Vectorizing with RVV

Challenges

Introduction and Rules

- We have prepared 10 vectorization Challenges of varying difficulty.
- The challenges self-check the validity of the result, but not the vectorization efficiency.

- You have three levels of help for the challenges:
  - **Hints**: e.g. “Look at the function parameters”
  - **Target metrics:** e.g. “The VectorMix should increase to X”
  - **Solutions:** A proposed solution for the challenge

- ⚠️ Solutions might not be the optimal, you are invited to find better solutions

- When you complete a challenge, you can call a teacher or compare it to the solution.

Challenges Links

Challenge #1 (SpMV): https://repo.hca.bsc.es/epic/z/rRFc_I

Challenge #2 (Pressures): https://repo.hca.bsc.es/epic/z/pdRbXR

Challenge #3 (Transpose): https://repo.hca.bsc.es/epic/z/XMU1IO

Challenge #4 (Axpy_2d): https://repo.hca.bsc.es/epic/z/dAJaaY

Challenge #5 (Complex): https://repo.hca.bsc.es/epic/z/Clgx5_

Challenge #6 (start_end): https://repo.hca.bsc.es/epic/z/zNT-ji

Challenge #7 (3D): https://repo.hca.bsc.es/epic/z/A6pROp

Challenge #8 (halfwork): https://repo.hca.bsc.es/epic/z/0CYa6i

Challenge #9 (vecclass): https://repo.hca.bsc.es/epic/z/rkVhVz

Challenge #10 (intrinsics): https://repo.hca.bsc.es/epic/z/XdxHtH

Help for Challenge #1 (SpMV): Hint

- Sometimes the compiler needs a little help from pragmas…

Help for Challenge #1 (SpMV): Target Metrics

```
Region #1: Event 1000 (code_region), Value 1 (SpMV),
    Moved bytes (Total): 3163138
        Moved bytes (scalar): 13314 (0.42 %)
        Moved bytes (vector): 3149824 (99.58 %)
    tot_instr: 22261
        scalar_instr: 13058 (58.66 %)
        vsetvl_instr: 3069 (13.79 %)
        vector_instr: 6134 (27.55 %)
            SEW 8 vector_instr: 512 (8.35 %)
                avg_VL: 1.00 elements
                Arith: 512 (100.00 %)
                    FP: 0 (0.00 %)
                    INT: 512 (100.00 %)
                Mem: 0 (0.00 %)
                    unit: 0 (0.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 5622 (91.65 %)
                avg_VL: 163.38 elements
                Arith: 2810 (49.98 %)
                    FP: 1278 (45.48 %)
                    INT: 1532 (54.52 %)
                Mem: 2810 (49.98 %)
                    unit: 2044 (72.74 %)
                    strided: 0 (0.00 %)
                    indexed: 766 (27.26 %)
                Mask: 0 (0.00 %)
                Other: 2 (0.04 %)
-----------------------------------------------
```

Help for Challenge #1 (SpMV): Solution

- Solution: https://repo.hca.bsc.es/epic/z/bTT1hq

Help for Challenge #2 (Pressures): Hints

- The cubeRoot function **won’t be vectorized**
- Do you need to compute everything in the same loop? Can work be distributed into a “non-vec” loop and a vectorized one?

Help for Challenge #2 (Pressures): Metrics

```
Region #1: Event 1000 (code_region), Value 1 (Pressure),
    Moved bytes (Total): 168002
        Moved bytes (scalar): 36930 (21.98 %)
        Moved bytes (vector): 131072 (78.02 %)
    tot_instr: 41227
        scalar_instr: 41099 (99.69 %)
        vsetvl_instr: 16 (0.04 %)
        vector_instr: 112 (0.27 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 112 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 48 (42.86 %)
                    FP: 48 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 64 (57.14 %)
                    unit: 64 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
-----------------------------------------------
```

Help for Challenge #2 (Pressures): Solution

- Solution: https://repo.hca.bsc.es/epic/z/tGw2Az

Help for Challenge #3 (Transpose): Hints

- Is the Average VL good?
- The inner-most loop is the vectorized one
- What are the sizes of N and M?

Help for Challenge #3 (Transpose): Metrics

```
Region #1: Event 1000 (code_region), Value 1 (Transpose),
    Moved bytes (Total): 524684
        Moved bytes (scalar): 396 (0.08 %)
        Moved bytes (vector): 524288 (99.92 %)
    tot_instr: 1601
        scalar_instr: 1089 (68.02 %)
        vsetvl_instr: 128 (8.00 %)
        vector_instr: 384 (23.99 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 384 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 128 (33.33 %)
                    FP: 128 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 256 (66.67 %)
                    unit: 128 (50.00 %)
                    strided: 128 (50.00 %)
                        Avg. Stride (B): 64.00
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
-----------------------------------------------
```

Help for Challenge #3 (Transpose): Solution

- Solution: https://repo.hca.bsc.es/epic/z/JIfFId

Help for Challenge #4 (Axpy_2d): Hints

- Is the Average VL good?
- You can convert a 2D array to a 1D pointer like this:
  - double * X_ptr = &X[0][0];

Help for Challenge #4 (Axpy_2d): Metrics

```
Region #1: Event 1000 (code_region), Value 1 (Transpose),
    Moved bytes (Total): 98354
        Moved bytes (scalar): 50 (0.05 %)
        Moved bytes (vector): 98304 (99.95 %)
    tot_instr: 185
        scalar_instr: 105 (56.76 %)
        vsetvl_instr: 16 (8.65 %)
        vector_instr: 64 (34.59 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 64 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 16 (25.00 %)
                    FP: 16 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 48 (75.00 %)
                    unit: 48 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
-----------------------------------------------
```

Help for Challenge #4 (Axpy_2d): Solution

- Solution: https://repo.hca.bsc.es/epic/z/5KBvw1

Help for Challenge #5 (Complex): Hints

- Avoid using strided loads
- Is an array of structures the best way to vectorize a code?

Help for Challenge #5 (Complex): Metrics

```
Region #1: Event 1000 (code_region), Value 1 (StartEnd),
    Moved bytes (Total): 196640
        Moved bytes (scalar): 32 (0.02 %)
        Moved bytes (vector): 196608 (99.98 %)
    tot_instr: 314
        scalar_instr: 136 (43.31 %)
        vsetvl_instr: 17 (5.41 %)
        vector_instr: 161 (51.27 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 161 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 64 (39.75 %)
                    FP: 64 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 96 (59.63 %)
                    unit: 96 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 1 (0.62 %)
-----------------------------------------------
```

Help for Challenge #5 (Complex): Solution

- Solution: https://repo.hca.bsc.es/epic/z/ZcktWe

Help for Challenge #6 (start_end): Hints

- There are many “Mask” and “Other” instructions
- Does R[0] and R[1] need to be set within the loop?

Help for Challenge #6 (start_end): Metrics

```
Region #1: Event 1000 (code_region), Value 1 (StartEnd),
    Moved bytes (Total): 98354
        Moved bytes (scalar): 50 (0.05 %)
        Moved bytes (vector): 98304 (99.95 %)
    tot_instr: 229
        scalar_instr: 132 (57.64 %)
        vsetvl_instr: 17 (7.42 %)
        vector_instr: 80 (34.93 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 80 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 32 (40.00 %)
                    FP: 32 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 48 (60.00 %)
                    unit: 48 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
-----------------------------------------------
```

Help for Challenge #6 (start_end): Solution

- Solution: https://repo.hca.bsc.es/epic/z/JKtaiV

Help for Challenge #7 (3D): Hints

- The Average VL is low
- There are strided operations
- Is the loop order the best it can be?

Help for Challenge #7 (3D): Metrics

```
Region #1: Event 1000 (code_region), Value 1 (foo),
    Moved bytes (Total): 525150
        Moved bytes (scalar): 606 (0.12 %)
        Moved bytes (vector): 524544 (99.88 %)
    tot_instr: 2526
        scalar_instr: 1147 (45.41 %)
        vsetvl_instr: 353 (13.97 %)
        vector_instr: 1026 (40.62 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 1026 (100.00 %)
                avg_VL: 232.14 elements
                Arith: 192 (18.71 %)
                    FP: 160 (83.33 %)
                    INT: 32 (16.67 %)
                Mem: 288 (28.07 %)
                    unit: 288 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 256 (24.95 %)
                Other: 290 (28.27 %)
-------------------------------------------------
```

Help for Challenge #7 (3D): Solution

Solution: https://repo.hca.bsc.es/epic/z/Yp-VCZ

Help for Challenge #8 (halfwork): Hints

- A pragma might help the compiler
- Once you vectorize the code, can you remove the mask instructions?
- Separating the work into two loops will help

Help for Challenge #8 (halfwork): Metrics

```
Region #1: Event 1000 (code_region), Value 1 (Halfwork),
    Moved bytes (Total): 28711
        Moved bytes (scalar): 39 (0.14 %)
        Moved bytes (vector): 28672 (99.86 %)
    tot_instr: 87
        scalar_instr: 58 (66.67 %)
        vsetvl_instr: 7 (8.05 %)
        vector_instr: 22 (25.29 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 22 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 8 (36.36 %)
                    FP: 8 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 14 (63.64 %)
                    unit: 10 (71.43 %)
                    strided: 4 (28.57 %)
                        Avg. Stride (B): 16.00
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
```

Help for Challenge #8 (halfwork)

Solution: https://repo.hca.bsc.es/epic/z/wCrarf

Help for Challenge #9 (vecclass): Hints

This code has two optimization routes:

Low vector mix:

- The vector mix is very low, but all operations are vectorize
- There are too many scalar instructions appearing “out of nowhere”
- Maybe they are copies? How can you avoid them?

Fusing loops:

- Does it make sense to have separate loops for each vector operation?

Help for Challenge #9 (vecclass): Metrics

Solution 1

```
Region #1: Event 1000 (code_region), Value 1 (Vector_Ops),
    Moved bytes (Total): 196696
        Moved bytes (scalar): 88 (0.04 %)
        Moved bytes (vector): 196608 (99.96 %)
    tot_instr: 470
        scalar_instr: 310 (65.96 %)
        vsetvl_instr: 32 (6.81 %)
        vector_instr: 128 (27.23 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 128 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 32 (25.00 %)
                    FP: 32 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 96 (75.00 %)
                    unit: 96 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
```

Solution 2

```
Region #1: Event 1000 (code_region), Value 1 (Vector_Ops),
    Moved bytes (Total): 65552
        Moved bytes (scalar): 16 (0.02 %)
        Moved bytes (vector): 65536 (99.98 %)
    tot_instr: 130
        scalar_instr: 74 (56.92 %)
        vsetvl_instr: 8 (6.15 %)
        vector_instr: 48 (36.92 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 48 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 16 (33.33 %)
                    FP: 16 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 32 (66.67 %)
                    unit: 32 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 0 (0.00 %)
```

Help for Challenge #9 (vecclass): Solution

- Solution 1: https://repo.hca.bsc.es/epic/z/VDEAw6
- Solution 2: https://repo.hca.bsc.es/epic/z/hXS14R

Help for Challenge #10 (intrinsics) Hints:

- Only the inner-most loop is vectorized
- This generates a reduction instruction
- Can you use intrinsics to vectorize the outer loop?

Help for Challenge #10 (intrinsics): Metrics

```
Region #1: Event 1000 (code_region), Value 2 (Reduction_outer
    Moved bytes (Total): 14345
        Moved bytes (scalar): 2057 (14.34 %)
        Moved bytes (vector): 12288 (85.66 %)
    tot_instr: 5169
        scalar_instr: 3103 (60.03 %)
        vsetvl_instr: 2 (0.04 %)
        vector_instr: 2064 (39.93 %)
            SEW 8 vector_instr: 0 (0.00 %)
            SEW 16 vector_instr: 0 (0.00 %)
            SEW 32 vector_instr: 0 (0.00 %)
            SEW 64 vector_instr: 2064 (100.00 %)
                avg_VL: 256.00 elements
                Arith: 2056 (99.61 %)
                    FP: 2056 (100.00 %)
                    INT: 0 (0.00 %)
                Mem: 6 (0.29 %)
                    unit: 6 (100.00 %)
                    strided: 0 (0.00 %)
                    indexed: 0 (0.00 %)
                Mask: 0 (0.00 %)
                Other: 2 (0.10 %)
-----------------------------------------------
```

Help for Challenge #10 (intrinsics)

- Solution: https://repo.hca.bsc.es/epic/z/NFIf1F