed?*

Given a **cell**, it and any objects it refers to—recursively, including internal components of atoms—will have their mark flag raised. No other objects will be affected and no other changes will be made to the objects which are. The global constants (specials) are ignored.

mark's main complication is that it's a linear implementation of a recursive algorithm. It can't use any of the real stacks to keep track of the recursion so it uses the individual cells its scanning as an impromptu stack. This heap mutation needs to have no visible external effect despite mutating every **cell** that's considered.

4. *What example inputs trigger different behaviors?*

Global constants and cells already marked vs. unmarked cells. Obviously different objects will be marked in their own way.

Constants aside, the different types of object come in one of 5 categories: pairs, vectors, atomic pairs, atomic lists (the car is opaque) and pure atoms (which are entirely opaque). These are referred to as P, V, A & L respectively.

5. *What set of tests will trigger each behavior and validate each guarantee?*

A test for each type of object—P, V, A & L as well as globals—created without any nesting and one for each recursive combination up to a depth of 3.

```
⟨ t/gc-mark.c 275 ⟩ ≡
  ⟨ Unit test header 237 ⟩
  enum llt_GC_Mark_flat {
    LLT_GC_MARK_SIMPLE_ATOM, LLT_GC_MARK_SIMPLE_LONG_ATOM, LLT_GC_MARK_SIMPLE_PAIR,
        LLT_GC_MARK_SIMPLE_VECTOR
  };
  enum llt_GC_Mark_recursion {
    LLT_GC_MARK_RECURSIVE_PA, LLT_GC_MARK_RECURSIVE_PL, LLT_GC_MARK_RECURSIVE_PP,
        LLT_GC_MARK_RECURSIVE_PV, LLT_GC_MARK_RECURSIVE_PLL, LLT_GC_MARK_RECURSIVE_VA,
        LLT_GC_MARK_RECURSIVE_VL, LLT_GC_MARK_RECURSIVE_VP, LLT_GC_MARK_RECURSIVE_VV,
        LLT_GC_MARK_RECURSIVE_VLL, LLT_GC_MARK_RECURSIVE_LL, LLT_GC_MARK_RECURSIVE_LLL,
        LLT_GC_MARK_RECURSIVE_PPA, LLT_GC_MARK_RECURSIVE_PPL, LLT_GC_MARK_RECURSIVE_PPP,
        LLT_GC_MARK_RECURSIVE_PPV, LLT_GC_MARK_RECURSIVE_PVA, LLT_GC_MARK_RECURSIVE_PVL,
        LLT_GC_MARK_RECURSIVE_PVP, LLT_GC_MARK_RECURSIVE_PVV, LLT_GC_MARK_RECURSIVE_VPA,
        LLT_GC_MARK_RECURSIVE_VPL, LLT_GC_MARK_RECURSIVE_VPP, LLT_GC_MARK_RECURSIVE_VPV,
        LLT_GC_MARK_RECURSIVE_VVA, LLT_GC_MARK_RECURSIVE_VVL, LLT_GC_MARK_RECURSIVE_VVP,
        LLT_GC_MARK_RECURSIVE_VVV
  };
  struct llt_Fixture {
    LLT_FIXTURE_HEADER;
    cell safe;
    char *copy;
    size_t len;
    boolean proper_pair_p;
    enum llt_GC_Mark_recursion complex;
    enum llt_GC_Mark_flat simplex;
  };
  ⟨ Unit test body 238 ⟩
  ⟨ Unit test: garbage collector mark 276 ⟩
```

$\textbf{llt\_fixture}$ $\mathit{Test\_Fixtures}[\,] = \{$
$\mathit{llt\_GC\_Mark\_\_Global}$, $\mathit{llt\_GC\_Mark\_\_Atom}$, $\mathit{llt\_GC\_Mark\_\_Long\_Atom}$, $\mathit{llt\_GC\_Mark\_\_Pair}$,
$\mathit{llt\_GC\_Mark\_\_Vector}$, $\mathit{llt\_GC\_Mark\_\_Recursive\_P}$, $\mathit{llt\_GC\_Mark\_\_Recursive\_V}$,
$\mathit{llt\_GC\_Mark\_\_Recursive\_L}$, $\mathit{llt\_GC\_Mark\_\_Recursive\_PP}$, $\mathit{llt\_GC\_Mark\_\_Recursive\_PV}$,
$\mathit{llt\_GC\_Mark\_\_Recursive\_VP}$, $\mathit{llt\_GC\_Mark\_\_Recursive\_VV}$, Λ
};

276. These tests work by serialising the object under test into a buffer before and after performing the test to check for changes, and recursively walking the data structure using C's stack to look for the mark flag.

⟨ Unit test: garbage collector *mark* 276 ⟩ ≡

```
boolean llt_GC_Mark_is_marked_p(cell c)
{
  return special_p(c) ∨ (mark_p(c)
      ∧ (¬acar_p(c) ∨ llt_GC_Mark_is_marked_p(car(c)))
      ∧ (¬acdr_p(c) ∨ llt_GC_Mark_is_marked_p(cdr(c))));
}
```

See also sections 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 301, 302, 303, 304, 305, 306, and 307.

This code is used in section 275.

277. Of course after the mark phase of garbage collection live objects *have* been changed because that's the whole point so serialising the post-mark object as-is wouldn't work. Instead the flag is (recursively) lowered first reverting the only change that *mark* should have made.

⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
void llt_GC_Mark_unmark_m(cell c)
{
  int i;
  if (special_p(c)) return;
  mark_clear(c);
  if (acar_p(c)) llt_GC_Mark_unmark_m(car(c));
  if (acdr_p(c)) llt_GC_Mark_unmark_m(cdr(c));
  if (vector_p(c))
    for (i = 0; i < vector_length(c); i++)
      llt_GC_Mark_unmark_m(vector_ref(c, i));
}
```

278. Objects need to be created in various combinations to create the recursive structures to test.

#define *llt_GC_Mark_mkatom* *sym*

⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
cell llt_GC_Mark_mklong(int x, int y)
{
  cell r;
  vms_push(int_new(y));
  r = int_new(x);
  cdr(r) = vms_pop();
  return r;
}
```

279. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
cell llt_GC_Mark_mklonglong(int x, int y, int z)
{
  cell r;

  vms_push(int_new(z));
  r = int_new(y);
  cdr(r) = vms_pop();
  vms_push(r);
  r = int_new(x);
  cdr(r) = vms_pop();
  return r;
}
```

280. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
cell llt_GC_Mark_mkpair(boolean proper_p)
{
  cell r = cons(VOID, UNDEFINED);

  if (proper_p) cdr(r) = NIL;
  return r;
}
```

281. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
cell llt_GC_Mark_mkvector(void)
{
  cell r;
  int i, j;

  r = vector_new_imp(abs(UNDEFINED), 0, NIL);
  for (i = 0, j = −1; j ≥ UNDEFINED; i++, j−−) vector_ref(r, i) = j;
  return r;
}
```

282. Preparing and running the tests. This is where the object under test (created below) gets serialised.

⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
void llt_GC_Mark_prepare(llt_Fixture *fix)
{
  fix→len = object_sizeof(fix→safe);
  ERR_OOM_P(fix→copy = malloc(fix→len));
  object_copy(fix→safe, fix→copy, btrue);
}
```

283. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
void llt_GC_Mark_destroy(llt_Fixture *fix)
{
  free(fix→copy);
}
```

284. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
void llt_GC_Mark_act(llt_Fixture *fix)
{
  mark(fix→safe);
}
```

285. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
boolean llt_GC_Mark_test(llt_Fixture *fix)
{
  char buf[TEST_BUFSIZE];
  boolean ok;
  ok = tap_ok(llt_GC_Mark_is_marked_p(fix→safe), fpmsgf("the␣object␣is␣fully␣marked"));
  llt_GC_Mark_unmark_m(fix→safe);
  tap_again(ok, object_compare(fix→copy, fix→len, fix→safe, btrue), fpmsgf("the␣object␣is␣unchanged"));
  return ok;
}
```

286. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark_fix(llt_Fixture *fix, const char *name)
{
  fix→name = name;
  fix→prepare = llt_GC_Mark_prepare;
  fix→destroy = llt_GC_Mark_destroy;
  fix→act = llt_GC_Mark_act;
  fix→test = llt_GC_Mark_test;
  fix→safe = NIL;
  return fix;
}
```

287. This defines 6 test cases, one for each global object, which need no further preparation.

⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
#define mkfix(n, o) do
  {
    llt_GC_Mark_fix(f + (n), __func__);
    f[(n)].suffix = #o;
    f[(n)].safe = (o);
  }
  while (0)
  llt_Fixture *llt_GC_Mark__Global(void)
  {
    llt_Fixture *f = llt_alloc(6);
    mkfix(0, NIL);
    mkfix(1, FALSE);
    mkfix(2, TRUE);
    mkfix(3, END_OF_FILE);
    mkfix(4, VOID);
    mkfix(5, UNDEFINED);
    return f;
  }
#undef mkfix
```

288. Four test cases test each of the other object types without triggering recursion.

⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
void llt_GC_Mark__PLAV_prepare(llt_Fixture *fix)
{
  switch (fix→simplex) {
  case LLT_GC_MARK_SIMPLE_ATOM:
    fix→safe = llt_GC_Mark_mkatom("forty-two");
    break;
  case LLT_GC_MARK_SIMPLE_LONG_ATOM:
    fix→safe = int_new(42);    /* nb. doesn't use mklong */
    break;
  case LLT_GC_MARK_SIMPLE_PAIR:
    fix→safe = llt_GC_Mark_mkpair(fix→proper_pair_p);
    break;
  case LLT_GC_MARK_SIMPLE_VECTOR:
    fix→safe = llt_GC_Mark_mkvector();
    break;
  }
  llt_GC_Mark_prepare(fix);
}
```

289. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Atom(void)
{
  llt_Fixture *f = llt_alloc(1);
  llt_GC_Mark_fix(f, __func__);
  f→simplex = LLT_GC_MARK_SIMPLE_ATOM;
  f→prepare = llt_GC_Mark__PLAV_prepare;
  return f;
}
```

290. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Long_Atom(void)
{
  llt_Fixture *f = llt_alloc(1);
  llt_GC_Mark_fix(f, __func__);
  f→simplex = LLT_GC_MARK_SIMPLE_LONG_ATOM;
  f→prepare = llt_GC_Mark__PLAV_prepare;
  return f;
}
```

291. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Pair(void)
{
  llt_Fixture *f = llt_alloc(2);
  llt_GC_Mark_fix(f + 0, __func__);
  llt_GC_Mark_fix(f + 1, __func__);
  f[0].simplex = f[1].simplex = LLT_GC_MARK_SIMPLE_PAIR;
  f[0].prepare = f[1].prepare = llt_GC_Mark__PLAV_prepare;
  f[0].proper_pair_p = btrue;
  return f;
}
```

292. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Vector(void)
{
  llt_Fixture *f = llt_alloc(1);
  llt_GC_Mark_fix(f, __func__);
  f→simplex = LLT_GC_MARK_SIMPLE_VECTOR;
  f→prepare = llt_GC_Mark__PLAV_prepare;
  return f;
}
```

293. Preparing the recursive test cases involves a lot of repetetive and methodical code.

⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
void llt_GC_Mark__Recursive_prepare_imp(llt_Fixture *fix, enum llt_GC_Mark_recursion c)
{
  switch (c) {
    ⟨ Unit test part: prepare plain pairs 294 ⟩
    ⟨ Unit test part: prepare plain vectors 295 ⟩
    ⟨ Unit test part: prepare atomic lists 296 ⟩
    ⟨ Unit test part: prepare pairs in pairs 297 ⟩
    ⟨ Unit test part: prepare vectors in pairs 298 ⟩
    ⟨ Unit test part: prepare pairs in vectors 299 ⟩
    ⟨ Unit test part: prepare vectors in vectors 300 ⟩
  }
}
void llt_GC_Mark__Recursive_prepare(llt_Fixture *fix)
{
  llt_GC_Mark__Recursive_prepare_imp(fix, fix→complex);
  Tmp_Test = NIL;
  llt_GC_Mark_prepare(fix);
}
```

294. ⟨ Unit test part: prepare plain pairs 294 ⟩ ≡
case `LLT_GC_MARK_RECURSIVE_PA`: *fix→safe* = *llt_GC_Mark_mkpair*(*bfalse*);
  car(*fix→safe*) = *llt_GC_Mark_mkatom*(`"forty-two"`);
  cdr(*fix→safe*) = *llt_GC_Mark_mkatom*(`"twoty-four"`);
  break;
case `LLT_GC_MARK_RECURSIVE_PL`: *fix→safe* = *llt_GC_Mark_mkpair*(*bfalse*);
  car(*fix→safe*) = *llt_GC_Mark_mklong*(2048, 42);
  cdr(*fix→safe*) = *llt_GC_Mark_mklong*(8042, 24);
  break;
case `LLT_GC_MARK_RECURSIVE_PP`: *fix→safe* = *llt_GC_Mark_mkpair*(*bfalse*);
  car(*fix→safe*) = *llt_GC_Mark_mkpair*(*btrue*);
  cdr(*fix→safe*) = *llt_GC_Mark_mkpair*(*bfalse*);
  break;
case `LLT_GC_MARK_RECURSIVE_PV`: *fix→safe* = *llt_GC_Mark_mkpair*(*bfalse*);
  car(*fix→safe*) = *llt_GC_Mark_mkvector*( );
  cdr(*fix→safe*) = *llt_GC_Mark_mkvector*( );
  break;
case `LLT_GC_MARK_RECURSIVE_PLL`: *fix→safe* = *llt_GC_Mark_mkpair*(*bfalse*);
  car(*fix→safe*) = *llt_GC_Mark_mklonglong*(1024, 2048, 42);
  cdr(*fix→safe*) = *llt_GC_Mark_mklonglong*(4201, 4820, 24);
  break;

This code is used in section 293.

295. ⟨ Unit test part: prepare plain vectors 295 ⟩ ≡
case `LLT_GC_MARK_RECURSIVE_VA`: *fix→safe* = *llt_GC_Mark_mkvector*( );
  vector_ref(*fix→safe*, 4) = *llt_GC_Mark_mkatom*(`"42"`);
  vector_ref(*fix→safe*, 2) = *llt_GC_Mark_mkatom*(`"24"`);
  break;
case `LLT_GC_MARK_RECURSIVE_VL`: *fix→safe* = *llt_GC_Mark_mkvector*( );
  vector_ref(*fix→safe*, 4) = *llt_GC_Mark_mklong*(2048, 42);
  vector_ref(*fix→safe*, 2) = *llt_GC_Mark_mklong*(8042, 24);
  break;
case `LLT_GC_MARK_RECURSIVE_VP`: *fix→safe* = *llt_GC_Mark_mkvector*( );
  vector_ref(*fix→safe*, 4) = *llt_GC_Mark_mkpair*(*btrue*);
  vector_ref(*fix→safe*, 2) = *llt_GC_Mark_mkpair*(*bfalse*);
  break;
case `LLT_GC_MARK_RECURSIVE_VV`: *fix→safe* = *llt_GC_Mark_mkvector*( );
  vector_ref(*fix→safe*, 4) = *llt_GC_Mark_mkvector*( );
  vector_ref(*fix→safe*, 2) = *llt_GC_Mark_mkvector*( );
  break;
case `LLT_GC_MARK_RECURSIVE_VLL`: *fix→safe* = *llt_GC_Mark_mkvector*( );
  vector_ref(*fix→safe*, 4) = *llt_GC_Mark_mklonglong*(1024, 2048, 42);
  vector_ref(*fix→safe*, 2) = *llt_GC_Mark_mklonglong*(4201, 4820, 24);
  break;

This code is used in section 293.

296. ⟨ Unit test part: prepare atomic lists 296 ⟩ ≡
case `LLT_GC_MARK_RECURSIVE_LL`: *fix→safe* = *llt_GC_Mark_mklong*(1024, 42);
  break;
case `LLT_GC_MARK_RECURSIVE_LLL`: *fix→safe* = *llt_GC_Mark_mklonglong*(1024, 2048, 42);
  break;

This code is used in section 293.

297. ⟨ Unit test part: prepare pairs in pairs 297 ⟩ ≡

```
case LLT_GC_MARK_RECURSIVE_PPA:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PA);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PA);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_PPL:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PL);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PL);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_PPP:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PP);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PP);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_PPV:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PV);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PV);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
```

This code is used in section 293.

298. ⟨ Unit test part: prepare vectors in pairs 298 ⟩ ≡

```
case LLT_GC_MARK_RECURSIVE_PVA:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VA);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VA);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_PVL:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VL);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VL);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_PVP:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VP);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VP);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_PVV:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VV);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VV);
  fix→safe = cons(fix→safe, Tmp_Test);
  break;
```

This code is used in section 293.

299. ⟨ Unit test part: prepare pairs in vectors 299 ⟩ ≡

```
case LLT_GC_MARK_RECURSIVE_VPA:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PA);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PA);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_VPL:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PL);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PL);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_VPP:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PP);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PP);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_VPV:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PV);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_PV);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
```

This code is used in section 293.

300. ⟨ Unit test part: prepare vectors in vectors 300 ⟩ ≡

```
case LLT_GC_MARK_RECURSIVE_VVA:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VA);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VA);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_VVL:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VL);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VL);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_VVP:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VP);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VP);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
case LLT_GC_MARK_RECURSIVE_VVV:
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VV);
  Tmp_Test = fix→safe;
  llt_GC_Mark__Recursive_prepare_imp(fix, LLT_GC_MARK_RECURSIVE_VV);
  Tmp_Test = cons(fix→safe, Tmp_Test);
  fix→safe = llt_GC_Mark_mkvector();
  vector_ref(fix→safe, 4) = car(Tmp_Test);
  vector_ref(fix→safe, 2) = cdr(Tmp_Test);
  break;
```

This code is used in section 293.

301. **#define** *llt_GC_Mark_recfix*(*f*, *n*, *c*) **do**

```
{
  llt_GC_Mark_fix(f + (n), __func__);
  f[(n)].prepare = llt_GC_Mark__Recursive_prepare;
  f[(n)].complex = (c);
  f[(n)].suffix = #c;
}
while (0)
```

⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Recursive_P(void)
{
  llt_Fixture *f = llt_alloc(5);
  llt_GC_Mark_recfix(f, 0, LLT_GC_MARK_RECURSIVE_PA);
  llt_GC_Mark_recfix(f, 1, LLT_GC_MARK_RECURSIVE_PL);
  llt_GC_Mark_recfix(f, 2, LLT_GC_MARK_RECURSIVE_PP);
  llt_GC_Mark_recfix(f, 3, LLT_GC_MARK_RECURSIVE_PV);
  llt_GC_Mark_recfix(f, 4, LLT_GC_MARK_RECURSIVE_PLL);
  return f;
}
```

302. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Recursive_V(void)
{
  llt_Fixture *f = llt_alloc(5);
  llt_GC_Mark_recfix(f, 0, LLT_GC_MARK_RECURSIVE_VA);
  llt_GC_Mark_recfix(f, 1, LLT_GC_MARK_RECURSIVE_VL);
  llt_GC_Mark_recfix(f, 2, LLT_GC_MARK_RECURSIVE_VP);
  llt_GC_Mark_recfix(f, 3, LLT_GC_MARK_RECURSIVE_VV);
  llt_GC_Mark_recfix(f, 4, LLT_GC_MARK_RECURSIVE_VLL);
  return f;
}
```

303. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Recursive_L(void)
{
  llt_Fixture *f = llt_alloc(2);
  llt_GC_Mark_recfix(f, 0, LLT_GC_MARK_RECURSIVE_LL);
  llt_GC_Mark_recfix(f, 1, LLT_GC_MARK_RECURSIVE_LLL);
  return f;
}
```

304. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡

```
llt_Fixture *llt_GC_Mark__Recursive_PP(void)
{
  llt_Fixture *f = llt_alloc(4);
  llt_GC_Mark_recfix(f, 0, LLT_GC_MARK_RECURSIVE_PPA);
  llt_GC_Mark_recfix(f, 1, LLT_GC_MARK_RECURSIVE_PPL);
  llt_GC_Mark_recfix(f, 2, LLT_GC_MARK_RECURSIVE_PPP);
  llt_GC_Mark_recfix(f, 3, LLT_GC_MARK_RECURSIVE_PPV);
  return f;
}
```

305. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡
llt_Fixture **llt_GC_Mark__Recursive_PV*(**void**)
{
  llt_Fixture **f* = *llt_alloc*(4);
  llt_GC_Mark_recfix(*f*, 0, LLT_GC_MARK_RECURSIVE_PVA);
  llt_GC_Mark_recfix(*f*, 1, LLT_GC_MARK_RECURSIVE_PVL);
  llt_GC_Mark_recfix(*f*, 2, LLT_GC_MARK_RECURSIVE_PVP);
  llt_GC_Mark_recfix(*f*, 3, LLT_GC_MARK_RECURSIVE_PVV);
  return *f*;
}

306. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡
llt_Fixture **llt_GC_Mark__Recursive_VP*(**void**)
{
  llt_Fixture **f* = *llt_alloc*(4);
  llt_GC_Mark_recfix(*f*, 0, LLT_GC_MARK_RECURSIVE_VPA);
  llt_GC_Mark_recfix(*f*, 1, LLT_GC_MARK_RECURSIVE_VPL);
  llt_GC_Mark_recfix(*f*, 2, LLT_GC_MARK_RECURSIVE_VPP);
  llt_GC_Mark_recfix(*f*, 3, LLT_GC_MARK_RECURSIVE_VPV);
  return *f*;
}

307. ⟨ Unit test: garbage collector *mark* 276 ⟩ +≡
llt_Fixture **llt_GC_Mark__Recursive_VV*(**void**)
{
  llt_Fixture **f* = *llt_alloc*(4);
  llt_GC_Mark_recfix(*f*, 0, LLT_GC_MARK_RECURSIVE_VVA);
  llt_GC_Mark_recfix(*f*, 1, LLT_GC_MARK_RECURSIVE_VVL);
  llt_GC_Mark_recfix(*f*, 2, LLT_GC_MARK_RECURSIVE_VVP);
  llt_GC_Mark_recfix(*f*, 3, LLT_GC_MARK_RECURSIVE_VVV);
  return *f*;
}

308. Sweep.

1. *What is the contract fulfilled by the code under test?*

All cells which are marked will become unmarked and otherwise unchanged. All other cells will be on the free list (in an insignificant order). The size of the free list will be returned.

2. *What preconditions are required, and how are they enforced?*

The pool need not have been initialised in which case the free list and return value are `NIL` and 0 respectively. Live objects should be already marked for which we define *llt_GC_Sweep_mark_m* which also counts the size of the object being marked.

3. *What postconditions are guaranteed?*

The *car* content of a cell put into the free list is unchanged but this doesn't matter.

4. *What example inputs trigger different behaviors?*

Exactly 2: The size of the pool and the set of marked cells.

5. *What set of tests will trigger each behavior and validate each guarantee?*

There are three tests. The simplest is to verify that *sweep* is effectively a no-op when there is no pool. The other two both prepare a dud object which should be returned to the free list and test whether *sweep* works correctly both with and without a live object.

```
⟨ t/gc-sweep.c 308 ⟩ ≡
  ⟨ Unit test header 237 ⟩
  struct llt_Fixture {
    LLT_FIXTURE_HEADER;
    boolean preinit_p;
    cell safe;
    cell *safe_buf;
    size_t expect;
    int ret_val;
  };
  ⟨ Unit test body 238 ⟩
  ⟨ Unit test: garbage collector sweep 309 ⟩
  llt_fixture Test_Fixtures[] = {
      llt_GC_Sweep__Empty_Pool, llt_GC_Sweep__Used_Pool, Λ
  };
```

309. ⟨ Unit test: garbage collector *sweep* 309 ⟩ ≡

```
  size_t llt_GC_Sweep_mark_m(cell c)
  {
    int i;
    size_t count = 0;
    if (special_p(c)) return 0;
    mark_set(c);
    count++;
    if (acar_p(c)) count += llt_GC_Sweep_mark_m(car(c));
    if (acdr_p(c)) count += llt_GC_Sweep_mark_m(cdr(c));
    if (vector_p(c))
      for (i = 0; i < vector_length(c); i++)
        count += llt_GC_Sweep_mark_m(vector_ref(c, i));
    return count;
  }
```

See also sections 310, 311, 312, 313, 314, 315, 316, and 317.

This code is used in section 308.

310. To test *sweep* when there is no pool there's no need to actually remove the pool. In other cases a few cells are consumed from the free list and ignored.

⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡
```
void llt_GC_Sweep_prepare(llt_Fixture *fix)
{
  if (fix→preinit_p) {
    Cells_Poolsize = 0;
    Cells_Free = UNDEFINED;
    return;
  }
  vms_push(cons(NIL, NIL));
  cons(NIL, vms_pop());
}
```

311. The VM is fully reset after every test.

⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡
```
void llt_GC_Sweep_destroy(llt_Fixture *fix __unused)
{
  free(fix→safe_buf);
  vm_init_imp();
}
```

312. ⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡
```
void llt_GC_Sweep_act(llt_Fixture *fix)
{
  fix→ret_val = sweep();
}
```

313. ⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡

```
boolean llt_GC_Sweep_test(llt_Fixture *fix)
{
  char buf[TEST_BUFSIZE];
  cell f;
  boolean ok, mark_ok_p, free_ok_p;
  int i, rem;
  rem = Cells_Poolsize − fix→expect;
  ok = tap_ok(fix→ret_val ≡ rem, fpmsgf("sweep␣returns␣the␣number␣of␣free␣cells␣(%d)", rem));
  i = 0;
  for (f = Cells_Free; ¬null_p(f); f = cdr(f))
    i++;
  tap_more(ok, i ≡ rem, fpmsgf("the␣number␣of␣free␣cells␣is␣correct␣(%d)", rem));
  mark_ok_p = btrue;
  for (i = 0; i < (int) fix→expect; i++)
    if (mark_p(fix→safe_buf[i]))
      mark_ok_p = bfalse;
  tap_more(ok, mark_ok_p, fpmsgf("the␣cells␣are␣unmarked"));
  free_ok_p = btrue;
  for (f = Cells_Free; ¬null_p(f); f = cdr(f))
    for (i = 0; i < (int) fix→expect; i++)
      if (fix→safe_buf[i] ≡ f)
        free_ok_p = bfalse;
  tap_more(ok, mark_ok_p, fpmsgf("the␣used␣cells␣are␣not␣in␣the␣free␣list"));
  return ok;
}
```

314. ⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡

```
llt_Fixture *llt_GC_Sweep_fix(llt_Fixture *fix, const char *name)
{
  fix→name = name;
  fix→prepare = llt_GC_Sweep_prepare;
  fix→destroy = llt_GC_Sweep_destroy;
  fix→act = llt_GC_Sweep_act;
  fix→test = llt_GC_Sweep_test;
  return fix;
}
```

315. ⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡

```
llt_Fixture *llt_GC_Sweep__Empty_Pool(void)
{
  llt_Fixture *f = llt_alloc(2);
  llt_GC_Sweep_fix(f + 0, __func__);
  llt_GC_Sweep_fix(f + 1, __func__);
  f[0].preinit_p = btrue;
  f[0].suffix = "no␣pool";
  f[1].suffix = "unused";
  return f;
}
```

316. References to the cells which make up the object are saved in *fix→safe_buf* to check that they were not put on the free list.

⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡

```
void llt_GC_Sweep__Used_Pool_prepare(llt_Fixture *fix)
{
  fix→safe = cons(VOID, UNDEFINED);
  vms_push(fix→safe);
  fix→expect = llt_GC_Sweep_mark_m(vms_ref());
  ERR_OOM_P(fix→safe_buf = malloc(fix→expect));
  object_copyref(fix→safe, fix→safe_buf);
  llt_GC_Sweep_prepare(fix);
  vms_pop();
}
```

317. ⟨ Unit test: garbage collector *sweep* 309 ⟩ +≡

```
llt_Fixture *llt_GC_Sweep__Used_Pool(void)
{
  llt_Fixture *f = llt_alloc(1);
  llt_GC_Sweep_fix(f + 0, __func__);
  f[0].prepare = llt_GC_Sweep__Used_Pool_prepare;
  return f;
}
```

318. Vectors.

1. *What is the contract fulfilled by the code under test?*

vector objects which are not live (pointed at by something in `ROOTS`) will have their *tag* changed to `TAG_NONE` and their cdr née offset changed to a pointer in the free list, as will all their contents.

Live *vectors* cell pointer, length and contents are unchanged. The offset will be reduced by the (full) size of any unused *vectors* prior to it in `VECTOR`.

The number of free cells in `VECTOR` is returned.

2. *What preconditions are required, and how are they enforced?*

Used vectors must be pointed to from something in `ROOTS`. They will be pushed into `VMS`.

The linear nature of *vector_new* is taken advantage of to create the holes in the `VECTOR` buffer that *gc_vectors* must defragment.

All other aspects of the garbage collector are assumed to work correctly.

3. *What postconditions are guaranteed?*

The VM is fully reset after each test so that they begin with `VECTOR` in a clean state.

4. *What example inputs trigger different behaviors?*

The only things to affect the way *vector* garbage collection works is whether or not each vector is live and where they exist in memory in relation to one another, ie. whether unused vectors will leave holes in `VECTOR` after collection.

5. *What set of tests will trigger each behavior and validate each guarantee?*

The `VECTOR` buffer will be packed with live/unused objects in various arrangements.

```
#define LLT_GC_VECTOR__SIZE "2718281828459"
#define LLT_GC_VECTOR__SHAPE "GNS"
```

```
⟨ t/gc-vector.c 318 ⟩ ≡
  ⟨ Unit test header 237 ⟩
  struct llt_Fixture {
    LLT_FIXTURE_HEADER;
    const char *pattern;
    int ret_val;
    size_t safe_bufsize;
    size_t safe_size;
    cell *cell_buf;
    cell *offset_buf;
    char *safe_buf;
    size_t *size_buf;
    size_t unsafe_bufsize;
    cell *unsafe_buf;
  };
  ⟨ Unit test body 238 ⟩
  ⟨ Unit test: garbage collector gc_vector 319 ⟩
  llt_fixture Test_Fixtures[] = {
      llt_GC_Vector__All, Λ
  };
```

319. These tests are highly repetetive so the *vector*s defined by the fixture are created programmatically according to the pattern in *fix→pattern* which is a simple language of `L` & `U` characters.

Each vector is created by taking a character from the pattern and `LLT_GC_VECTOR__SIZE` in turn to decide on the size of the vector and whether it is live or unused. `LLT_GC_VECTOR__SHAPE` is then cycled through to populate each *vector* with a variety of data.

Live *vector*s are pushed onto `VMS` to keep them safe from collection. Vectors which will be considered unused are pushed onto `CTS` to keep them safe from collection while the fixture is being prepared.

```
⟨ Unit test: garbage collector gc_vector 319 ⟩ ≡
    /* There are too many one-letter variables in this function which then get reused */
  void llt_GC_Vector_prepare(llt_Fixture *fix)
  {
    cell g, v;
    char buf[TEST_BUFSIZE], *p, *s, *t;
    int i, n, z;
    if (¬fix→pattern) fix→pattern = "L";
    g = NIL;
    n = SCHAR_MAX;
    s = LLT_GC_VECTOR__SIZE;
    t = LLT_GC_VECTOR__SHAPE;
    for (p = (char *) fix→pattern; *p; p++) {
      if (*s ≡ '\0') s = LLT_GC_VECTOR__SIZE;
      ⟨ Unit test part: build a “random” vector 320 ⟩
      if (*p ≡ 'L') {
        ⟨ Unit test part: serialise a live vector into the fixture 321 ⟩
        vms_push(v);
      } else
        cts_push(v);
    }
    ⟨ Unit test part: complete live vector serialisation 322 ⟩
    ⟨ Unit test part: save unused vector references 323 ⟩
    cts_reset();
  }
```

See also sections 324, 327, 328, 329, and 330.

This code is used in section 318.

320. Each time a global variable is requested g is decremented, cycling from `NIL` down to `UNDEFINED`. Each new number and symbol is also unique using a counter n that starts high enough to create numbers not protected by *Small_Int*.

⟨ Unit test part: build a "random" *vector* 320 ⟩ ≡
```
v = vector_new((z = *s++ − '0'), NIL);
for (i = 0; i < z; i++) {
  if (*t ≡ '\0') t = LLT_GC_VECTOR__SHAPE;
  switch (*t++) {
  case 'G':
    vector_ref(v, i) = g−−;
    if (g < UNDEFINED) g = NIL;
    break;
  case 'N':
    vector_ref(v, i) = int_new(n += 42);
    break;
  case 'S':
    snprintf(buf, TEST_BUFSIZE, "testing-%d", n += 42);
    vector_ref(v, i) = sym(buf);
    break;
  }
}
```
This code is used in section 319.

321. The offset of a *vector* may change if there are unused *vector*s to collect so it's saved into *fix→offset_buf* instead and the live *vector*s are serialised without recording it.

⟨ Unit test part: serialise a live *vector* into the fixture 321 ⟩ ≡
```
fix→safe_size++;
fix→cell_buf = reallocarray(fix→cell_buf, fix→safe_size, sizeof(cell));
fix→offset_buf = reallocarray(fix→offset_buf, fix→safe_size, sizeof(cell));
fix→size_buf = reallocarray(fix→size_buf, fix→safe_size, sizeof(size_t));
fix→cell_buf[fix→safe_size − 1] = v;
fix→offset_buf[fix→safe_size − 1] = vector_offset(v);
fix→safe_bufsize += fix→size_buf[fix→safe_size − 1] = object_sizeof(v);
```
This code is used in section 319.

322. The list of live objects saved in `VMS` is reversed so that the order matches that in *fix→pattern* then they are serialised sequentially into *fix→safe_buf*.

⟨ Unit test part: complete live *vector* serialisation 322 ⟩ ≡
```
fix→safe_buf = calloc(fix→safe_bufsize, sizeof(char));
VMS = list_reverse_m(VMS, btrue);
n = 0;
for (v = VMS; ¬null_p(v); v = cdr(v))
  n += object_copy(car(v), fix→safe_buf + n, bfalse);
```
This code is used in section 319.

323. Unused objects don't need to be serialised; their cell references only are saved to verify that they have been returned to the free list.

⟨ Unit test part: save unused *vector* references 323 ⟩ ≡

```
fix→unsafe_bufsize = 0;
for (v = CTS; ¬null_p(v); v = cdr(v))
  fix→unsafe_bufsize += object_sizeofref(car(v));
fix→unsafe_buf = calloc(fix→unsafe_bufsize, sizeof(cell));
i = 0;
for (v = CTS; ¬null_p(v); v = cdr(v))
  i += object_copyref(car(v), fix→unsafe_buf + i);
```

This code is used in section 319.

324. ⟨ Unit test: garbage collector *gc_vector* 319 ⟩ +≡

```
boolean llt_GC_Vector_test(llt_Fixture *fix)
{
  char buf[TEST_BUFSIZE], *p, *s;
  boolean ok, liveok, freeok, tagok, *freelist;
  int delta, live, serial, unused, f, i;
  cell j;

  freelist = calloc(Cells_Poolsize, sizeof(boolean));
  for (j = Cells_Free; ¬null_p(j); j = cdr(j)) freelist[j] = btrue;
  delta = live = serial = unused = 0;
  s = LLT_GC_VECTOR__SIZE;
  ok = btrue;
  for (i = 0, p = (char *) fix→pattern; *p; i++, p++) {
    if (*s ≡ '\0') s = LLT_GC_VECTOR__SIZE;
    if (*p ≡ 'L') { ⟨ Unit test part: test a live vector 325 ⟩ }
    else { ⟨ Unit test part: test an unused vector 326 ⟩ }
    s++;
  }
  return ok;
}
```

325. ⟨ Unit test part: test a live *vector* 325 ⟩ ≡

```
liveok = object_compare(fix→safe_buf + serial, fix→size_buf[live], fix→cell_buf[live], bfalse);
tap_more(ok, liveok, fpmsgf("(L-%d)␣object␣is␣unchanged", live));
liveok = vector_offset(fix→cell_buf[live]) ≡ fix→offset_buf[live] − delta;
tap_more(ok, liveok, fpmsgf("(L-%d),␣object␣is␣defragmented", live));
serial += fix→size_buf[live];
live++;
```

This code is used in section 324.

326. ⟨ Unit test part: test an unused *vector* 326 ⟩ ≡

```
f = *s − '0';
delta += f ? vector_realsize(f) : 0;
tagok = freeok = btrue;
for (i = 0; i < (int) fix→unsafe_bufsize; i++) {
  j = fix→unsafe_buf[i];
  if (special_p(j) ∨ symbol_p(j) ∨ smallint_p(j)) continue;
  tagok = (tag(j) ≡ TAG_NONE) ∧ tagok;
  freeok = freelist[i] ∧ freeok;
}
tap_more(ok, tagok, fpmsgf("(U-%d)␣object's␣tag␣is␣cleared", unused));
tap_more(ok, freeok, fpmsgf("(U-%d)␣object␣is␣in␣the␣free␣list", unused));
unused++;
```

This code is used in section 324.

327. ⟨ Unit test: garbage collector *gc_vector* 319 ⟩ +≡

```
void llt_GC_Vector_destroy(llt_Fixture *fix)
{
  free(fix→cell_buf);
  free(fix→offset_buf);
  free(fix→safe_buf);
  free(fix→size_buf);
  free(fix→unsafe_buf);
  vm_init_imp();
}
```

328. ⟨ Unit test: garbage collector *gc_vector* 319 ⟩ +≡

```
void llt_GC_Vector_act(llt_Fixture *fix)
{
  fix→ret_val = gc_vectors();
}
```

329. ⟨ Unit test: garbage collector *gc_vector* 319 ⟩ +≡

```
llt_Fixture *llt_GC_Vector_fix(llt_Fixture *fix, const char *name)
{
  fix→name = name;
  fix→prepare = llt_GC_Vector_prepare;
  fix→destroy = llt_GC_Vector_destroy;
  fix→act = llt_GC_Vector_act;
  fix→test = llt_GC_Vector_test;
  return fix;
}
```

330. The tests themselves are then defined with a list of combinations of `L` & `U` that are built into the fixtures.

⟨ Unit test: garbage collector *gc_vector* 319 ⟩ +≡

```
llt_Fixture *llt_GC_Vector__All(void)
{
    static char *test_patterns[] = {"L", "LL", "LLL", "LLLU", "LLLUUU", "LLUL", "LLUUUL", "LULL",
        "LUULL", "LULUL", "LUULUL", "LUULUUL", "LLLULLLULLL", "LLLUUUULLLUUUULLL", "UL", "ULLL",
        "ULLLU", "UUULLL", "UUULLLUUU", "UUULLLUUULLL", "UUULLLUUULLLUUU", Λ};
    char **p;
    llt_Fixture *f;
    int c, i;
    for (c = 0, p = test_patterns; *p; c++, p++) ;
    f = llt_alloc(c);
    for (i = 0; i < c; i++) {
        llt_GC_Vector_fix(f + i, __func__);
        f[i].suffix = f[i].pattern = test_patterns[i];
    }
    return f;
}
```

331. Objects.

332. Pair Integration. With the basic building blocks' interactions tested we arrive at the critical integration between the compiler and the interpreter.

Calling the following tests integration tests may be thought of as a bit of a misnomer; if so consider them unit tests of the integration tests which are to follow in pure LossLess code.

Starting with *pair*s tests that *cons*, *car*, *cdr*, *null?*, *pair?*, *set-car!* & *set-cdr!* return their result and don't do anything strange. This code is extremely boring and repetetive.

```
⟨ t/pair.c 332 ⟩ ≡
  ⟨ Old test executable wrapper 225 ⟩
  void test_main(void)
  {
    boolean ok, okok;
    cell marco, polo, t, water;      /* t is not saved from destruction */
    char *prefix = Λ;
    char msg[TEST_BUFSIZE] = {0};
    marco = sym("marco?");
    polo = sym("polo!");
    water = sym("fish␣out␣of␣water!");
    ⟨ Test integrating cons 333 ⟩
    ⟨ Test integrating car 334 ⟩
    ⟨ Test integrating cdr 335 ⟩
    ⟨ Test integrating null? 336 ⟩
    ⟨ Test integrating pair? 339 ⟩
    ⟨ Test integrating set-car! 342 ⟩
    ⟨ Test integrating set-cdr! 343 ⟩
  }
```

333. These tests could perhaps be made more thorough but I'm not sure what it would achieve. Testing the non-mutating calls is basically the same: Prepare & interpret code that will call the operator and then test that the result is correct and that internal state is (not) changed as expected.

```
⟨ Test integrating cons 333 ⟩ ≡
  vm_reset();
  Acc = read_cstring(prefix = "(cons␣24␣42)");
  interpret();
  ok = tap_ok(pair_p(Acc), tmsgf("pair?"));
  tap_again(ok, integer_p(car(Acc)) ∧ int_value(car(Acc)) ≡ 24, tmsgf("car"));
  tap_again(ok, integer_p(car(Acc)) ∧ int_value(cdr(Acc)) ≡ 42, tmsgf("cdr"));
  test_vm_state_full(prefix);
```

This code is used in section 332.

334.

```
⟨ Test integrating car 334 ⟩ ≡
  vm_reset();
  t = cons(int_new(42), polo);
  t = cons(synquote_new(t), NIL);
  Tmp_Test = Acc = cons(sym("car"), t);
  prefix = "(car␣'(42␣.␣polo))";
  interpret();
  tap_ok(integer_p(Acc) ∧ int_value(Acc) ≡ 42, tmsgf("integer?"));
  test_vm_state_full(prefix);
```

This code is used in section 332.

335. ⟨ Test integrating cdr 335 ⟩ ≡
vm_reset( );
Acc = *cons*(*sym*("cdr"), *t*);
prefix = "(cdr␣'(42␣.␣polo))";
interpret( );
tap_ok(*symbol_p*(*Acc*) ∧ *Acc* ≡ *polo*, *tmsgf*("symbol?"));
test_vm_state_full(*prefix*);

This code is used in section 332.

336. ⟨ Test integrating null? 336 ⟩ ≡
vm_reset( );
t = *cons*(NIL, NIL);
Acc = *cons*(*sym*("null?"), *t*);
prefix = "(null?␣())";
interpret( );
tap_ok(*true_p*(*Acc*), *tmsgf*("true?"));
test_vm_state_full(*prefix*);

See also sections 337 and 338.

This code is used in section 332.

337. ⟨ Test integrating null? 336 ⟩ +≡
vm_reset( );
t = *cons*(*synquote_new*(*polo*), NIL);
Acc = *cons*(*sym*("null?"), *t*);
prefix = "(null?␣'polo!)";
interpret( );
tap_ok(*false_p*(*Acc*), *tmsgf*("false?"));
test_vm_state_full(*prefix*);

338. ⟨ Test integrating null? 336 ⟩ +≡
vm_reset( );
t = *synquote_new*(*cons*(NIL, NIL));
Acc = *cons*(*sym*("null?"), *cons*(*t*, NIL));
prefix = "(null?␣'(()))";
interpret( );
tap_ok(*false_p*(*Acc*), *tmsgf*("false?"));
test_vm_state_full(*prefix*);

339. ⟨ Test integrating pair? 339 ⟩ ≡
vm_reset( );
Acc = *cons*(*sym*("pair?"), *cons*(NIL, NIL));
prefix = "(pair?␣())";
interpret( );
tap_ok(*false_p*(*Acc*), *tmsgf*("false?"));
test_vm_state_full(*prefix*);

See also sections 340 and 341.

This code is used in section 332.

340. ⟨ Test integrating pair? 339 ⟩ +≡

```
vm_reset();
t = cons(synquote_new(polo), NIL);
Acc = cons(sym("pair?"), t);
prefix = "(pair?␣'polo!)";
interpret();
tap_ok(false_p(Acc), tmsgf("false?"));
test_vm_state_full(prefix);
```

341. ⟨ Test integrating pair? 339 ⟩ +≡

```
vm_reset();
t = synquote_new(cons(NIL, NIL));
Acc = cons(sym("pair?"), cons(t, NIL));
prefix = "(pair?␣'(()))";
interpret();
tap_ok(true_p(Acc), tmsgf("true?"));
test_vm_state_full(prefix);
```

342. Testing that pair mutation works correctly requires some more work. A pair is created and saved in *Tmp_Test* then the code which will be interpreted is created by hand to inject that pair directly and avoid looking for its value in an *environment*.

TODO: duplicate these tests for symbols that are looked up.

⟨ Test integrating set-car! 342 ⟩ ≡

```
vm_reset();
Tmp_Test = cons(marco, water);
t = cons(synquote_new(polo), NIL);
t = cons(synquote_new(Tmp_Test), t);
Acc = cons(sym("set-car!"), t);
prefix = "(set-car!␣'(marco␣.␣|fish␣out␣of␣water!|)␣'polo!)";
interpret();
ok = tap_ok(void_p(Acc), tmsgf("void?"));
okok = tap_ok(ok ∧ pair_p(Tmp_Test), tmsgf("(pair?␣T)"));
tap_again(ok, symbol_p(car(Tmp_Test)) ∧ car(Tmp_Test) ≡ polo, tmsgf("(eq?␣(car␣T)␣'polo!)"));
tap_again(okok, symbol_p(cdr(Tmp_Test)) ∧ cdr(Tmp_Test) ≡ water,
      tmsgf("(eq?␣(cdr␣T)␣'|fish␣out␣of␣water!|)"));
```

This code is used in section 332.

343. ⟨ Test integrating set-cdr! 343 ⟩ ≡

```
vm_reset();
Tmp_Test = cons(water, marco);
t = cons(synquote_new(polo), NIL);
t = cons(synquote_new(Tmp_Test), t);
Acc = cons(sym("set-cdr!"), t);
prefix = "(set-cdr!␣'(|fish␣out␣of␣water!|␣.␣marco)␣'polo!)";
interpret();
ok = tap_ok(void_p(Acc), tmsgf("void?"));
okok = tap_ok(ok ∧ pair_p(Tmp_Test), tmsgf("(pair?␣T)"));
tap_again(ok, symbol_p(car(Tmp_Test)) ∧ car(Tmp_Test) ≡ water,
      tmsgf("(eq?␣(car␣T)␣'|fish␣out␣of␣water!|)"));
tap_again(okok, symbol_p(cdr(Tmp_Test)) ∧ cdr(Tmp_Test) ≡ polo, tmsgf("(eq?␣(cdr␣T)␣'polo!)"));
```

This code is used in section 332.

344. Integrating eval. Although useful to write, and they weeded out some dumb bugs, the real difficulty is in ensuring the correct *environment* is in place at the right time.

We'll skip **error** for now and start with **eval**. Again this test isn't thorough but I think it's good enough for now. The important tests are that the arguments to **eval** are evaluated in the compile-time environment in which the **eval** is located, and that the program which the first argument evaluates to is itself evaluated in the environment the second argument evaluates to.

```
⟨ t/eval.c 344 ⟩ ≡
  ⟨ Old test executable wrapper 225 ⟩
  void test_main(void)
  {
    cell t, m, p;
    char *prefix;
    char msg[TEST_BUFSIZE] = {0};
    ⟨ Test integrating eval 345 ⟩
  }
```

See also section 350.

345. The first test of **eval** calls into it without needing to look up any of its arguments. The program to be evaluated calls *test!probe* and its result is examined. First evaluating in the current environment which is here *Root*.

```
⟨ Test integrating eval 345 ⟩ ≡
  vm_reset();
  Acc = read_cstring((prefix = "(eval␣'(test!probe))"));
  interpret();
  t = assoc_value(Acc, sym("Env"));
  tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣T␣'Env))"));
  tap_ok(t ≡ Root, tmsgf("(eq?␣(assoc-value␣T␣'Env)␣Root)"));
    /* TODO: Is it worth testing that Acc ≡ Prog ≡ [ OP_TEST_PROBE OP_RETURN ]? */
  test_vm_state_full(prefix);
```

See also sections 346, 347, and 348.

This code is used in section 344.

346. And then testing with a second argument of an artificially-constructed environment.

The probing symbol is given a different name to shield against it being found in *Root* and fooling the tests into passing.

```
⟨ Test integrating eval 345 ⟩ +≡
  vm_reset();
  Tmp_Test = env_empty();
  env_set(Tmp_Test, sym("alt-test!probe"), env_search(Root, sym("test!probe")), TRUE);
  Acc = read_cstring((prefix = "(eval␣'(alt-test!probe)␣E)"));
  cddr(Acc) = cons(Tmp_Test, NIL);
  interpret();
  t = assoc_value(Acc, sym("Env"));
  tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣T␣'Env))"));
  tap_ok(t ≡ Tmp_Test, tmsgf("(eq?␣(assoc-value␣T␣'Env)␣E)"));
  test_vm_state_full(prefix);
```

347. Testing that **eval**'s arguments are evaluated in the correct *environment* is a little more difficult. The *environment* with variables to supply **eval**'s arguments is constructed. These are the program source and another artificial *environment* which the program should be evaluated in.

t, *m* & *p* are protected throughout as they are only links to somewhere in the outer *environment* which is protected by *Tmp_Test*.

⟨ Test integrating **eval** 345 ⟩ +≡

```
Tmp_Test = env_empty();    /* outer environment */
env_set(Tmp_Test, sym("eval"), env_search(Root, sym("eval")), TRUE);
env_set(Tmp_Test, sym("alt-test!probe"), env_search(Root, sym("error")), TRUE);
t = read_cstring("(alt-test!probe␣'oops)");    /* program; oops in case we end up in error */
env_set(Tmp_Test, sym("testing-program"), t, TRUE);
m = env_empty();    /* evaluation environmant */
env_set(Tmp_Test, sym("testing-environment"), m, TRUE);
env_set(m, sym("alt-test!probe"), env_search(Root, sym("test!probe")), TRUE);
env_set(m, sym("testing-environment"), env_empty(), TRUE);
p = read_cstring("(error␣wrong-program)");
env_set(m, sym("testing-program"), p, TRUE);
```

348. **eval** is then called in the newly-constructed *environment* by putting it in *Env* before calling *interpret*, mimicking what *frame_push* would do when entering the closure the *environment* represents.

⟨ Test integrating **eval** 345 ⟩ +≡

```
vm_reset();
prefix = "(eval␣testing-program␣testing-environment)";
Acc = read_cstring(prefix);
Env = Tmp_Test;
interpret();
t = assoc_value(Acc, sym("Env"));
tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣T␣'Env))"));
tap_ok(t ≡ m, tmsgf("(eq?␣(assoc-value␣T␣'Env)␣E)"));
test_integrate_eval_unchanged(prefix, Tmp_Test, m);
test_vm_state_normal(prefix);
tap_ok(Env ≡ Tmp_Test, tmsgf("(unchanged?␣Env)"));
```

349. Neither of the two environments should be changed at all. That is *inner* should have exactly `alt-test!probe`, `testing-environment` & `testing-program`, *outer* should have the same symbols with the different values as above and also **eval**.

⟨ Function declarations 8 ⟩ +≡

```
#define TEST_EVAL_FOUND(var)
  if (undefined_p(var)) (var) = cadar(t);
  else fmore = btrue;
#define TEST_EVAL_FIND
  feval = fprobe = fenv = fprog = UNDEFINED;
  fmore = bfalse;
  while (¬null_p(t)) {
    if (caar(t) ≡ sym("alt-test!probe")) { TEST_EVAL_FOUND(fprobe); }
    else if (caar(t) ≡ sym("eval")) { TEST_EVAL_FOUND(feval); }
    else if (caar(t) ≡ sym("testing-environment")) { TEST_EVAL_FOUND(fenv); }
    else if (caar(t) ≡ sym("testing-program")) { TEST_EVAL_FOUND(fprog); }
    else fmore = btrue;
    t = cdr(t);
  }
  void test_integrate_eval_unchanged(char *, cell, cell);
```

350. ⟨ t/eval.c 344 ⟩ +≡

```
  void test_integrate_eval_unchanged(char *prefix, cell outer, cell inner)
  {
    boolean oki, oko, fmore;
    cell fenv, feval, fprobe, fprog;
    cell oeval, oprobe;
    cell iprobe;
    cell t;
    char msg[TEST_BUFSIZE] = {0};
    ⟨ Test the outer environment when testing eval 351 ⟩
    ⟨ Test the inner environment when testing eval 352 ⟩
  }
```

351. ⟨ Test the outer environment when testing **eval** 351 ⟩ ≡

```
  oko = tap_ok(environment_p(outer), tmsgf("(environment?␣outer)"));
  tap_ok(env_root_p(outer), tmsgf("(environment.is-root?␣outer)"));
  if (oko) {
    oeval = env_search(Root, sym("eval"));
    oprobe = env_search(Root, sym("error"));
    t = env_layer(outer);
    TEST_EVAL_FIND
    if (¬undefined_p(fprog)) oki = list_p(fprog, FALSE, &t) ∧ int_value(t) ≡ 2;
        /* TODO: write for match(fprog, read_cstring("(alt-test!probe␣'oops)")) */
  }
  tap_again(oko, ¬fmore ∧ feval ≡ oeval ∧ fprobe ≡ oprobe ∧ fenv ≡ inner,
      tmsgf("outer␣environment␣is␣unchanged"));
```

This code is used in section 350.

352. ⟨ Test the inner environment when testing **eval** 352 ⟩ ≡

```
oki = tap_ok(environment_p(inner), tmsgf("(environment?␣inner)"));
tap_ok(env_root_p(inner), tmsgf("(environment.is-root?␣inner)"));
if (oki) {
  iprobe = env_search(Root, sym("test!probe"));
  t = env_layer(inner);
  TEST_EVAL_FIND
  if (¬undefined_p(fprog)) oki = list_p(fprog, FALSE, &t) ∧ int_value(t) ≡ 2;
}
tap_again(oki, ¬fmore ∧ undefined_p(feval) ∧ fprobe ≡ iprobe ∧ env_empty_p(fenv),
    tmsgf("inner␣environment␣is␣unchanged"));
```

This code is used in section 350.

353. Conditional Integration. Before testing conditional interaction with *environment*s it's reassuring to know that **if**'s syntax works the way that's expected of it, namely that when only the conequent is provided without an alternate it is as though the alternate was the value `VOID`, and that a call to it has no unexpected side-effects.

```
⟨ t/if.c 353 ⟩ ≡
  ⟨ Old test executable wrapper 225 ⟩
  void test_main(void)
  {
    cell fcorrect, tcorrect, fwrong, twrong;
    cell talt, tcons, tq;
    cell marco, polo, t;
    char *prefix = Λ;
    char msg[TEST_BUFSIZE] = {0};
    fcorrect = sym("correct-false");
    fwrong = sym("wrong-false");
    tcorrect = sym("correct-true");
    twrong = sym("wrong-true");
    talt = sym("test-alternate");
    tcons = sym("test-consequent");
    tq = sym("test-query");
    marco = sym("marco?");
    polo = sym("polo!");
    ⟨ Sanity test if's syntax 354 ⟩
    ⟨ Test integrating if 358 ⟩
  }
```

354. Four tests make sure **if**'s arguments work as advertised. These are the only tests of the 2-argument form of **if**.

`(if #t 'polo!)` ⇒ `polo!`:

```
⟨ Sanity test if's syntax 354 ⟩ ≡
  vm_reset();
  t = cons(synquote_new(polo), NIL);
  t = cons(TRUE, t);
  Acc = cons(sym("if"), t);
  prefix = "(if␣#t␣'polo!)";
  interpret();
  tap_ok(symbol_p(Acc) ∧ Acc ≡ polo, tmsgf("symbol?"));
  test_vm_state_full(prefix);
```

See also sections 355, 356, and 357.

This code is used in section 353.

355. `(if #f 'marco?)` ⇒ `VOID`:

```
⟨ Sanity test if's syntax 354 ⟩ +≡
  vm_reset();
  t = cons(synquote_new(marco), NIL);
  t = cons(FALSE, t);
  Acc = cons(sym("if"), t);
  prefix = "(if␣#f␣'marco?)";
  interpret();
  tap_ok(void_p(Acc), tmsgf("void?"));
  test_vm_state_full(prefix);
```

356. `(if #t 'marco? 'polo!)` ⇒ `marco?`:

⟨ Sanity test **if**'s syntax 354 ⟩ +≡

$vm\_reset()$;
$t = cons(synquote\_new(polo), \texttt{NIL})$;
$t = cons(synquote\_new(marco), t)$;
$t = cons(\texttt{TRUE}, t)$;
$Acc = cons(sym(\texttt{"if"}), t)$;
$prefix = \texttt{"(if␣\#t␣'marco?␣'polo!)"}$;
$interpret()$;
$tap\_ok(symbol\_p(Acc) \wedge Acc \equiv marco, tmsgf(\texttt{"symbol?"}))$;
$test\_vm\_state\_full(prefix)$;

357. `(if #f 'marco? 'polo!)` ⇒ `polo!`:

⟨ Sanity test **if**'s syntax 354 ⟩ +≡

$vm\_reset()$;
$t = cons(synquote\_new(polo), \texttt{NIL})$;
$t = cons(synquote\_new(marco), t)$;
$t = cons(\texttt{FALSE}, t)$;
$Acc = cons(sym(\texttt{"if"}), t)$;
$prefix = \texttt{"(if␣\#f␣'marco?␣'polo!)"}$;
$interpret()$;
$tap\_ok(symbol\_p(Acc) \wedge Acc \equiv polo, tmsgf(\texttt{"symbol?"}))$;
$test\_vm\_state\_full(prefix)$;

358. To confirm that **if**'s arguments are evaluated in the correct *environment* *Root* is replaced with a duplicate and invalid variants of the symbols inserted into it. This is then extended into a new *environment* with the desired version of the four symbols **if**, `test-query`, `test-consequent` and `test-alternate`.

⟨ Test integrating **if** 358 ⟩ ≡

$t = env\_layer(Tmp\_Test = Root)$;
$Root = env\_empty()$;
for $(; \neg null\_p(t); t = cdr(t))$
  if $(caar(t) \neq sym(\texttt{"if"}))$ $env\_set(Root, caar(t), cadar(t), btrue)$;
$env\_set(Root, sym(\texttt{"if"}), env\_search(Tmp\_Test, sym(\texttt{"error"})), btrue)$;
$env\_set(Root, talt, fwrong, btrue)$;
$env\_set(Root, tcons, twrong, btrue)$;
$env\_set(Root, tq, \texttt{VOID}, btrue)$;
$Env = env\_extend(Root)$;
$env\_set(Env, sym(\texttt{"if"}), env\_search(Tmp\_Test, sym(\texttt{"if"})), btrue)$;
$env\_set(Env, talt, fcorrect, btrue)$;
$env\_set(Env, tcons, tcorrect, btrue)$;
$env\_set(Env, tq, \texttt{VOID}, btrue)$;

See also sections 359, 360, and 361.
This code is used in section 353.

359. The test is performed with *test-query* resolving to `#f` & `#t`.

⟨ Test integrating **if** 358 ⟩ +≡
```
vm_reset();
env_set(Env, tq, FALSE, bfalse);
t = cons(talt, NIL);
t = cons(tcons, t);
t = cons(tq, t);
Acc = cons(sym("if"), t);
prefix = "(let␣((query␣#f))␣(if␣query␣consequent␣alternate))";
t = Env;
interpret();
tap_ok(symbol_p(Acc) ∧ Acc ≡ fcorrect, tmsgf("symbol?"));
test_vm_state_normal(prefix);
tap_ok(Env ≡ t, tmsgf("(unchanged?␣Env)"));
```

360. ⟨ Test integrating **if** 358 ⟩ +≡
```
vm_reset();
env_set(Env, tq, TRUE, bfalse);
t = cons(talt, NIL);
t = cons(tcons, t);
t = cons(tq, t);
Acc = cons(sym("if"), t);
prefix = "(let␣((query␣#t))␣(if␣query␣consequent␣alternate))";
t = Env;
interpret();
tap_ok(symbol_p(Acc) ∧ Acc ≡ tcorrect, tmsgf("symbol?"));
test_vm_state_normal(prefix);
tap_ok(Env ≡ t, tmsgf("(unchanged?␣Env)"));
```

361. It is important that the real *Root* is restored at the end of these tests in order to perform any more testing.

⟨ Test integrating **if** 358 ⟩ +≡
```
Root = Tmp_Test;
```

362. Applicatives. Testing **lambda** here is mostly concerned with verifying that the correct environment is stored in the closure it creates and then extended when it is entered.

These tests (and **vov**, below) could be performed using higher-level testing and *current-environment* but a) there is no practically usable `LossLess` language yet and b) I have a feeling I may want to write deeper individual tests.

```
⟨ t/lambda.c 362 ⟩ ≡
  ⟨ Old test executable wrapper 225 ⟩
  void test_main(void)
  {
    boolean ok;
    cell ie, oe, len;
    cell t, m, p;
    cell sn, si, sin, sinn, so, sout, soutn;
    char *prefix;
    char msg[TEST_BUFSIZE] = {0};
      /* Although myriad these variables' scope is small and they are not used between the sections */
    sn = sym("n");
    si = sym("inner");
    sin = sym("in");
    sinn = sym("in-n");
    so = sym("outer");
    sout = sym("out");
    soutn = sym("out-n");
    ⟨ Test calling lambda 363 ⟩
    ⟨ Test entering an applicative closure 364 ⟩
    ⟨ Applicative test passing an applicative 365 ⟩
    ⟨ Applicative test passing an operative 366 ⟩
    ⟨ Applicative test returning an applicative 367 ⟩
    ⟨ Applicative test returning an operative 368 ⟩
  }
```

363. An applicative closes over the local *environment* that was active at the point **lambda** was compiled.

```
#define TEST_AB "(lambda␣x)"
#define TEST_AB_PRINT "(lambda␣x␣...)"
⟨ Test calling lambda 363 ⟩ ≡
  Env = env_extend(Root);
  Tmp_Test = test_copy_env();
  Acc = read_cstring(TEST_AB);
  prefix = TEST_AB_PRINT;
  vm_reset();
  interpret();
  ok = tap_ok(applicative_p(Acc), tmsgf("applicative?"));
  tap_again(ok, applicative_formals(Acc) ≡ sym("x"), tmsgf("formals"));
  if (ok) t = applicative_closure(Acc);
  tap_again(ok, environment_p(car(t)), tmsgf("environment?"));
  tap_again(ok, test_is_env(car(t), Tmp_Test), tmsgf("closure"));
  if (ok) t = cdr(t);
  tap_again(ok, car(t) ≠ Prog, tmsgf("prog"));      /* & what? */
  test_vm_state_normal(prefix);
  tap_ok(test_compare_env(Tmp_Test), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 362.

364. When entering an applicative closure the *environment* it closed over at compile-time is extended (into a new frame which is removed when leaving the closure).

```
#define TEST_AC "(lambda␣x␣(test!probe))"
#define TEST_AC_PRINT "("TEST_AC")"
⟨ Test entering an applicative closure 364 ⟩ ≡
  Env = env_extend(Root);
  Tmp_Test = cons(test_copy_env(), NIL);
  Acc = read_cstring(TEST_AC);
  vm_reset();
  interpret();
  Env = env_extend(Root);
  cdr(Tmp_Test) = test_copy_env();
  t = read_cstring("(LAMBDA)");
  car(t) = Acc;
  Acc = t;
  prefix = TEST_AC_PRINT;
  vm_reset();
  interpret();
  t = assoc_value(Acc, sym("Env"));
  ok = tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣T␣'Env))"));
  tap_again(ok, test_is_env(env_parent(t), car(Tmp_Test)),
      tmsgf("(eq?␣(assoc-value␣T␣'Env)␣(env.parent␣E))"));
  test_vm_state_normal(prefix);
  tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 362.

365. Given that we can compile and enter an applicative closure, this test assures that we can correctly enter a closure that's passed as an argument to it. The expression being evaluated is: `((lambda`$_0$ `(L`$_1$ `. x0) (L`$_1$ `(test!probe`$_0$`))) (lambda`$_1$ `(T`$_0$ `. x1) (test!probe`$_1$`)))` except that the same technique as the previous test compiles each expression in its own *environment*.

Entering the outer closure extends the *environment* E_0 to E_1 which will be contained in the probe result that's an argument to the inner closure.

Entering the inner closure extends its *environment* E_2 to E_3.

```
#define TEST_ACA_INNER "(lambda␣(T␣.␣x1)␣(test!probe))"
#define TEST_ACA_OUTER "(lambda␣(L␣.␣x0)␣(L␣(test!probe)))"
#define TEST_ACA "("TEST_ACA_OUTER"LAMBDA)"
#define TEST_ACA_PRINT "("TEST_ACA_OUTER"␣(LAMBDA))"
```

⟨ Applicative test passing an *applicative* 365 ⟩ ≡

```
Env = env_extend(Root);    /* E2 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_ACA_INNER);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);    /* E0 */
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_ACA);
cadr(Acc) = vms_pop();
prefix = TEST_ACA_PRINT;
vm_reset();
interpret();
t = assoc_value(Acc, sym("Env"));    /* E3 */
ok = tap_ok(environment_p(t), tmsgf("(environment?␣inner)"));
if (ok) p = env_search(t, sym("T"));
if (ok) m = assoc_value(p, sym("Env"));    /* E1 */
tap_again(ok, environment_p(m), tmsgf("(environment?␣outer)"));
tap_again(ok, m ≠ t, tmsgf("(eq?␣outer␣inner)"));
tap_again(ok, test_is_env(env_parent(m), cdr(Tmp_Test)), tmsgf("(parent?␣outer)"));
tap_again(ok, test_is_env(env_parent(t), car(Tmp_Test)), tmsgf("(parent?␣inner)"));
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 362.

366. This is the same test, passing/entering an *operative*. The key difference is that the inner *operative* must evaluate its arguments itself. Additionally *test!probe* is an operative so an applicative variant is called: `(vov ((A vov/args) (E vov/env)) (test!probe-applying (eval (car A) E)))`.

The same *environments* are in play as in the previous test with the addition that E_1 will be passed into the inner closure in *vov/environment*.

```
#define TEST_ACO_INNER_BODY "(test!probe-applying␣(eval␣(car␣A)␣E))"
#define TEST_ACO_INNER "(vov␣((A␣vov/args)␣(E␣vov/env))"
          TEST_ACO_INNER_BODY")"
#define TEST_ACO_OUTER "(lambda␣(V␣.␣x0)␣(V␣(test!probe)))"
#define TEST_ACO "("TEST_ACO_OUTER"VOV)"
#define TEST_ACO_PRINT "((LAMBDA)␣(vov␣(...)␣"TEST_ACO_INNER_BODY")"
```

⟨ Applicative test passing an *operative* 366 ⟩ ≡

```
Env = env_extend(Root);      /* E2 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_ACO_INNER);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);      /* E0 */
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_ACO);
cadr(Acc) = vms_pop();
prefix = TEST_ACO_PRINT;
vm_reset();
interpret();
t = assoc_value(Acc, sym("Env"));      /* E3 */
p = car(assoc_value(Acc, sym("Args")));
m = assoc_value(p, sym("Env"));      /* E1 */
ok = tap_ok(environment_p(m), tmsgf("(environment?␣outer)"));
tap_again(ok, test_is_env(env_parent(m), cdr(Tmp_Test)), tmsgf("(parent?␣outer)"));
ok = tap_ok(environment_p(t), tmsgf("(environment?␣inner)"));
if (ok) p = env_search(t, sym("E"));      /* E1 */
tap_again(ok, environment_p(p), tmsgf("(environment?␣E)"));
tap_again(ok, test_is_env(p, m), tmsgf("operative␣environment"));
tap_ok(¬test_is_env(m, t), tmsgf("(eq?␣outer␣inner)"));
tap_again(ok, test_is_env(env_parent(t), car(Tmp_Test)), tmsgf("(parent?␣inner)"));
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 362.

367. Similar to applicatives which call into another closure are applicatives which return one. Starting with an *applicative*-returning-*applicative* `(lambda (outer n) (lambda (inner n) (test!probe)))`.

This is a function which takes two arguments, *outer* and *n* and creates another function which closes over them and takes two of its own arguments, *inner* and *n*.

The test calls this by evaluating `((X 'out 'out-n) 'in 'in-n)` with the above code inserted in the `X` position.

When the inner lambda is evaluating *test!probe* its local *environment* E_2 should be an extension of the dynamic *environment* E_1 that was created when entering the outer closure. E_1 should be an extension of the run-time *environment* E_0 when the closure was built.

```
#define TEST_ARA_INNER  "(lambda␣(inner␣n)␣(test!probe))"
#define TEST_ARA_BUILD  "(lambda␣(outer␣n)␣"TEST_ARA_INNER")"
#define TEST_ARA_PRINT  TEST_ARA_BUILD
#define TEST_ARA_CALL  "((LAMBDA␣'out␣'out-n)␣'in␣'in-n)"
```

⟨ Applicative test returning an *applicative* 367 ⟩ ≡

```
Env = env_extend(Root);    /* E0 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_ARA_BUILD);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_ARA_CALL);
caar(Acc) = vms_pop();
prefix = TEST_ARA_PRINT;
vm_reset();
interpret();
ie = assoc_value(Acc, sym("Env"));    /* E2 */
ok = tap_ok(environment_p(ie), tmsgf("(environment?␣inner)"));
tap_again(ok, env_search(ie, sn) ≡ sinn, tmsgf("(eq?␣n␣'in-n)"));
tap_again(ok, env_search(ie, si) ≡ sin, tmsgf("(eq?␣inner␣'in)"));
tap_again(ok, env_search(ie, so) ≡ sout, tmsgf("(eq?␣outer␣'out)"));
if (ok) oe = env_parent(ie);    /* E1 */
tap_again(ok, environment_p(oe), tmsgf("(environment?␣outer)"));
tap_again(ok, env_search(oe, sn) ≡ soutn, tmsgf("(eq?␣n␣'out-n)"));
tap_again(ok, undefined_p(env_search(oe, si)), tmsgf("(defined?␣inner)"));
tap_again(ok, env_search(oe, so) ≡ sout, tmsgf("(eq?␣outer␣'out)"));
tap_again(ok, test_is_env(env_parent(oe), car(Tmp_Test)), tmsgf("(parent?␣outer)"));    /* E0 */
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 362.

368. Finally, an applicative closing over an operative it returns looks similar: `(vov ((A vov/args) (E vov/env)) (test!probe-applying A E))`

Again the same *environments* are in play although this time the operative's arguments are unevaluated and E_3, the run-time environment, is passed in *vov/environment*.

```
#define TEST_ARO_INNER_BODY "(test!probe-applying␣A␣E)"
#define TEST_ARO_INNER "(vov␣((A␣vov/args)␣(E␣vov/env))"TEST_ARO_INNER_BODY")"
#define TEST_ARO_BUILD "(lambda␣(outer␣n)"TEST_ARO_INNER")"
#define TEST_ARO_CALL "((LAMBDA␣'out␣'out-n)␣'in␣'in-n)"
#define TEST_ARO_PRINT "(LAMBDA␣(vov␣(...)␣"TEST_ARO_INNER_BODY"))"
```

⟨ Applicative test returning an *operative* 368 ⟩ ≡

```
Env = env_extend(Root);   /* E0 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_ARO_BUILD);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);   /* E3 */
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_ARO_CALL);
caar(Acc) = vms_pop();
prefix = TEST_ARO_PRINT;
vm_reset();
interpret();
ie = assoc_value(Acc, sym("Env"));   /* E2 */
ok = tap_ok(environment_p(ie), tmsgf("(environment?␣inner)"));
tap_again(ok, undefined_p(env_here(ie, sn)), tmsgf("(lifted?␣n)"));
tap_again(ok, undefined_p(env_here(ie, so)), tmsgf("(lifted?␣outer)"));
tap_again(ok, env_search(ie, sn) ≡ soutn, tmsgf("(eq?␣n␣'out-n)"));
tap_again(ok, env_search(ie, so) ≡ sout, tmsgf("(eq?␣outer␣'out)"));
if (ok) oe = env_parent(ie);   /* E1 */
tap_again(ok, environment_p(oe), tmsgf("(environment?␣outer)"));
tap_again(ok, env_search(ie, sn) ≡ soutn, tmsgf("(eq?␣n␣'out-n)"));
tap_again(ok, env_search(ie, so) ≡ sout, tmsgf("(eq?␣outer␣'out)"));
tap_again(ok, undefined_p(env_search(oe, sym("A"))), tmsgf("(defined?␣A)"));
tap_again(ok, undefined_p(env_search(oe, sym("E"))), tmsgf("(defined?␣E)"));
tap_again(ok, test_is_env(env_parent(oe), car(Tmp_Test)), tmsgf("(parent?␣outer)"));   /* E0 */
if (ok) t = env_search(ie, sym("A"));
tap_again(ok, true_p(list_p(t, FALSE, &len)), tmsgf("(list?␣A)"));
tap_again(ok, int_value(len) ≡ 2, tmsgf("length"));
tap_again(ok, syntax_p(car(t)) ∧ cdar(t) ≡ sin ∧ syntax_p(cadr(t)) ∧ cdadr(t) ≡ sinn,
    tmsgf("unevaluated"));
tap_again(ok, test_is_env(env_search(ie, sym("E")), cdr(Tmp_Test)), tmsgf("(eq?␣E␣Env)"));
  /* E3 */
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 362.

369. Operatives. Testing **vov** follows the same plan as **lambda** with the obvious changes to which environment is expected to be found where and care taken to ensure that arguments are evaluated when appropriate.

```
⟨ t/vov.c 369 ⟩ ≡
  ⟨ Old test executable wrapper 225 ⟩
  void test_main(void)
  {
    boolean ok;
    cell t, m, p;
    cell sn, si, sin, sinn, so, sout, soutn;
    char *prefix;
    char msg[TEST_BUFSIZE] = {0};
    sn = sym("n");
    si = sym("inner");
    sin = sym("in");
    sinn = sym("in-n");
    so = sym("outer");
    sout = sym("out");
    soutn = sym("out-n");
    ⟨ Test calling vov 370 ⟩
    ⟨ Test entering an operative closure 371 ⟩
    ⟨ Operative test passing an applicative 372 ⟩
    ⟨ Operative test passing an operative 373 ⟩
    ⟨ Operative test returning an applicative 374 ⟩
    ⟨ Operative test returning an operative 375 ⟩
  }
```

370.

```
#define TEST_OB "(vov␣((E␣vov/env)))"
#define TEST_OB_PRINT "(vov␣((E␣vov/env))␣...)"
⟨ Test calling vov 370 ⟩ ≡
  Env = env_extend(Root);
  Tmp_Test = test_copy_env();
  Acc = read_cstring(TEST_OB);
  prefix = TEST_OB_PRINT;
  vm_reset();
  interpret();
  ok = tap_ok(operative_p(Acc), tmsgf("operative?"));
  tap_again(ok, pair_p(t = operative_formals(Acc)), tmsgf("formals"));
  if (ok) t = operative_closure(Acc);
  tap_again(ok, environment_p(car(t)), tmsgf("environment?"));
  tap_again(ok, car(t) ≡ Env, tmsgf("closure"));
  if (ok) t = cdr(t);
  tap_again(ok, car(t) ≠ Prog, tmsgf("prog"));    /* & what? */
  test_vm_state_normal(prefix);
  tap_ok(test_compare_env(Tmp_Test), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 369.

371. Upon entering an operative closure:

1. The run-time *environment* E_0 when it was created is extended to a new *environment* E_1 contaning the 1-3 **vov** arguments.

2. The run-time *environment* E_2 when it was entered is passed to the **vov** in the argument in the *vov/environment* (or *vov/env*) position.

3. Upon leaving it the stack and the run-time *environment* are restored unchanged.

```
#define TEST_OC "(vov␣((A␣vov/args)␣(E␣vov/env))␣(test!probe-applying␣A␣E))"
#define TEST_OC_PRINT "((vov␣(...)␣(test!probe-applying␣A␣E)))"
```

⟨ Test entering an *operative* closure 371 ⟩ ≡

```
Env = env_extend(Root);    /* E0 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_OC);
vm_reset();
interpret();
Env = env_extend(Root);    /* E2 */
cdr(Tmp_Test) = test_copy_env();
t = read_cstring("(VOV)");
car(t) = Acc;
Acc = t;
prefix = TEST_OC_PRINT;
vm_reset();
interpret();
t = assoc_value(Acc, sym("Env"));    /* E1 */
ok = tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣T␣'Env))"));
tap_again(ok, test_is_env(env_parent(t), car(Tmp_Test)),
    tmsgf("(eq?␣(assoc-value␣T␣'Env)␣(env.parent␣E))"));
if (ok) p = env_search(t, sym("E"));    /* E2 */
tap_again(ok, environment_p(p), tmsgf("(environment?␣E)"));
tap_again(ok, test_is_env(p, cdr(Tmp_Test)), tmsgf("(eq?␣T␣(current-environment))"));
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 369.

372. Calling an applicative inside an operative closure is no different from any other function call. An operative closure is entered with the result of **lambda** as an argument: `((VOV) (lambda x1 (test!probe)))`.

Operative's arguments are not evaluated so whether a **lambda** expression, variable lookup or whatever the operative evaluates its argument in the caller's *environment* then calls into it along with its own probe: `((vov (...) (cons ((eval (car A) E)) (test!probe))) (LAMBDA))`.

The operative's compile-time *environment* E_0 is extended up entering it to E_1. The run-time *environment* E_2 is extended when entering the callee's applicative and is passed to the operative.

```
#define TEST_OCA_INNER "(lambda␣x1␣(test!probe))"
#define TEST_OCA_OUTER
        "(vov␣((A␣vov/args)␣(E␣vov/env))""(cons␣((eval␣(car␣A)␣E))␣(test!probe)))"
#define TEST_OCA "("TEST_OCA_OUTER"LAMBDA)"
#define TEST_OCA_PRINT "((VOV)␣"TEST_OCA_INNER")"
```

⟨ Operative test passing an *applicative* 372 ⟩ ≡

```
Env = env_extend(Root);    /* E0 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_OCA_INNER);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);    /* E2 */
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_OCA);
cadr(Acc) = vms_pop();
prefix = TEST_OCA_PRINT;
vm_reset();
interpret();
t = assoc_value(cdr(Acc), sym("Env"));    /* E1 */
ok = tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣(cdr␣T)␣'Env))"));
tap_again(ok, test_is_env(env_parent(t), cdr(Tmp_Test)), tmsgf("(parent?␣E)"));    /* E0 */
tap_again(ok, test_is_env(env_search(t, sym("E")), cdr(Tmp_Test)), tmsgf("(eq?␣E␣vov/env)"));
p = assoc_value(car(Acc), sym("Env"));    /* E3 */
ok = tap_ok(environment_p(p), tmsgf("(environment?␣(assoc-value␣(car␣T)␣'Env))"));
tap_again(ok, test_is_env(env_parent(p), car(Tmp_Test)), tmsgf("(parent?␣E')"));    /* E2 */
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 369.

373. To verify calling an operative argument to an operative closure there are three tests to perform:

1. The run-time *environment* E_2 in the inner operative is an extension of the one it was originally created with E_1.
2. The run-time *environment* E_1 in the outer operative is an extension of its compile-time *environment* E_0.
3. E_1 is the *vov/environment* argument of the inner operative.

```
#define TEST_OCO_INNER "(vov␣((yE␣vov/env))␣(test!probe))"
#define TEST_OCO_OUTER "(vov␣((xA␣vov/args)␣(xE␣vov/env))"
        "(cons␣((eval␣(car␣xA)␣xE))␣(test!probe)))"
#define TEST_OCO "("TEST_OCO_OUTERTEST_OCO_INNER")"
#define TEST_OCO_PRINT "((VOV)␣"TEST_OCO_INNER")"
```

⟨ Operative test passing an *operative* 373 ⟩ ≡

```
Env = env_extend(Root);     /* E0 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_OCO_INNER);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_OCO);
cadr(Acc) = vms_pop();
prefix = TEST_OCO_PRINT;
vm_reset();
interpret();
t = assoc_value(car(Acc), sym("Env"));     /* E2 */
ok = tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣(car␣T)␣'Env))"));
tap_again(ok, test_is_env(env_parent(t), car(Tmp_Test)), tmsgf("(parent?␣E)"));     /* E1 */
m = env_here(t, sym("yE"));     /* E1 */
tap_again(ok, ¬undefined_p(m), tmsgf("(env.exists?␣E␣yE)"));
p = assoc_value(cdr(Acc), sym("Env"));     /* E1 */
ok = tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣(cdr␣T)␣'Env))"));
tap_again(ok, test_is_env(m, p), tmsgf("operative␣environment"));
tap_ok(¬test_is_env(p, t), tmsgf("(eq?␣E'␣E)"));
tap_again(ok, test_is_env(env_parent(p), cdr(Tmp_Test)), tmsgf("(parent?␣E')"));     /* E0 */
tap_again(ok, ¬undefined_p(env_here(p, sym("xE"))), tmsgf("(env.exists?␣E'␣xE)"));
tap_again(ok, ¬undefined_p(env_here(p, sym("xA"))), tmsgf("(env.exists?␣E'␣xA)"));
tap_ok(test_is_env(p, m), tmsgf("(eq?␣E'␣yE)"));
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 369.

374. Building applicatives and operatives within an operative requires extra care to evaluate code in the correct *environment*.

The *environment* E_1 that a returned applicative closes over, and will extend into E_2 when it's entered, is the local *environment* of the operative.

The outer operative evaluates its two arguments in its caller's *environment* E_0, saving them in *outer* and *n* in turn, and then calls **lambda**.

```
#define TEST_ORA_INNER "(lambda␣(inner␣n)␣(test!probe))"
#define TEST_ORA_MIXUP "(define!␣(current-environment)␣inner␣'out)""(define!␣(current-e\
        nvironment)␣outer␣(eval␣(car␣yA)␣yE))""(define!␣(current-environment)␣n␣(eval␣\
        (car␣(cdr␣yA))␣yE))"
#define TEST_ORA_BUILD
        "(vov␣((yA␣vov/args)␣(yE␣vov/env))"TEST_ORA_MIXUPTEST_ORA_INNER")"
#define TEST_ORA_CALL "((VOV␣'out␣'out-n)␣'in␣'in-n)"
#define TEST_ORA_PRINT "(vov␣(...)␣(lambda␣(inner␣n)␣(test!probe)))"
```

⟨ Operative test returning an *applicative* 374 ⟩ ≡

```
Env = env_extend(Root);   /* E0 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_ORA_BUILD);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_ORA_CALL);
caar(Acc) = vms_pop();
prefix = TEST_ORA_PRINT;
vm_reset();
interpret();
t = assoc_value(Acc, sym("Env"));   /* E2 */
ok = tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣(cdr␣T)␣'Env))"));
m = env_here(t, sym("n"));
tap_again(ok, m ≡ sinn, tmsgf("(eq?␣(env.here␣E␣n)␣'in-n)"));
m = env_here(t, sym("inner"));
tap_again(ok, m ≡ sin, tmsgf("(eq?␣(env.here␣E␣inner)␣'in)"));
tap_again(ok, undefined_p(env_here(t, sym("outer"))), tmsgf("(exists-here?␣E␣outer)"));
m = env_search(t, sym("outer"));
tap_again(ok, m ≡ sout, tmsgf("(eq?␣(env.lookup␣E␣inner)␣'out)"));
if (ok) p = env_parent(t);   /* E1 */
tap_again(ok, ¬undefined_p(env_here(p, sym("yE"))), tmsgf("(exists?␣(env.parent␣E)␣yE)"));
tap_again(ok, test_is_env(env_parent(p), car(Tmp_Test)), tmsgf("(env.parent?␣E')"));   /* E0 */
m = env_here(p, sym("n"));
tap_again(ok, m ≡ soutn, tmsgf("(eq?␣(env.here␣E␣n)␣'out-n)"));
m = env_here(p, sym("inner"));
tap_again(ok, m ≡ sout, tmsgf("(eq?␣(env.here␣E␣inner)␣'out)"));
m = env_here(p, sym("outer"));
tap_again(ok, m ≡ sout, tmsgf("(eq?␣(env.lookup␣E␣inner)␣'out)"));
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 369.

375. Closing over an operative within an operative requires even more care that the correct environment is used so that the returned operative has access to its creator's local environment.

The creating operative extends the *environment* E_0 it closes over and this *environment* E_1 is then closed over by the returned operative. E_1 is extended upon entering the inner operative into *environment* E_2.

The same run-time *environment* E_3 is passed as an argument to the each operative.

```
#define TEST_ORO_INNER_BODY "(test!probe-applying␣(eval␣'(test!probe)␣oE))"
#define TEST_ORO_INNER "(vov␣((oE␣vov/env))"TEST_ORO_INNER_BODY")"
#define TEST_ORO_BUILD "(vov␣((A␣vov/args)␣(E␣vov/env))"TEST_ORO_INNER")"
#define TEST_ORO_CALL "((VOV␣'out␣'out-n)␣'in␣'in-n)"
#define TEST_ORO_PRINT "(VOV␣(vov␣(...)␣(test!probe␣(eval␣'(test!probe)␣E)))"
```

⟨ Operative test returning an *operative* 375 ⟩ ≡

```
Env = env_extend(Root);    /* E0 */
Tmp_Test = cons(test_copy_env(), NIL);
Acc = read_cstring(TEST_ORO_BUILD);
vm_reset();
interpret();
vms_push(Acc);
Env = env_extend(Root);    /* E3 */
cdr(Tmp_Test) = test_copy_env();
Acc = read_cstring(TEST_ORO_CALL);
caar(Acc) = vms_pop();
prefix = TEST_ORO_PRINT;
vm_reset();
interpret();
t = assoc_value(Acc, sym("Env"));    /* E2 */
ok = tap_ok(environment_p(t), tmsgf("(environment?␣(assoc-value␣T␣'Env))"));
if (ok) m = env_here(t, sym("oE"));    /* E3 */
tap_again(ok, environment_p(m), tmsgf("(environment?␣oE)"));
tap_again(ok, m ≡ cdr(Tmp_Test), tmsgf("(eq?␣E␣Env)"));
if (ok) m = env_parent(t);    /* E1 */
tap_again(ok, ¬undefined_p(env_here(m, sym("A"))), tmsgf("(env.exists?␣E'␣A)"));
if (ok) p = env_here(m, sym("E"));    /* E3 */
tap_again(ok, ¬undefined_p(env_here(m, sym("E"))), tmsgf("(env.exists?␣E'␣E)"));
tap_again(ok, p ≡ cdr(Tmp_Test), tmsgf("(eq?␣E''␣Env)"));
tap_again(ok, env_parent(m) ≡ car(Tmp_Test), tmsgf("(eq?␣(env.parent␣E')␣Env)"));
test_vm_state_normal(prefix);
tap_ok(test_compare_env(cdr(Tmp_Test)), tmsgf("(unchanged?␣Env)"));
```

This code is used in section 369.

376. Exceptions. When an error occurs at run-time it has the option (unimplemented) to be handled at run-time but if it isn't then control returns to before the beginning of the main loop. Each time around the main loop, *interpret* begins by calling *vm_reset* but that explicitely *doesn't* change the *environment* to allow for run-time mutation and expects that well-behaved code will clear the stack correctly.

These exception tests enter a closure, which creates a stack frame, and call **error** within it. The tests then ensure that the *environment* and stack are ready to compute again.

There is no actual support for exception handlers so the interpreter will halt and jump back *Goto_Begin*.

```
#define GOTO_FAIL "((lambda␣x␣(error␣fail)))"
```

```
⟨ t/exception.c 376 ⟩ ≡
  ⟨ Old test executable wrapper 225 ⟩
  void test_main(void)
  {
    volatile boolean first = btrue;
    volatile boolean failed = bfalse;     /* WARNING: ERROR: SUCCESS */
    boolean ok;

    Error_Handler = btrue;
    vm_prepare();
    if (first) {
      first = bfalse;
      vm_reset();
      Acc = read_cstring(GOTO_FAIL);
      interpret();
    }
    else failed = btrue;
    ok = tap_ok(failed, "an␣error␣is␣raised");
    test_vm_state(GOTO_FAIL, TEST_VMSTATE_RUNNING | TEST_VMSTATE_NOT_INTERRUPTED |
        TEST_VMSTATE_ENV_ROOT | TEST_VMSTATE_STACKS);
  }
```

377. TODO.

```
⟨ List of opcode primitives 377 ⟩ ≡   /* Core: */
  {"error", compile_error}, {"eval", compile_eval}, {"if", compile_conditional}, {"lambda",
      compile_lambda}, {"vov", compile_vov}, {"quote", compile_quote}, {"quasiquote",
      compile_quasiquote},   /* Pairs: */
  {"car", compile_car}, {"cdr", compile_cdr}, {"cons", compile_cons}, {"null?", compile_null_p},
      {"pair?", compile_pair_p}, {"set-car!", compile_set_car_m}, {"set-cdr!", compile_set_cdr_m},
    /* Mutation: */
  {"current-environment", compile_env_current}, {"root-environment", compile_env_root}, {"set!",
      compile_set_m}, {"define!", compile_define_m},
#ifdef LL_TEST
  ⟨ Testing primitives 221 ⟩
#endif
```

This code is used in section 89.

378. REPL. The *main* loop is a simple repl.

```
⟨ repl.c 378 ⟩ ≡
#include "lossless.h"
  int main(int argc, char **argv __unused)
  {
    vm_init();
    if (argc > 1) {
      printf("usage: %s", argv[0]);
      return EXIT_FAILURE;
    }
    vm_prepare();
    while (1) {
      vm_reset();
      printf("> ");
      Acc = read_form();
      if (eof_p(Acc) ∨ Interrupt) break;
      interpret();
      if (¬void_p(Acc)) {
        write_form(Acc, 0);
        printf("\n");
      }
    }
    if (Interrupt) printf("Interrupted");
    return EXIT_SUCCESS;
  }
```

379. Association Lists.

⟨ Function declarations 8 ⟩ +≡

```
cell assoc_member(cell, cell);
cell assoc_content(cell, cell);
cell assoc_value(cell, cell);
```

380.

```
cell assoc_member(cell alist, cell needle)
{
  if (¬symbol_p(needle)) error (ERR_ARITY_SYNTAX, NIL);
  if (¬list_p(alist, FALSE, Λ)) error (ERR_ARITY_SYNTAX, NIL);
  for ( ; pair_p(alist); alist = cdr(alist))
    if (caar(alist) ≡ needle) return car(alist);
  return FALSE;
}
cell assoc_content(cell alist, cell needle)
{
  cell r;
  r = assoc_member(alist, needle);
  if (¬pair_p(r)) error (ERR_UNEXPECTED, r);
  return cdr(r);
}
cell assoc_value(cell alist, cell needle)
{
  cell r;
  r = assoc_member(alist, needle);
  if (¬pair_p(cdr(r))) error (ERR_UNEXPECTED, r);
  return cadr(r);
}
```

381. Misc.

382. **#define** *synquote_new*(o) *atom*$(Sym\_SYNTAX\_QUOTE, (o),$ `FORMAT_SYNTAX`$)$
$/* */$

383. Index.

LossLess Programming Environment