# **Vectorization**

Write-up 1: Look at the assembly code above. The compiler has translated the code to set the start index at  $-2^{16}$  and adds to it for each memory access. Why doesn't it set the start index to 0 and use small positive offsets?

答:每一次循环,i都需要和SIZE (即2的16次方)作比较。先将内存置为-2的16次方,加上i后直接判断正负来比较大小,可以减少开销。

**Write-up 2:** This code is still not aligned when using AVX2 registers. Fix the code to make sure it uses aligned moves for the best performance.

答:修改代码如下,使程序使用32位对齐寄存器性能最优

```
#include <stdint.h>
#include <stdlib.h>
#include <math.h>

#define SIZE (1L << 32)

void test(uint8_t* restrict a, uint8_t* restrict b) {
    a = __builtin_assume_aligned(a, 32);
    b = __builtin_assume_aligned(b, 32);

for (uint64_t i = 0; i < SIZE; i++) {
    a[i] += b[i];
    }
}</pre>
```

**Write-up 3:** Provide a theory for why the compiler is generating dramatically different assembly.

答: 三元运算符不会按元素比较数组。

**Write-up 4:** Inspect the assembly and determine why the assembly does not include instructions with vector registers. Do you think it would be faster if it did vectorize? Explain.

# 运行生成的汇编代码如下:

```
.text
    .file
          "example3.c"
                                    # -- Begin function test
    .globl test
    .p2align 4, 0x90
    .type test,@function
                                        # @test
test:
.Lfunc_begin0:
    .file 1 "/home/zhangkeer/a4/recitation3" "example3.c"
           1 9 0
                                   # example3.c:9:0
    .cfi_startproc
# %bb.0:
    #DEBUG_VALUE: test:a <- $rdi</pre>
    #DEBUG_VALUE: test:a <- $rdi</pre>
    #DEBUG_VALUE: test:b <- $rsi</pre>
    #DEBUG_VALUE: test:b <- $rsi</pre>
    pushq %rax
    .cfi_def_cfa_offset 16
.Ltmp0:
    #DEBUG_VALUE: test:i <- 0</pre>
    .loc 1 12 3 prologue_end
                                   # example3.c:12:3
    addq $1, %rsi
.Ltmp1:
    .loc 1 13 10
                                   # example3.c:13:10
           $65536, %edx
                                   \# \text{ imm} = 0 \times 10000
    #DEBUG_VALUE: test:a <- $rdi</pre>
    callq memcpy
.Ltmp2:
    #DEBUG_VALUE: test:i <- undef</pre>
    .loc 1 15 1
                                   # example3.c:15:1
    popq
         %rax
    .cfi_def_cfa_offset 8
    retq
.Ltmp3:
.Lfunc_end0:
    .size
          test, .Lfunc_end0-test
    .cfi_endproc
                                        # -- End function
    .file 2 "/usr/include/x86_64-linux-gnu/bits" "types.h"
    .file 3 "/usr/include/x86_64-linux-gnu/bits" "stdint-uintn.h"
    .section
                .debug_str,"MS",@progbits,1
.Linfo_string0:
    .asciz "clang version 10.0.0-4ubuntu1 " # string offset=0
.Linfo_string1:
    .asciz "example3.c"
                                   # string offset=31
.Linfo_string2:
    .asciz "/home/zhangkeer/a4/recitation3" # string offset=42
.Linfo_string3:
    .asciz "test"
                                   # string offset=73
.Linfo_string4:
    .asciz "a"
                                    # string offset=78
.Linfo_string5:
    .asciz "unsigned char" # string offset=80
```

```
.Linfo_string6:
   .asciz "__uint8_t" # string offset=94
.Linfo_string7:
   .asciz "uint8_t"
                               # string offset=104
.Linfo_string8:
   .asciz "b"
                                # string offset=112
.Linfo_string9:
   .asciz "i"
                                # string offset=114
.Linfo_string10:
   .asciz "long unsigned int"
                                # string offset=116
.Linfo_string11:
   .asciz "__uint64_t" # string offset=134
.Linfo_string12:
   .asciz "uint64_t"
                                # string offset=145
   .section
              .debug_loc,"",@progbits
.Ldebug_loc0:
   .quad .Lfunc_begin0-.Lfunc_begin0
   .quad .Ltmp2-.Lfunc_begin0
   .short 1
                                # Loc expr size
   .byte 85
                                 # DW_OP_reg5
   .quad 0
   .quad 0
.Ldebug_loc1:
   .quad .Lfunc_begin0-.Lfunc_begin0
   .quad .Ltmp1-.Lfunc_begin0
   .short 1
                                 # Loc expr size
   .byte 84
                                 # DW_OP_reg4
   .quad 0
   .quad 0
.Ldebug_loc2:
   .quad .Ltmp0-.Lfunc_begin0
   .quad .Ltmp2-.Lfunc_begin0
   .short 2
                                # Loc expr size
   .byte 48
                                # DW_OP_lit0
   .byte 159
                                # DW_OP_stack_value
   .quad 0
   .quad
               .debug_abbrev,"",@progbits
   .section
                                 # Abbreviation Code
   .byte 1
   .byte 17
                                 # DW_TAG_compile_unit
   .byte 1
                                 # DW_CHILDREN_yes
   .byte
         37
                                 # DW_AT_producer
   .byte 14
                                 # DW_FORM_strp
   .byte 19
                                 # DW_AT_language
   .byte 5
                                 # DW_FORM_data2
   .byte 3
                                 # DW_AT_name
   .byte
         14
                                 # DW_FORM_strp
   .byte
         16
                                 # DW_AT_stmt_list
   .byte
         23
                                 # DW_FORM_sec_offset
   .byte
         27
                                 # DW_AT_comp_dir
   .byte
         14
                                 # DW_FORM_strp
   .byte
          17
                                 # DW_AT_low_pc
   .byte 1
                                 # DW_FORM_addr
   .byte 18
                                 # DW_AT_high_pc
   .byte 6
                                 # DW_FORM_data4
   .byte 0
                                 # EOM(1)
   .byte
          0
                                 # EOM(2)
                                 # Abbreviation Code
   .byte
```

```
.byte
        46
                                 # DW_TAG_subprogram
                                 # DW_CHILDREN_yes
.byte
        1
        17
                                 # DW_AT_low_pc
.byte
.byte
        1
                                 # DW_FORM_addr
.byte
        18
                                 # DW_AT_high_pc
        6
.byte
                                 # DW_FORM_data4
.byte
        64
                                 # DW_AT_frame_base
.byte
        24
                                 # DW_FORM_exprloc
       "\227B"
                                 # DW_AT_GNU_all_call_sites
.ascii
.byte
        25
                                 # DW_FORM_flag_present
.byte
                                 # DW_AT_name
        3
.byte
        14
                                 # DW_FORM_strp
.byte
        58
                                 # DW_AT_decl_file
        11
                                 # DW_FORM_data1
.byte
.byte
        59
                                 # DW_AT_decl_line
.byte
        11
                                 # DW_FORM_data1
        39
                                 # DW_AT_prototyped
.byte
.byte
        25
                                 # DW_FORM_flag_present
                                 # DW_AT_external
.byte
        63
.byte
        25
                                 # DW_FORM_flag_present
                                 # EOM(1)
.byte
        0
        0
                                 # EOM(2)
.byte
.byte
        3
                                 # Abbreviation Code
       5
.byte
                                 # DW_TAG_formal_parameter
.byte
        0
                                 # DW_CHILDREN_no
        2
.byte
                                 # DW_AT_location
.byte
        23
                                 # DW_FORM_sec_offset
        3
                                 # DW_AT_name
.byte
.byte
        14
                                 # DW_FORM_strp
.byte
        58
                                 # DW_AT_decl_file
.byte
        11
                                 # DW_FORM_data1
.byte
        59
                                 # DW_AT_decl_line
.byte
        11
                                 # DW_FORM_data1
.byte
        73
                                 # DW_AT_type
.byte
        19
                                 # DW_FORM_ref4
.byte
        0
                                 # EOM(1)
        0
                                 # EOM(2)
.byte
        4
.byte
                                 # Abbreviation Code
.byte
        52
                                 # DW_TAG_variable
.byte
        0
                                 # DW_CHILDREN_no
        2
                                 # DW_AT_location
.byte
.byte
        23
                                 # DW_FORM_sec_offset
.byte
        3
                                 # DW_AT_name
.byte
        14
                                 # DW_FORM_strp
.byte
        58
                                 # DW_AT_decl_file
                                 # DW_FORM_data1
.byte
        11
.byte
        59
                                 # DW_AT_decl_line
.byte
        11
                                 # DW_FORM_data1
        73
.byte
                                 # DW_AT_type
                                 # DW_FORM_ref4
.byte
        19
                                 # EOM(1)
.byte
        0
.byte
        0
                                 # EOM(2)
        5
.byte
                                 # Abbreviation Code
.byte
        55
                                 # DW_TAG_restrict_type
.byte
        0
                                 # DW_CHILDREN_no
.byte
        73
                                 # DW_AT_type
.byte
        19
                                 # DW_FORM_ref4
.byte
                                 \# EOM(1)
```

```
.byte
           0
                                   # EOM(2)
    .byte
           6
                                   # Abbreviation Code
    .byte
           15
                                   # DW_TAG_pointer_type
   .byte
           0
                                   # DW_CHILDREN_no
    .byte
          73
                                   # DW_AT_type
   .byte
          19
                                   # DW_FORM_ref4
   .byte
           0
                                   # EOM(1)
   .byte
           0
                                   # EOM(2)
          7
                                   # Abbreviation Code
   .byte
    .byte
          22
                                   # DW_TAG_typedef
   .byte
                                   # DW_CHILDREN_no
          0
   .byte
          73
                                   # DW_AT_type
   .byte
           19
                                   # DW_FORM_ref4
   .byte
          3
                                   # DW_AT_name
   .byte
           14
                                   # DW_FORM_strp
          58
   .byte
                                   # DW_AT_decl_file
           11
                                   # DW_FORM_data1
   .byte
   .byte
           59
                                   # DW_AT_decl_line
   .byte
          11
                                   # DW_FORM_data1
   .byte
                                   # EOM(1)
          0
                                   # EOM(2)
   .byte
          8
                                   # Abbreviation Code
   .byte
   .byte
          36
                                   # DW_TAG_base_type
   .byte
         0
                                   # DW_CHILDREN_no
   .byte
          3
                                   # DW_AT_name
   .byte
          14
                                   # DW_FORM_strp
                                   # DW_AT_encoding
   .byte
          62
   .byte 11
                                   # DW_FORM_data1
   .byte 11
                                   # DW_AT_byte_size
          11
   .byte
                                   # DW_FORM_data1
   .byte 0
                                   # EOM(1)
   .byte
           0
                                   # EOM(2)
   .byte
           0
                                   # EOM(3)
               .debug_info,"",@progbits
   .section
.Lcu_begin0:
    .long
           .Ldebug_info_end0-.Ldebug_info_start0 # Length of Unit
.Ldebug_info_start0:
                                  # DWARF version number
    .short 4
                                  # Offset Into Abbrev. Section
   .long
           .debug_abbrev
   .byte
                                  # Address Size (in bytes)
                                  # Abbrev [1] 0xb:0xa7 DW_TAG_compile_unit
   .byte
          1
                                 # DW_AT_producer
   .long
           .Linfo_string0
   .short 12
                                  # DW_AT_language
   .long
          .Linfo_string1
                                  # DW_AT_name
   .long .Lline_table_start0 # DW_AT_stmt_list
          .Linfo_string2
   .long
                                  # DW_AT_comp_dir
   .quad
           .Lfunc_begin0
                                   # DW_AT_low_pc
   .long
           .Lfunc_end0-.Lfunc_begin0 # DW_AT_high_pc
   .byte
                                   # Abbrev [2] 0x2a:0x43 DW_TAG_subprogram
   .quad
          .Lfunc_begin0
                                   # DW_AT_low_pc
           .Lfunc_end0-.Lfunc_begin0 # DW_AT_high_pc
   .long
   .byte
           1
                                   # DW_AT_frame_base
   .byte
           87
                                       # DW_AT_GNU_all_call_sites
    .long
           .Linfo_string3
                                   # DW_AT_name
    .byte
           1
                                   # DW_AT_decl_file
    .byte
           9
                                   # DW_AT_decl_line
                                       # DW_AT_prototyped
```

```
# DW_AT_external
   .byte
          3
                                # Abbrev [3] 0x3f:0xf
DW_TAG_formal_parameter
   .long
         .Ldebug_loc0
                                # DW_AT_location
   .long
          .Linfo_string4
                                # DW_AT_name
   .byte 1
                                # DW_AT_decl_file
   .byte
         9
                                # DW_AT_decl_line
         109
   .long
                                # DW_AT_type
   .byte 3
                                # Abbrev [3] 0x4e:0xf
DW_TAG_formal_parameter
   .long .Ldebug_loc1
                                # DW_AT_location
   .long .Linfo_string8
                               # DW_AT_name
   .byte
           1
                                # DW_AT_decl_file
   .byte 9
                                # DW_AT_decl_line
   .long
         109
                                # DW_AT_type
                               # Abbrev [4] 0x5d:0xf DW_TAG_variable
   .byte 4
                            # DW_AT_location
# DW_AT_name
         .Ldebug_loc2
   .long
   .long .Linfo_string9
   .byte 1
                                # DW_AT_decl_file
   .byte
         10
                                 # DW_AT_decl_line
   .long 148
                                # DW_AT_type
         0
                                # End Of Children Mark
   .byte
   .byte 5
                                # Abbrev [5] 0x6d:0x5 DW_TAG_restrict_type
   .long 114
                                # DW_AT_type
   .byte
                                # Abbrev [6] 0x72:0x5 DW_TAG_pointer_type
   .long
         119
                                # DW_AT_type
         7
   .byte
                                # Abbrev [7] 0x77:0xb DW_TAG_typedef
   .long 130
                                # DW_AT_type
   .long .Linfo_string7
                                # DW_AT_name
   .byte
                                # DW_AT_decl_file
   .byte 24
                               # DW_AT_decl_line
   .byte
         7
                                # Abbrev [7] 0x82:0xb DW_TAG_typedef
   .long
         141
                                # DW_AT_type
   .long .Linfo_string6
                                # DW_AT_name
   .byte
                                # DW_AT_decl_file
   .byte
         38
                                # DW_AT_decl_line
                                # Abbrev [8] 0x8d:0x7 DW_TAG_base_type
   .byte
                             # DW_AT_name
   .long
         .Linfo_string5
   .byte
                                # DW_AT_encoding
   .byte
         1
                                # DW_AT_byte_size
         7
                                # Abbrev [7] 0x94:0xb DW_TAG_typedef
   .byte
   .long
          159
                                # DW_AT_type
                              # DW_AT_name
   .long
         .Linfo_string12
   .byte
          3
                                # DW_AT_decl_file
   .byte
          27
                                # DW_AT_decl_line
   .byte
         7
                                # Abbrev [7] 0x9f:0xb DW_TAG_typedef
   .long
          170
                                # DW_AT_type
   .long
         .Linfo_string11
                               # DW_AT_name
   .byte
          2
                                # DW_AT_decl_file
   .byte
         45
                                # DW_AT_decl_line
                                # Abbrev [8] 0xaa:0x7 DW_TAG_base_type
   .byte
         .Linfo_string10
   .long
                             # DW_AT_name
    .byte
         7
                                # DW_AT_encoding
    .byte
          8
                                # DW_AT_byte_size
                                 # End Of Children Mark
    .byte
           0
.Ldebug_info_end0:
    .ident "clang version 10.0.0-4ubuntu1"
    .section ".note.GNU-stack","",@progbits
```

```
.addrsig
  .section   .debug_line,"",@progbits
.Lline_table_start0:
```

内存未对齐。可以将所有元素移动一位来重新对齐b数组,或者使用偏移量。向量化后程序运行会更快。

**Write-up 5:** Check the assembly and verify that it does in fact vectorize properly. Also what do you notice when you run the command

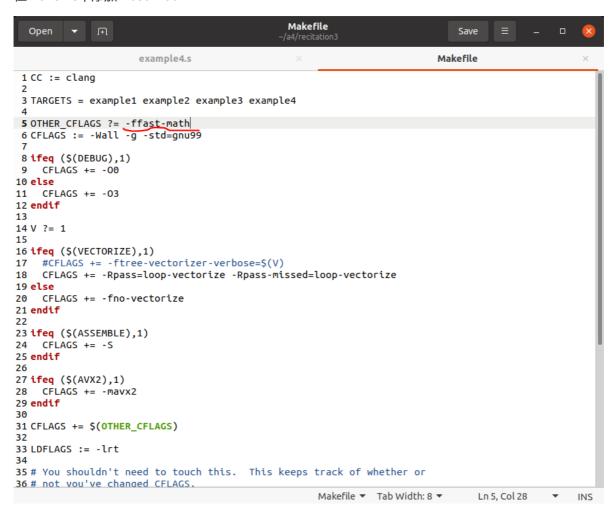
```
$ clang -03 example4.c -o example4; ./example4
```

with and without the -ffast-math flag? Specifically, why do you a see a difference in the output.

## 答:

对于汇编代码,材料已列举使用addsd命令的非向量化代码。

在makefile中添加-ffast-math



重新编译,可以看到确实应用了向量化:

```
F
                          zhangkeer@ubuntu: ~/a4/recitation3
                                                            Q
zhangkeer@ubuntu:~/a4/recitation3$ make clean; make ASSEMBLE=1 VECTORIZE=1 examp
le4.o
rm -rf
rm -f *.o *.s .cflags perf.data *.lst
clang -Wall -g -std=gnu99 -03 -Rpass=loop-vectorize -Rpass-missed=loop-vectorize
  -S -ffast-math -c example4.c
example4.c:17:3: remark: vectorized loop (vectorization width: 2, interleaved co
unt: 2) [-Rpass=loop-vectorize]
 for (i = 0; i < SIZE; i++) {
example4.c:25:3: remark: the cost-model indicates that vectorization is not bene
ficial [-Rpass-missed=loop-vectorize]
  for (int i = 0; i < SIZE; i++) {
example4.c:25:3: remark: the cost-model indicates that interleaving is not benef
icial [-Rpass-missed=loop-vectorize]
example4.c:17:3: remark: vectorized loop (vectorization width: 2, interleaved co
unt: 2) [-Rpass=loop-vectorize]
 for (i = 0; i < SIZE; i++) {
zhangkeer@ubuntu:~/a4/recitation3$
```

汇编代码中的命令也由addsd变成了addpd:

```
example4.s
         Save
 Open
 1
         .text
                    "example4.c"
 2
            .file
            .globl test
                                               # -- Begin function test
 3
 4
            .p2align
                             4. 0x90
 5
           .type test,@function
 6 test:
                                                # @test
 7 .Lfunc_begin0:
           .file 1 "/home/zhangkeer/a4/recitation3" "example4.c"
.loc 1 10 0 # example4.c:10:0
 9
10
           .cfi_startproc
11 # %bb.0:
12
           #DEBUG_VALUE: test:a <- $rdi</pre>
           #DEBUG_VALUE: test:a <- $rdi</pre>
13
           xorpd %xmm0, %xmm0
14
15
          xorl
                   %eax, %eax
16 .Ltmp0:
           #DEBUG_VALUE: test:i <- 0</pre>
17
18
           #DEBUG_VALUE: test:y <- 0.000000e+00
19
           #DEBUG_VALUE: test:x <- undef</pre>
           xorpd %xmm1, %xmm1
20
21 .Ltmp1:
           .p2align
                            4. 0x90
22
23 .LBB0 1:
                                                # =>This Inner Loop Header: Depth=1
           #DEBUG_VALUE: test:y <- 0.000000e+00
24
25
           #DEBUG_VALUE: test:i <- 0</pre>
26
           #DEBUG_VALUE: test:a <- $rdi</pre>
                  1 18 7 prologue_end
                                               # example4.c:18:7
27
           .loc
           addpd (%rdi,%rax,8), %xmm0 
16(%rdi,%rax,8), %xmm1 
32(%rdi,%rax,8), %xmm0 
48(%rdi,%rax,8), %xmm1
28
29
30
31
32
           addpd
                   64(%rdi,%rax,8), %xmm0
33
           addpd
                   80(%rdi,%rax,8), %xmm1
           addpd
                    96(%rdi,%rax,8), %xmm0
34
35
           addpd
                   112(%rdi,%rax,8), %xmm1
36 .Ltmp2:
37
           .loc
                    1 17 26
                                               # example4.c:17:26
                                                       Plain Text ▼ Tab Width: 8 ▼ Ln 1, Col 1 ▼ INS
```

不添加编译时标志 -ffast-math

```
clang -03 example4.c -o example4; ./example4
```

## 输出结果为:

```
zhangkeer@ubuntu:~/a4/recitation3$ clang -03 example4.c -o example4; ./example4
The decimal floating point sum result is: 11.667578
The raw floating point sum result is: 0x1.755cccec10aa5p+3
zhangkeer@ubuntu:~/a4/recitation3$
```

添加编译时标志 -ffast-math

```
clang -03 -ffast-math example4.c -o example4; ./example4
```

#### 输出结果为:

```
zhangkeer@ubuntu:~/a4/recitation3$ clang -03 -ffast-math example4.c -o example4
; ./example4
The decimal floating point sum result is: 11.667578
The raw floating point sum result is: 0x1.755cccec10aa3p+3
zhangkeer@ubuntu:~/a4/recitation3$
```

输出结果不同,因为我们设置标志位允许了 clang 重新排序我们给它的操作。

Write-up 6: What speedup does the vectorized code achieve over the unvectorized code? What additional speedup does using -mavx2 give? You may wish to run this experiment several times and take median elapsed times; you can report answers to the nearest 100% (e.g.,  $2\times$ ,  $3\times$ , etc). What can you infer about the bit width of the default vector registers on the awsrun machines? What about the bit width of the AVX2 vector registers? *Hint*: aside from speedup and the vectorization report, the most relevant information is that the data type for each array is uint32\_t.

#### Without vectorization:

```
$ make
$ time ./loop
```

## 运行结果:

# 修改1的值,多次运行,使结果更有代表性:

## With vectorization:

```
$ make clean
$ make VECTORIZE=1
$ time ./loop
```

## 运行结果:

```
zhangkeer@ubuntu:~/a4/homework3$ make clean
rm -f loop *.o *.s .cflags perf.data */perf.data
zhangkeer@ubuntu:~/a4/homework3$ make VECTORIZE=1
clang -Wall -std=gnu99 -g -O3 -DNDEBUG -Rpass=loop-vectorize -Rpass-missed=loop-
vectorize -ffast-math -c loop.c
loop.c:70:9: remark: vectorized loop (vectorization width: 4, interleaved count:
2) [-Rpass=loop-vectorize]
        for (j = 0; j < N; j++) {
clang -o loop loop.o -lrt
zhangkeer@ubuntu:~/a4/homework3$ time ./loop
Elapsed execution time: 0.021711 sec; N: 1024, I: 100000, __OP__: +, __TYPE__: u
int32_t
real
        0m0.024s
        0m0.024s
user
        0m0.000s
sys
zhangkeer@ubuntu:~/a4/homework3$
```

# 修改|的值,多次运行,使结果更有代表性:

# With AVX instructions:

```
$ make clean
$ make VECTORIZE=1 AVX2=1
$ time ./loop
```

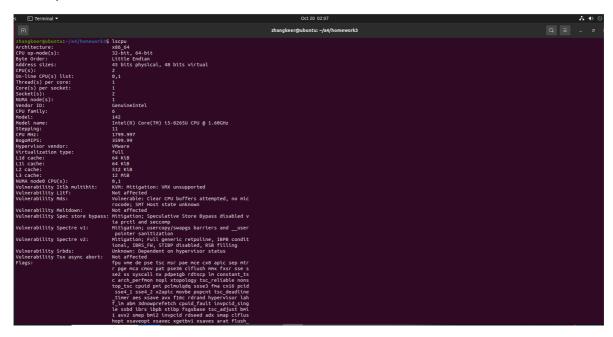
## 运行结果:

```
zhangkeer@ubuntu:~/a4/homework3$ make clean
rm -f loop *.o *.s .cflags perf.data */perf.data
zhangkeer@ubuntu:~/a4/homework3$ make VECTORIZE=1 AVX2=1
clang -Wall -std=gnu99 -g -O3 -DNDEBUG -Rpass=loop-vectorize -Rpass-missed=loop-
vectorize -ffast-math -mavx2 -c loop.c
loop.c:70:9: remark: vectorized loop (vectorization width: 8, interleaved count:
4) [-Rpass=loop-vectorize]
        for (j = 0; j < N; j++) {
clang -o loop loop.o -lrt
zhangkeer@ubuntu:~/a4/homework3$ time ./loop
Elapsed execution time: 0.006121 sec; N: 1024, I: 100000, __OP__: +, __TYPE__: u
int32_t
real
        0m0.008s
user
        0m0.004s
sys
        0m0.004s
zhangkeer@ubuntu:~/a4/homework3$
```

修改|的值,多次运行,使结果更有代表性:

```
zhangkeer@ubuntu:~/a4/homework3$ time ./loop
Elapsed execution time: 0.597142 sec; N: 1024, I: 10000000, __OP__: +, __TYPE__: uint32_t
real    0m0.599s
user    0m0.594s
sys    0m0.005s
zhangkeer@ubuntu:~/a4/homework3$
```

# 执行Iscpu的结果是:



综合上面的结果,向量化代码将非向量化代码的运行速度提升了约四倍,使用AVX2向量寄存器又在向量化代码的基础上加速了两倍左右。

由于每个数组元素的数据类型是uint32\_t, 默认向量寄存器的位宽应该是4×32=128 (性能提升四倍), AVX2 向量寄存器的位宽应该是2×128=256 (性能又提升两倍)。