# Intel<sup>®</sup> oneAPI VTune<sup>™</sup> Profiler 2021.1.1 Gold

**Elapsed Time:** 0.045s

Application execution time is too short. Metrics data may be unreliable. Consider reducing the sampling interval or increasing your application execution time.

**Clockticks:** 153,360,000 **Instructions Retired:** 140,220,000

**CPI Rate:** 1.094

The CPI may be too high. This could be caused by issues such as memory stalls, instruction starvation, branch misprediction or long latency instructions. Explore the other hardware-related metrics to identify what is causing high CPI.

MUX Reliability: 0.986

**Retiring:** 24.9% of Pipeline Slots 26.9% of Pipeline Slots

 FP Arithmetic:
 0.0% of uOps

 FP x87:
 0.0% of uOps

 FP Scalar:
 0.0% of uOps

 FP Vector:
 0.0% of uOps

 Other:
 100.0% of uOps

**Heavy Operations:** 0.0% of Pipeline Slots **Microcode Sequencer:** 7.4% of Pipeline Slots **Assists:** 0.0% of Pipeline Slots

**Front-End Bound:** 21.7% of Pipeline Slots

Issue: A significant portion of Pipeline Slots is remaining empty due to issues in the Front-End.

Tips: Make sure the code working size is not too large, the code layout does not require too many memory accesses per cycle to get enough instructions for filling four pipeline slots, or check for microcode assists.

# **Front-End Latency:** 24.9% of Pipeline Slots

This metric represents a fraction of slots during which CPU was stalled due to front-end latency issues, such as instruction-cache misses, ITLB misses or fetch stalls after a branch misprediction. In such cases, the front-end delivers no uOps.

0.0% of Clockticks ICache Misses: ITLB Overhead: 1.1% of Clockticks 0.0% of Clockticks **Branch Resteers: Mispredicts Resteers:** 0.0% of Clockticks **Clears Resteers:** 0.0% of Clockticks 0.0% of Clockticks Unknown Branches: 0.0% of Clockticks DSB Switches: **Length Changing Prefixes:** 0.0% of Clockticks MS Switches: 0.0% of Clockticks

Issue: A significant fraction of cycles was stalled due to switches of uOp delivery to the Microcode Sequencer (MS). Commonly used instructions are optimized for delivery by the DSB or MITE pipelines. Certain operations cannot be handled natively by the execution pipeline, and must be performed by microcode (small programs injected into the execution stream). Switching to the MS too often can negatively impact performance. The MS is designated to deliver long uOp flows required by CISC instructions like CPUID, or uncommon conditions like Floating Point Assists when dealing with Denormals. Note that this metric value may be highlighted due to Microcode Sequencer issue.

Front-End Bandwidth: 0.0% of Pipeline Slots Front-End Bandwidth MITE: 20.7% of Clockticks Front-End Bandwidth DSB: 4.1% of Clockticks

(Info) DSB Coverage: 31.6%

Branch Mispredict:
Machine Clears:

3.1% of Pipeline Slots
0.0% of Pipeline Slots
3.1% of Pipeline Slots **Bad Speculation:** 0.0% of Pipeline Slots 3.1% of Pipeline Slots **Back-End Bound:** 50.3% of Pipeline Slots

A significant portion of pipeline slots are remaining empty. When operations take too long in the back-end, they introduce bubbles in the pipeline that ultimately cause fewer pipeline slots containing useful work to be retired per cycle than the machine is capable to support. This opportunity cost results in slower execution. Long-latency operations like divides and memory operations can cause this, as can too many operations being directed to a single execution port (for example, more multiply operations arriving in the back-end per cycle than the execution unit can support).

Memory Bound: 33.5% of Pipeline Slots The metric value is high. This can indicate that the significant fraction of execution pipeline slots could be stalled due to demand memory load and stores. Use Memory Access analysis to have the metric breakdown by memory hierarchy, memory bandwidth information, correlation by memory objects.

#### **L1 Bound:** 14.1% of Clockticks

This metric shows how often machine was stalled without missing the L1 data cache. The L1 cache typically has the shortest latency. However, in certain cases like loads blocked on older stores, a load might suffer a high latency even though it is being satisfied by the L1. Note that this metric value may be highlighted due to DTLB Overhead or Cycles of 1 Port Utilized issues.

DTLB Overhead: 0.7% of Clockticks
Load STLB Hit: 0.0% of Clockticks
Load STLB Miss: 0.7% of Clockticks

**Loads Blocked by Store Forwarding:** 0.0% of Clockticks **Lock Latency:** 0.0% of Clockticks

A significant fraction of CPU cycles spent handling cache misses due to lock operations. Due to the microarchitecture handling of locks, they are classified as L1 Bound regardless of what memory source satisfied them. Note that this metric value may be highlighted due to Store Latency issue.

**Split Loads:** 0.0% of Clockticks **4K Aliasing:** 0.0% of Clockticks **FB Full:** 0.0% of Clockticks

This metric does a rough estimation of how often L1D Fill Buffer unavailability limited additional L1D miss memory access requests to proceed. The higher the metric value, the deeper the memory hierarchy level the misses are satisfied from. Often it hints on approaching bandwidth limits (to L2 cache, L3 cache or external memory). Avoid adding software prefetches if indeed memory BW limited.

L2 Bound:

Contested Accesses:
Data Sharing:
L3 Latency:
SQ Full:

DRAM Bound:

0.0% of Clockticks

Memory Bandwidth: 14.1% of Clockticks
Memory Latency: 21.1% of Clockticks

Store Bound: 0.0% of Clockticks

Store Latency:

False Sharing:

Split Stores:

DTLB Store Overhead:

2.00% of Clockticks
0.0% of Clockticks
1.2% of Clockticks

**Store STLB Hit:** 0.0% of Clockticks **Store STLB Hit:** 1.2% of Clockticks

**Core Bound:** 16.8% of Pipeline Slots

This metric represents how much Core non-memory issues were of a bottleneck. Shortage in hardware compute resources, or dependencies software's instructions are both categorized under Core Bound. Hence it may indicate the machine ran out of an 000 resources, certain execution units are overloaded or dependencies in program's data- or instruction- flow are limiting the performance (e.g. FP-chained long-latency arithmetic operations).

**Divider:** 0.0% of Clockticks **Port Utilization:** 8.8% of Clockticks

**Cycles of 0 Ports Utilized:** 22.8% of Clockticks **Serializing Operations:** 21.1% of Clockticks

Mixing Vectors: 0.0% of uOps

Cycles of 1 Port Utilized: 4.1% of Clockticks

Cycles of 2 Ports Utilized: 14.5% of Clockticks

Cycles of 3+ Ports Utilized: 16.6% of Clockticks

ALU Operation Utilization: 13.5% of Clockticks
Port 0: 8.3% of Clockticks
Port 5: 8.3% of Clockticks
Port 6: 29.0% of Clockticks
Load Operation Utilization: 4.1% of Clockticks

Port 2: 8.3% of Clockticks
Port 3: 8.3% of Clockticks
Store Operation Utilization: 8.3% of Clockticks
Port 4: 8.3% of Clockticks
0.0% of Clockticks

**Vector Capacity Usage (FPU):** 0.0%

**Average CPU Frequency:** 1.198 GHz

**Total Thread Count:** 9 **Paused Time:** 0s

# **Effective Physical Core Utilization:** 59.5% (2.381 out of 4)

The metric value is low, which may signal a poor physical CPU cores utilization caused by:

- load imbalance
- threading runtime overhead
- contended synchronization
- thread/process underutilization
- incorrect affinity that utilizes logical cores instead of physical cores

Explore sub-metrics to estimate the efficiency of MPI and OpenMP parallelism or run the Locks and Waits analysis to identify parallel bottlenecks for other parallel runtimes.

## **Effective Logical Core Utilization:** 35.0% (2.801 out of 8)

The metric value is low, which may signal a poor logical CPU cores utilization. Consider improving physical core utilization as the first step and then look at opportunities to utilize logical cores, which in some cases can improve processor throughput and overall performance of multi-threaded applications.

#### **Collection and Platform Info:**

**Application Command Line:** ./codecs/HHI-VVC/decoder/vvdecapp "-b"

"./bin/HHI-VVC/randomaccess\_fast.cfg/CLASS\_C/RaceHorses\_416x240\_30\_QP\_27\_HHI-VVC.bin"

**User Name:** root

**Operating System:** 5.4.0-72-generic DISTRIB\_ID=Ubuntu

DISTRIB RELEASE=18.04 DISTRIB CODENAME=bionic

DISTRIB DESCRIPTION="Ubuntu 18.04.5 LTS"

**Computer Name:** eimon

**Result Size:** 14.5 MB

**Collection start time:** 22:29:07 18/04/2021 UTC

**Collection stop time:** 22:29:07 18/04/2021 UTC

**Collector Type:** Event-based sampling driver

CPU:

Name: Intel(R) Processor code named Kabylake

ULX

**Frequency:** 1.992 GHz

**Logical CPU Count:** 8

**Cache Allocation Technology:** 

**Level 2 capability:** not detected

**Level 3 capability:** not detected

# Intel<sup>®</sup> oneAPI VTune<sup>™</sup> Profiler 2021.1.1 Gold

**Elapsed Time:** 0.025s

Application execution time is too short. Metrics data may be unreliable. Consider reducing the sampling interval or increasing your application execution time.

**Clockticks:** 173,160,000 **Instructions Retired:** 151,740,000

**CPI Rate:** 1.141

The CPI may be too high. This could be caused by issues such as memory stalls, instruction starvation, branch misprediction or long latency instructions. Explore the other hardware-related metrics to identify what is causing high CPI.

MUX Reliability: 0.967

**Retiring:** 46.8% of Pipeline Slots

A high fraction of pipeline slots was utilized by useful work. While the goal is to make this metric value as big as possible, a high Retiring value for non-vectorized code could prompt you to consider code vectorization. Vectorization enables doing more computations without significantly increasing the number of instructions, thus improving the performance. Note that this metric value may be highlighted due to Microcode Sequencer (MS) issue, so the performance can be improved by avoiding using the MS.

**Light Operations:** 32.9% of Pipeline Slots

 FP Arithmetic:
 0.0% of uOps

 FP x87:
 0.0% of uOps

 FP Scalar:
 0.0% of uOps

 FP Vector:
 0.0% of uOps

 Other:
 100.0% of uOps

**Heavy Operations:** 13.9% of Pipeline Slots

CPU retired heavy-weight operations (instructions that required 2+ uops) in a significant fraction of cycles.

**Microcode Sequencer:** 1.1% of Pipeline Slots 0.0% of Pipeline Slots

**Front-End Bound:** 36.3% of Pipeline Slots

Issue: A significant portion of Pipeline Slots is remaining empty due to issues in the Front-End.

Tips: Make sure the code working size is not too large, the code layout does not require too many memory accesses per cycle to get enough instructions for filling four pipeline slots, or check for microcode assists.

## **Front-End Latency:** 28.1% of Pipeline Slots

This metric represents a fraction of slots during which CPU was stalled due to front-end latency issues, such as instruction-cache misses, ITLB misses or fetch stalls after a branch misprediction. In such cases, the front-end delivers no uOps.

#### **ICache Misses:** 6.2% of Clockticks

Issue: A significant portion of instruction fetches is missing in the instruction cache.

#### Tips:

- 1. Use profile-guided optimization to reduce the size of hot code regions.
- 2. Consider compiler options to reorder functions so that hot functions are located together.
- 3. If your application makes significant use of macros, try to reduce this by either converting the relevant macros to functions or using linker options to eliminate repeated code.
- 4. Consider the 0s/01 optimization level or the following subset of optimizations to decrease your code footprint:
- use inlining only when it decreases the footprint
  - disable loop unrolling
  - disable intrinsic inlining

#### Optimization examples:

Instruction Cache Misses recipe from Intel VTune Profiler Performance Analysis Cookbook

ITLB Overhead:
Branch Resteers:
Mispredicts Resteers:
Clears Resteers:
Unknown Branches:
DSB Switches:
Length Changing Prefixes:

1.2% of Clockticks
0.0% of Clockticks
0.0% of Clockticks
0.0% of Clockticks
0.0% of Clockticks

Issue: A significant fraction of cycles was stalled due to switches of uOp delivery to the Microcode Sequencer (MS). Commonly used instructions are optimized for delivery by the DSB or MITE pipelines. Certain operations cannot be handled natively by the execution pipeline, and must be performed by microcode (small programs injected into the execution stream). Switching to the MS too often can negatively impact performance. The MS is designated to deliver long uOp flows required by CISC instructions like CPUID, or uncommon conditions like Floating Point Assists when dealing with Denormals. Note that this metric value may be highlighted due to Microcode Sequencer issue.

**Front-End Bandwidth:** 8.2% of Pipeline Slots **Front-End Bandwidth MITE:** 28.1% of Clockticks **Front-End Bandwidth DSB:** 4.7% of Clockticks

(Info) DSB Coverage: 24.0%

**Bad Speculation:**Branch Mispredict:
Machine Clears:
3.5% of Pipeline Slots
0.0% of Pipeline Slots
3.5% of Pipeline Slots

```
Back-End Bound:
                            13.5% of Pipeline Slots
                              9.5% of Pipeline Slots
  Memory Bound:
     L1 Bound:
                                 9.4% of Clockticks
        DTLB Overhead:
                                    0.6% of Clockticks
           Load STLB Hit:
                                       0.0% of Clockticks
           Load STLB Miss:
                                       0.6% of Clockticks
        Loads Blocked by Store Forwarding: 0.0% of Clockticks
                                    0.0% of Clockticks
        Lock Latency:
        Split Loads:
                                    0.0% of Clockticks
        4K Aliasing:
                                    0.0% of Clockticks
        FB Full:
                                    0.0% of Clockticks
     L2 Bound:
                                 3.1% of Clockticks
     L3 Bound:
                                 6.2% of Clockticks
        Contested Accesses:
                                    0.0% of Clockticks
                                    0.0% of Clockticks
        Data Sharing:
        L3 Latency:
                                    0.0% of Clockticks
        SO Full:
                                    0.0% of Clockticks
     DRAM Bound:
                                 0.0% of Clockticks
        Memory Bandwidth:
                                    0.0% of Clockticks
        Memory Latency:
                                    18.7% of Clockticks
     Store Bound:
                                 0.0% of Clockticks
        Store Latency:
                                    12.5% of Clockticks
        False Sharing:
                                    0.0% of Clockticks
        Split Stores:
                                    0.0% of Clockticks
        DTLB Store Overhead:
                                    0.7% of Clockticks
           Store STLB Hit:
                                       0.0% of Clockticks
           Store STLB Hit:
                                       0.7% of Clockticks
  Core Bound:
                              3.9% of Pipeline Slots
     Divider:
                                 0.0% of Clockticks
     Port Utilization:
                                 7.7% of Clockticks
        Cycles of 0 Ports Utilized: 25.7% of Clockticks
           Serializing Operations:
                                       12.5% of Clockticks
           Mixing Vectors:
                                       0.0% of uOps
        Cycles of 1 Port Utilized:
                                    9.4% of Clockticks
        Cycles of 2 Ports Utilized: 14.0% of Clockticks
        Cycles of 3+ Ports Utilized: 18.7% of Clockticks
           ALU Operation Utilization: 15.2% of Clockticks
              Port 0:
                                          9.4% of Clockticks
                                          9.4% of Clockticks
              Port 1:
                                          9.4% of Clockticks
              Port 5:
                                          32.7% of Clockticks
              Port 6:
           Load Operation Utilization: 9.4% of Clockticks
              Port 2:
                                          14.0% of Clockticks
              Port 3:
                                          9.4% of Clockticks
           Store Operation Utilization: 9.4% of Clockticks
              Port 4:
                                          9.4% of Clockticks
              Port 7:
                                          4.7% of Clockticks
        Vector Capacity Usage (FPU): 0.0%
Average CPU Frequency: 2.244 GHz
Total Thread Count:
                            9
```

**Paused Time:** 0s

# **Effective Physical Core Utilization:** 49.6% (1.984 out of 4)

The metric value is low, which may signal a poor physical CPU cores utilization caused by:

- load imbalance
- threading runtime overhead
- contended synchronization
- thread/process underutilization
- incorrect affinity that utilizes logical cores instead of physical cores

Explore sub-metrics to estimate the efficiency of MPI and OpenMP parallelism or run the Locks and Waits analysis to identify parallel bottlenecks for other parallel runtimes.

#### **Effective Logical Core Utilization:** 37.2% (2.975 out of 8)

The metric value is low, which may signal a poor logical CPU cores utilization. Consider improving physical core utilization as the first step and then look at opportunities to utilize logical cores, which in some cases can improve processor throughput and overall performance of multi-threaded applications.

#### **Collection and Platform Info:**

**Application Command Line:** ./codecs/HHI-VVC/decoder/vvdecapp "-b" "./bin/HHI-VVC/randomaccess\_fast.cfg/CLASS\_C/ RaceHorses 416x240 30 QP 27 HHI-VVC.bin"

User Name: root

**Operating System:** 5.4.0-72-generic DISTRIB\_ID=Ubuntu DISTRIB\_RELEASE=18.04 DISTRIB\_CODENAME=bionic DISTRIB\_DESCRIPTION="Ubuntu 18.04.5 LTS"

**Computer Name:** eimon

**Result Size:** 14.1 MB

**Collection start time:** 07:49:28 19/04/2021 UTC

**Collection stop time:** 07:49:28 19/04/2021 UTC

**Collector Type:** Event-based sampling driver

CPU:

Name: Intel(R) Processor code named Kabylake

ULX

**Frequency:** 1.992 GHz

**Logical CPU Count:** 8

**Cache Allocation Technology:** 

Level 2 capability: not detected

**Level 3 capability:** not detected