



**HAL**  
open science

## Tips for making the most of 64-bit architectures in language design, libraries or garbage collection

Benoît Sonntag, Dominique Colnet

### ► To cite this version:

Benoît Sonntag, Dominique Colnet. Tips for making the most of 64-bit architectures in language design, libraries or garbage collection. 2023. hal-04245259

**HAL Id: hal-04245259**

**<https://hal.science/hal-04245259v1>**

Preprint submitted on 23 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Tips for making the most of 64-bit architectures in language design, libraries or garbage collection

Benoît Sonntag\*  
Benoit.Sonntag@lisaac.org  
Université de Strasbourg  
Strasbourg, France

Dominique Colnet  
Dominique.Colnet@loria.fr  
Université de Lorraine  
Nancy, France

## Abstract

The 64-bit architectures that have become standard today offer unprecedented low-level programming possibilities. For the first time in the history of computing, the size of address registers far exceeded the physical capacity of their bus. After a brief reminder of the possibilities offered by the small size of addresses compared to the available 64 bits, we develop three concrete examples of how the vacant bits of these registers can be used. Among these examples, two of them concern the implementation of a library for a new statically typed programming language. Firstly, the implementation of multi-precision integers, with the aim of improving performance in terms of both calculation speed and RAM savings. The second example focuses on the library's handling of UTF-8 character strings. Here, the idea is to make indexing easier by ignoring the physical size of each UTF-8 characters. Finally, the third example is a possible enhancement of garbage collectors, in particular the mark & sweep for the object marking phase.

**Keywords:** 64 bits architecture, address bus, large integer, UTF-8 strings, GC marking bit

## 1 The revenge of address registers

Several times in its history, in the competition to increase the power of computers, we've seen the emergence of tricks to first increase the size of the address bus before questioning the whole architecture.

For example, the mythical Z80 processor, with its 8-bit data bus and 16-bit addressing bus, should have been limited to 64 KB of RAM. But, with its two-memory bank access principle, the Amstrad 6128 extends its memory to 128KB, doubling the capacity of its predecessor, the 464.

As part of the 8086 family with its 16-bit address register architecture, this processor has been equipped with a 20-bit address bus thanks to the addition of segment registers. The address is made up of a pair of registers {segment, offset}. The offset provides access to a contiguous 64KB range in memory. As for the segment registers, they allow memory jumps in 16-byte steps. Thus, the segment registers provide the 4 bits of high address required to reach the megabyte of RAM (1 MB). The final address calculation is given by:

$$Address = R_{segment} \times 16 + R_{offset}$$

After a brief appearance of 24-bit processors with the 80286, came the 32-bit 80386 architecture, with a bus capable of addressing 4 GB of memory. But again, in the last years of his reign, with a clever combination of segmentation and pagination, the limit of this architecture has been pushed back to 36-bit addressing.

Then came 64-bit architecture, which for the first time in the history of computing has address registers that can go far beyond the memory capacities currently physically available. The astronomical number of 16 Eio ( $2^{64} = 1.8e19$  bytes) that such a register could theoretically address is so far out of reach, even in the distant future, that designers preferred to truncate the logical address to 48 bits. In this apparently arbitrary choice, the number of indirections needed to manage pagination must also be taken into account. In fact, with 4 KB pages, we have the 12 least significant bits addressable contiguously, then 4 indirection tables of 9 bits each must be consulted to reach a 48-bit physical address<sup>1</sup>.

So, quite surprisingly, this 48-bit logical address leaves a 16-bit high-order range unused for each address pointer location. In addition, we can observe that memory allocators always allocate structures aligned at least with the machine word size. So, if we consider a structure address, we also have the 3 least significant bits, which are always 0. Interestingly, on a 64-bit architecture, all structure pointers have only 45 significant bits. However, their use requires a few precautions before considering them as pointers. It's worth noting that using a mask with a binary-and (&) or a 3-bit binary shift to the left (<<) are manipulations that have a marginal cost at runtime. For each address, we therefore have 16 high-order bits at our disposal, and potentially 3 low-order bits that are free for other uses!

The thread running through this article is how best to use these insignificant bits in addresses to store precise information in a given context. This practice is already widely used in the design of architectures and operating systems. For example, in a pagination indirection table, each page address is aligned on 4 KB, so the 12 least significant bits are ignored when the MMU reads the page address. These 12 unused bits contain other indicators, such as the right to write in this

\*Both authors contributed equally to this research.

<sup>1</sup>512 entrées par table d'indirection de 4KB:  $2^9 \times 64 \text{ bits} = 4\text{KB}$ .

very memory page. There are also other indicators defined and used solely by the operating system.

Further away from hardware and operating systems, we also find this kind of approach at the software level, or more precisely at the language and compilation level in [1]. In Bonds et al (1992), *Tracking bad apples: Reporting the origin of null and undefined value errors*, the proposal is to internally replace the Null value of a source program with deliberately invalid pointers to encode information. Here, the information is used to trace the origin of the Null in the event of an application crash.

Let's now look at how we can take advantage of all these findings concerning address format, starting with the management of numbers without overflow.

## 2 Towards integers that don't overflow

For the vast majority of programming languages, built-in integer types are limited to a certain size. For example, Java's `int` type is limited to 32 bits. A few rare cases of overflow are sometimes detected by the Java compiler only when the values are statically determinable. In all cases, no overflow test is performed at runtime. In this way, a positive value of type `int` can be made negative by simply incrementing by 1. This is often surprising, especially for novice programmers.

Another example from a more recent language is Rust, which notes this type `i32` or `u32` for the unsigned version. In addition to the more appropriate type name, Rust in *debug* mode only, offers overflow control. In *release* mode, overflow control is not performed. As you might expect, Rust's objective is to take full advantage of the processor's power. There are even specialized functions to bypass overflow problems regardless of the compilation mode<sup>2</sup>. For large numbers without overflow, Rust also offers types such as `BigInt` or `BigUint`.

Historically, Smalltalk was the first language to natively integrate the concept of an integer that never overflows. This choice is perfectly understandable in terms of comfort for programmers and for people unfamiliar with hardware constraints. For Python, which is also interpreted, the choice is similar to Smalltalk: no possible overflow and numbers that can grow in memory size as and when required. In our opinion, there are two major drawbacks to this choice: slowness and the impossibility of easily taking hardware into account.

### 2.1 Hardware limited types and flexible general type

In the Smalltalk language virtual machine, each 32-bit word representing an object is split into two parts[2]. The first part stores the information in 30 bits, the remaining two bits being the object's basic type. We therefore have 4 internal

object categories, one of which is reserved for a small 30-bit integer. At runtime, if a calculation exceeds 30 bits, the small integer becomes a complex object with a pointer to model a larger integer via a real object which uses an array. Smalltalk's object architecture was particularly well thought-out for its time, in the context of a 32-bit architecture and a pure, non-statically typed object environment. However, we can only note a 30-bit address limit representing only 1 GB of accessible memory. We've taken inspiration from this type of partitioning in the context of 64-bit architecture on a compiled language with static typing.

A good programming language needs to offer both programming comfort and full hardware speed. The choice of the Rust language, which makes it possible to preserve types that can exactly match the characteristics of the hardware, must be maintained (i.e. built-in types `i8`, `i32` and `i64` for signed and built-in types `u8`, `u32` and `u64` for unsigned). Rather than resorting to general types specialized in handling numbers without overflow, we propose a hybrid solution that allows, according to variations in calculations, to retain almost all the power of the small, limited numbers that exist natively. In fact, depending on the evolution of calculations, for example, two very large numbers that are subtracted from each other can return to the 32-bit representation interval. In such a case, it's interesting to get closer to the performance of native types for the result of this operation.

Unlike Smalltalk, which is completely and uniquely typed at runtime, we're working with a statically typed language. For our proposal, it doesn't matter whether the typing is explicit or based on type inference. Knowing that a given variable can only contain signed integers means that we can reuse the Smalltalk implementation idea with less variability in the entities represented. So, unlike Smalltalk, and thanks to static typing, we have the whole machine word to best encode our integer. In the following, we'll assume that the programming language offers a native type for handling signed integers, which we'll call  $\mathbb{Z}$ . The idea is to always use a 64-bit machine word for any variable of type  $\mathbb{Z}$ .

Internally, and completely transparently to the user, there are three possible representations for the  $\mathbb{Z}$  type. The aim is to make the most of 64-bit and get the best performance out of it. The choice between these 3 representations is directly related to the size of the integer you need to model. Figure 1 illustrates the following explanations of our 3 encoding formats for integers.

### 2.2 Tiny `int`: $n \leq 63$ bits - see top of figure 1

The first encoding format, *Tiny*, is for a small integer that can be encoded on 63 bits, that is in two's complement the range from  $(-2^{62})$  to  $(2^{62} - 1)$ . This format is distinguished from the other two by the last bit 63 being set to 0. Unsurprisingly, the use of the other 63 least significant bits stores our integer. With a barely perceptible increase in computing

<sup>2</sup>`u32.wrapping_add`, `u64.wrapping_add`, `u64.wrapping_add_signed`, `u64.wrapping_sub`, etc.

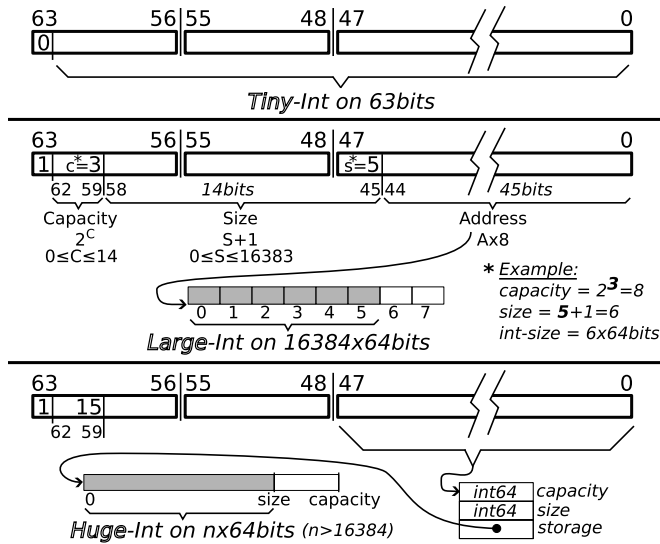


Figure 1. The 3 integer encoding formats of the new  $\mathbb{Z}$  type.

time compared to using a 64-bit raw machine word, it enables fast management of small integers. As with Smalltalk, when this capacity is exceeded, we dynamically migrate the integer to the second format we named Large.

### 2.3 Large int: $64bits \leq n \leq 2^{20} bits$ - middle of figure 1

The second format for the  $\mathbb{Z}$  type, which we've named Large, is the most complex and is particularly compact. It can encode integers requiring a maximum of  $16384 \times 64$  bits (i.e. an integer of 131 072 bytes or 1 Mbits). Thanks to its already particularly wide range, it's more than sufficient for most applications. The 64-bit machine word representing the integer stores 3 pieces of information:

- The address of a contiguous memory area containing the integer using word of 64 bits.
- The maximum capacity of this memory area. If necessary, the memory area is reallocated with a capacity twice that of the previous one<sup>3</sup>.
- The size actually used in the memory area. In other words, the number of machine words needed to represent the integer in binary form.

The distribution of information within the 64 bits is as follows:

- 63 Set to 1 to avoid being identified as a Tiny encoding.
- 59-62 The capacity in power of 2. This 4-bit number can be used to encode capacities ranging from  $2^0$  to  $2^{14}$ . Setting the 4 bits to 1 ( $2^{15}$ ) is forbidden, and we reserve this value to identify the Huge encoding format. We therefore have an array with a maximum capacity of 16384 64-bit cells.

<sup>3</sup>The well-known heuristic of doubling the capacity of a dynamic array is especially relevant to our use case. In this case, the multiplication of two  $n$ -bit integers uses  $2 \times n$  bits.

45-58 This 14-bit range encodes the exact size actually used, from 1 to 16384.

0-44 A 45-bit range representing the address of the corresponding allocated area. As this area is 64-bit aligned, a 3-bit left shift gives the exact 48-bit valid address.

### 2.4 Huge int: $2^{20} < n < \text{whole memory}$ - bottom fig. 1

The last format for type  $\mathbb{Z}$  is shown at the bottom of figure 1 and is of a more standard design. It is identified by the presence of 1 on all bits from 59 to 63. The least significant part of the first 48 bits is an address to a standard object structure. The corresponding object contains a capacity, size and storage field for the usual implementation of a dynamic size array containing our Huge integer<sup>4</sup>.

### 2.5 Integer decoding pseudo-code

Algorithm 1 illustrates in pseudo-code the decoding of different formats of the type  $\mathbb{Z}$ . In this algorithm  $W$  is the machine

#### Algorithm 1 Reading an integer format Tiny, Large or Huge

```

if (W ≥ 0) then
  // Format TINY
  INT ← (((signed64)W) << 1) >> 1)
else
  cap ← (W >> 58) & Fh
  if (cap = 1111b) then
    // Format HUGE
    a ← W & FFFF FFFF FFFFh
    buf ← a.storage
    siz ← a.size
    cap ← a.capacity
  else
    // Format LARGE
    buf ← (W & 1FFF FFFF FFFFh) << 3
    siz ← ((W >> 44) & 3FFFh) + 1
    cap ← 1 << cap
  end if
  INT ← tab(buf, siz, cap)
end if

```

word representing the integer encoded in one of the 3 formats, and  $INT$  represents bit access to the integer. Note that when using a small integer of less than 64 bits, the extra cost compared to the standard 64-bit basic type is just sign detection (or the position of bit 63) and a jump.

Note also that our representation is deliberately canonical. Depending on its size in number of bits, a given number has to be represented in just one of the three categories, Tiny, Large or Huge. This makes it easy to compare numbers with each other. In particular, the comparison between two

<sup>4</sup>This representation is named ArrayList in the Java library. In C++ this data structure is also known as `std::vector`.

numbers Tiny is made with the usual machine instruction for comparing two memory words.

### 3 String indexing in UTF-8 format

The use of the UTF-8 format for strings has become widespread. The main advantage is its compatibility with the standard ASCII format. It also offers compact encoding of unicode characters strings, with variable character sizes. Its encoding allows a string to be read in both directions, which means that many older, highly efficient string processing algorithms remain applicable. The penalizing management of UTF-16 in native as in Java is no longer desirable. Today, the development of a new programming language clearly suggests native management of UTF-8.

Compared with ASCII and despite its advantages, handling UTF-8 remains difficult and rather unnatural due to the variable character size. The conversion of an UTF-8 character into unicode contained in an array of bytes is performed by the standard algorithm 2. Processing and calculating a string

---

#### Algorithm 2 Standard UTF-8 to unicode conversion.

---

```

if  $ptr[0] < 80h$  then
     $ucode \leftarrow ptr[0]$ 
else if  $(ptr[0] \& E0h) = C0h$  and
     $(ptr[1] \& C0h) = 80h$  then
     $ucode \leftarrow ((ptr[0] \& 1Fh) \ll 6) | (ptr[1] \& 3Fh)$ 
else if  $(ptr[0] \& F0h) = E0h$  and
     $(ptr[1] \& C0h) = 80h$  and
     $(ptr[2] \& C0h) = 80h$  then
     $ucode \leftarrow ((ptr[0] \& 0Fh) \ll 12) |$ 
     $((ptr[1] \& 3Fh) \ll 6) | (ptr[2] \& 3Fh)$ 
else if  $(ptr[0] \& F8h) = F0h$  and
     $(ptr[1] \& C0h) = 80h$  and
     $(ptr[2] \& C0h) = 80h$  and
     $(ptr[3] \& C0h) = 80h$  then
     $ucode \leftarrow ((ptr[0] \& 07h) \ll 18) |$ 
     $((ptr[1] \& 3Fh) \ll 12) | ((ptr[2] \& 3Fh) \ll 6) |$ 
     $(ptr[3] \& 3Fh)$ 
else
    print "Invalid UTF-8 sequence"
end if
Output  $ucode$ 

```

---

index is much trickier. A number of libraries are currently available (ICU, UTF8-CPP, libunistring, libiconv,...) and are essentially based on two approaches to the treatment of UTF-8.

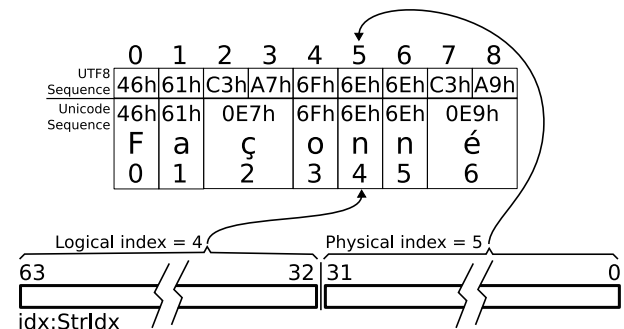
The first approach is to provide an interface for converting from UTF-8 to unicode and vice versa. Thus, the string translated into unicode is made up of 32-bit words, i.e. one 32-bit word for each character. With the exception of very large strings, processing becomes as easy as in the days of the ASCII format. The disadvantages of this technique are, of

course, the translation from one format to another, and the fact that in the vast majority of cases, memory representation is quite costly.

The second approach to managing UTF-8 strings is to offer the user high-level functions that don't require indexes. Its principle is based on the use of regular expressions similar to those found in `grep`, `sed` ou `awk`. It's a good choice in many situations. However, these are complex mechanisms and their intensive use can quickly become a drag in terms of performance. Furthermore, the programmer can quickly become frustrated at not having direct and easy access to the internal structure of the string via character-by-character indexing.

We propose a different approach to virtualize the notion of index. Our technique is necessarily slightly more expensive than a simple ASCII string, but no more expensive than a standard implementation of UTF-8 string management. The programmer can use indexes to consider a character jump as always being a single logical step. The corresponding physical step in memory is variable, but remains transparent to the user.

Our library is based on a single object type `StrIdx` modeled by a single 64-bit word. The idea is to represent the logical index and the physical index in the same word, split into two parts of 32 bits each. In the example shown in figure 2, the logical index of the first letter 'n' is 4 and its physical index is 5, due to the presence of the letter 'ç' before it. With



**Figure 2.** Type `StrIdx` includes both logical and physical index encoding. All in 64-bit memory.

this representation, in the worst case, the ratio between a UTF-8 character and its memory occupancy is one to 4. With an arbitrary division of 32 bits for the physical part and 32 bits for the logical part, we can process contiguous strings of 1 to 4 billion characters.

So, with a single 64-bit word, both the logical index and the corresponding physical index are stored. Note that if `idx` is of type `StrIdx`, the expression `(idx >> 32)` gives access to the logical index. The physical index is obtained with the `(idx & FFFF FFFFh)` expression. Algorithm 3 calculates the next `StrIdx` index from a given `StrIdx` index. Moving back one character is also easy to calculate (see algorithm 4). It's

---

**Algorithm 3** Move idx of type StrIdx one step forward.

---

```
code ← str.at((int32)idx)
idx ← idx + 1 0000 0001h
if (code & 80h) then
  repeat
    code ← code << 1
    idx ← idx + 1
  until (code & 40h = 0)
end if
Output idx
```

---



---

**Algorithm 4** Move idx of type StrIdx one step backward.

---

```
idx ← idx - 1 0000 0001h
while (str.at((int32)idx) & 80h) do
  idx ← idx - 1
end while
Output idx
```

---

worth noting that there's no extra cost involved in managing our index compared to what's usually done when handling the UTF-8 format. With advanced languages capable of defining binary operators, it is possible to facilitate the use of the two preceding algorithms. With the definition of a '+' and '-' operator taking as receiver an index of type StrIdx and as parameter an integer, we can move the index by simple arithmetic operation by any number of characters, regardless of their size. Finally, we propose Algorithm 5, which is an

---

**Algorithm 5** Move forward and compute unicode skipped.

---

```
unicode ← str.at((int32)idx)
idx ← idx + 1 0000 0001h
if (unicode & 80h) then
  msk ← 40h
  repeat
    unicode ← (unicode << 6) | (str.at((int32)idx) & 3Fh)
    msk ← msk << 5
    idx ← idx + 1
  until (unicode & msk = 0)
  unicode ← unicode & (msk - 1)
end if
Output unicode, idx
```

---

optimized version of Algorithm 2, enabling a very frequent operation: calculate the current unicode and advance the index to the next character.

## 4 The marking bit for the garbage collector

Among programming languages, the mark-and-sweep algorithm is still the most widely used for automatic memory management [3], [4], [5]. Without going into detail, this algorithm stops the program in progress, and performs two distinct phases:

1. Starting with the objects directly accessible at time  $t$  (stack, global variables) during program execution, the marking phase consists of recursively traversing these objects and marking them with a flag as being a live object.
2. In a second step, a linear memory run frees unmarked objects and cancels the marking of other objects.

After these two phases, the program can resume execution until the garbage collector is triggered again. Many variants of this algorithm exist, in particular to avoid program interruption, but the use of a marking bit is still required for its implementation.

The question for any language designer intending to implement this type of garbage collector is where to store this marking bit without polluting the structure of each allocated object. In many cases, this is not such a problem: adding an attribute within the object structure is easy to do. The waste-conscious designer may be sensitive to the systematic allocation of 64 bits for each object, when only one bit is necessary. But that's not the only reason to dwell on this issue. In low-level programming, it's not uncommon for a critical section of an application to be made up of objects with a frozen, hardware-imposed structure. Low-level programming is not incompatible with high-level object-oriented languages like `Lisaac` [6], [7]. If this is the case, there's no question of shamelessly adding a field to the structure for memory management.

In [8],[9] we see that it is possible to specialize the marking code according to the type of the object. This work focuses on how to recursively traverse the object graph. Here, our aim is to perform a specialization to change the value of this marking bit. With this specialization, we can find a specific place for each category or type of object.

Our idea is to find one or more unused bits in each structure representing a given type. In the following, we present several places where you can easily find this space.

### 4.1 Use an attribute that references another object

It's very common for the structure of an object to have attributes that are references to other objects. We therefore have the 16 most significant bits which are not used by the corresponding pointer. So, for example, the most significant bit can be used as the marking bit for garbage collection. Algorithm 6 shows how and where you can specialize the

---

**Algorithm 6** Setting the mark bit during the mark phase.

---

```
mark_type_name(obj) ←
ptr ← obj.attribute
if (ptr & 8000 0000 0000 0000h = 0) then
  obj.attribute ← ptr | 8000 0000 0000 0000h
  mark_attribute.type_name(ptr)
  ...Handling of other attributes...
end if
```

---

marking routine to further specialize the code in [8]. To avoid complicating the code unnecessarily, we’ve assumed that the attribute type is monomorphic. Of course, the most

---

**Algorithm 7** Clearing the mark bit during sweep phase.

---

```

sweep_type_name(obj) ←
if (obj.attribute & 8000 0000 0000 0000h) then
    obj.attribute ← obj.attribute & 7FFF FFFF FFFF FFFFh
else
    free(obj)
end if

```

---

significant bit of the pointer in question must be cleared during the sweep phase (see algorithm 7).

#### 4.2 Use the dynamic dispatch identification field

If an object’s type is subject to polymorphic calls, there is a hidden attribute at the beginning of its structure to identify its type. This field is necessary to handle dynamic dispatch in polymorphic calls. Without going into the details described in [10], this identifier is a simple integer with a maximum size directly related to the number of polymorphic types to be discriminated. In the optimization context described in [11], even if we globalize this identifier to the number of types possible in a given application, this number very rarely exceeds 8 bits. We therefore have all the space we need to

---

**Algorithm 8** The marking bit inside the dynamic dispatch information field.

---

```

ptr ← obj.attribute
id ← ptr.id
if (id & 8000 0000 0000 0000h = 0) then
    ptr.id ← id | 8000 0000 0000 0000h
    dispatch_and_mark(id, ptr)
    ...Handling of other attributes...
end if

```

---

install the marking bit at this point, always using specialized methods. However, it is important to note that it is necessary to take into account this bit potentially marked during the first phase of the recursive path (see algorithm 8).

#### 4.3 Use the re-alignment space in the data structure

Finally, if none of the above is possible, we can consider re-alignment zones. An attribute with a size of less than 64 bits will be compile-aligned to a 64-bit word. By adding a one-byte attribute at this precise location, the overall size of the structure is not increased at all. This byte can then be used as a marker bit.

In the example below, the two structures `obj_A` and `obj_B` are both 16 bytes long. Figure 3 illustrates the alignment and memory representation of the `obj_B` structure.

---

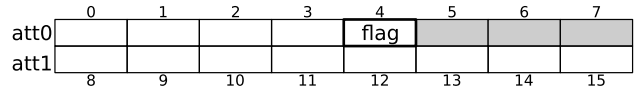
```

struct obj_A {
    int att0;
    long att1;
}

struct obj_B {
    int att0;
    char flag;
    long att1;
}

```

---



**Figure 3.** Memory representation of the `obj_B` structure.

As we’ve shown, it’s very often possible to find enough unused space. These free slots can be perfectly used for more complex variants of the *mark-and-sweep* requiring more than one marking bit. This technique is particularly well suited to garbage collectors whose code is specialized at compile time.

This free space can also be used for the implementation of monitors. Like Java’s synchronized blocks, for example. In fact, it’s usually easy to find 8 or even 16 bits within an object. This value can then be used as an index in a global table of pointers to the corresponding threads.

## 5 Conclusion

We’ve known since its emergence that 64-bit architectures tend to make executables and data structures mechanically larger than 32-bit architectures. Many bits are indeed wasted, often due to allocation alignments. But this wastage is also present in the fact that a pointer is limited to 48 bits, whereas it is represented in memory by a 64-bit word.

As part of the library implementation of a new language, we present here three very interesting uses for these wasted bits.

First, we give different physical representations of a multi-precision integer taking advantage of 64-bit architectures to gain performance in section 2.

In section 3, we look at a new approach to UTF-8 string manipulation by index virtualization using a 64-bit word.

Finally, section 4 shows how to optimize storage of the marking bit required in the implementation of the *mark & sweep* garbage collector algorithm.

## References

- [1] Michael D. Bond, Nicholas Nethercote, Stephen W. Kent, and Samuel Z. Guyer. Tracking bad apples: Reporting the origin of null and undefined value errors. In *In 22th Annual ACM Conference on Object-Oriented Programming Systems Languages and Applications (OOPSLA’2007)*, pages 405–422. ACM Press, 2007.
- [2] Adele Goldberg and David Robson. *Smalltalk-80: the language and its implementation*. Addison-Wesley, 1983. ISBN 0201113716, 1366 pages.
- [3] Paul R. Wilson. Uniprocessor Garbage Collector Techniques. International Workshop on Memory Management (IWMM’92), 1992.
- [4] Richard Jones and Rafael Lins. *Garbage Collection*. Wiley, 1996.
- [5] Richard Jones, Antony Hosking, and Eliot Moss. *The Garbage Collection Handbook: The Art of Automatic Memory Management*. Taylor & Francis, 2011.

## Tips for making the most of 64-bit architectures

- [6] Benoit Sonntag, Dominique Colnet, and Olivier Zendra. Dynamic inheritance: a powerful mechanism for operating system design. In *5th ECOOP Workshop on Object-Oriented and Operating Systems - ECOOP-OOOSWS'2002*, Lecture Notes in Computer Science, page 5 p, Malaga, Espagne, June 2002. Springer Verlag. Colloque avec actes et comité de lecture. internationale.
- [7] Benoit Sonntag and Dominique Colnet. Lisaac: The power of simplicity at work for operating systems. *International Conference on Technology of Object-Oriented Languages and Systems, Sydney, Australia, 2002*.
- [8] Dominique Colnet, Philippe Coucaud, and Olivier Zendra. Compiler Support to Customize the Mark & Sweep Algorithm. In *ACM SIGPLAN International Symposium on Memory Management - ISMM'98*, pages 154–165, Vancouver, British Columbia, Canada, 1998. ACM Special Interest Group on Programming Languages (SIGPLAN), ACM Press. Colloque avec actes et comité de lecture.
- [9] Dominique Colnet and Benoit Sonntag. Exploiting array manipulation habits to optimize garbage collection and type flow analysis. *Software: Practice and Experience*, pages 1639–1657, 2014.
- [10] Olivier Zendra, Dominique Colnet, and Suzanne Collin. Efficient Dynamic Dispatch without Virtual Function Tables. The SmallEiffel Compiler. In *In 12th Annual ACM Conference on Object-Oriented Programming Systems Languages and Applications (OOPSLA'97)*, pages 125–141. ACM Press, 1997.
- [11] Benoit Sonntag and Dominique Colnet. Efficient compilation strategy for object-oriented languages under the closed-world assumption. *Software: Practice and Experience*, 44(5):565–592, 2013.