

# インテル® oneAPI ツールキットが 提供するパフォーマンス解析ツール の新機能のご紹介

# インテル® oneAPI ベース・ツールキット

インテルの CPU、GPU、FPGA  
向けにアプリケーションを開発  
するための基本ツールとライブ  
ラリーの基本セットを含む。

インテル® VTune™ プロファイ  
ラーおよびインテル® Advisor  
はベース・ツールキットに同梱

## インテル® oneAPI ベース・ツールキット

### ダイレクト・プログラミング

インテル® oneAPI DPC++/C++  
コンパイラ

インテル® DPC++ 互換性ツール

インテル® ディストリビュー  
ション for Python\*

oneAPI ベース・ツールキット  
用インテル® FPGA アドオン

### API ベースのプログラミング

インテル® oneAPI DPC++  
ライブラリー  
(インテル® oneDPL)

インテル® oneAPI マス  
カーネル・ライブラリー  
(インテル® oneMKL)

インテル® oneAPI データ  
アナリティクス・ライブラリー  
(インテル® oneDAL)

インテル® oneAPI スレッディング  
・ビルディング・ブロック  
(インテル® oneTBB)

インテル® oneAPI ビデオ  
・プロセッシング・ライブラリー  
(インテル® oneVPL)

インテル® oneAPI コレクティブ  
・コミュニケーション・ライブラリー  
(インテル® oneCCL)

インテル® oneAPI ディープ・ニュー  
ラル・ネットワーク・ライブラリー  
(インテル® oneDNN)

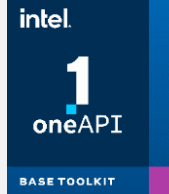
インテル® インテグレートッド  
・パフォーマンス・プリミティブ  
(インテル® IPP)

### 解析/デバッグツール

インテル® VTune™  
プロファイラー

インテル® Advisor

インテル® ディストリビュー  
ション for GDB

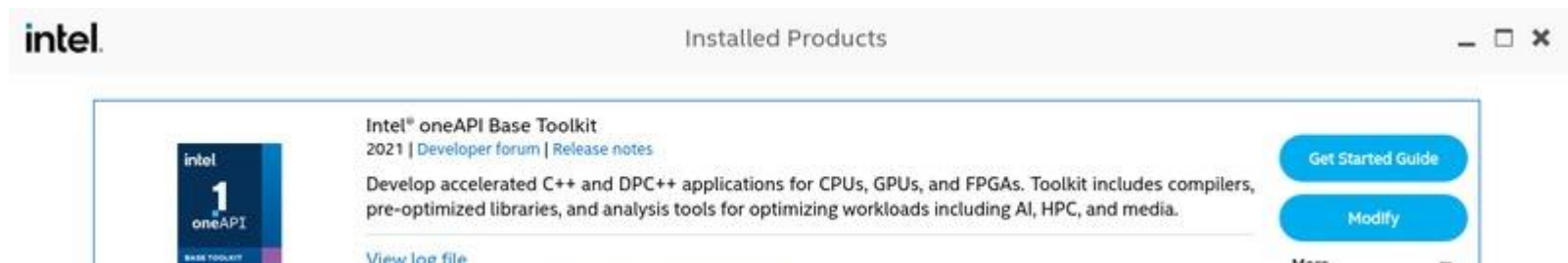


# インストール製品の一覧 バージョン確認

現時点の最新バージョン

インテル® VTune™ プロファイラー: 2021.4.0 – 6月リリース

インテル® Advisor: 2021.2 – 3月リリース



# インテル® VTune™ プロファイラー

# インテル® VTune™ プロファイラー

## DPC++ プロファイル - CPU、GPU、FPGA 向けにチューニング

### ■ データ並列 C++ (DPC++) の解析機能を追加

最も多くの時間を費やしている DPC++ の行を確認

インテルの GPU、CPU、FPGA 向けのチューニング - 複数 GPU を利用する環境をサポート  
サポートしているハードウェア・アクセラレーター向けに最適化

### ■ オフロードされた計算処理に対する解析機能のサポート

OpenMP\* オフロードのパフォーマンスをチューニング

### ■ 広範なパフォーマンス・プロファイルを追加

CPU、GPU、FPGA、スレッド、メモリー、キャッシュ、ストレージ

### ■ DPC++ 言語の追加

DPC++、C、C++、Fortran、Python\*、Go\*、Java\*、混在したコード

# 主な解析タイプ一覧

Performance Snapshot

Hotspots

Memory Consumption

Microarchitecture Exploration

Memory Access

Threading

HPC Performance Characterization

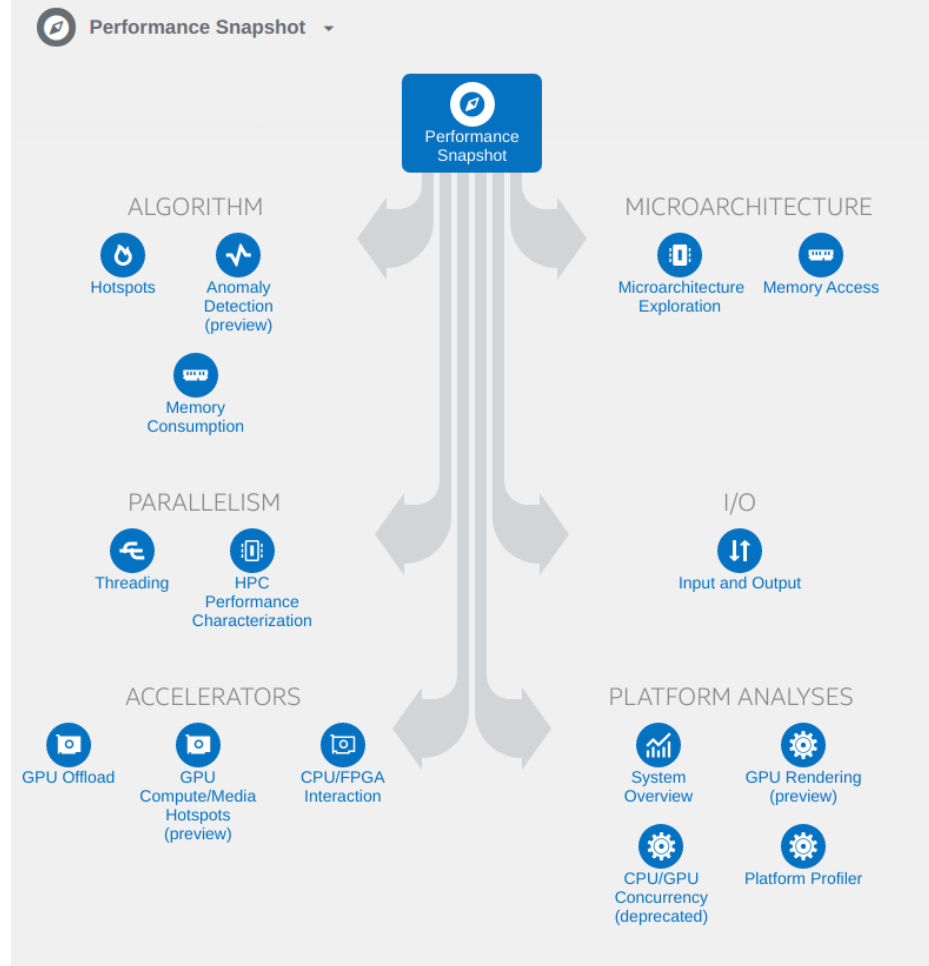
Input and Output

GPU Offload

GPU Compute/Media Hotspots (preview)

CPU/FPGA Interaction

Anomaly Detection (preview)



# パフォーマンス・スナップショットを使用した解析タイプの選択

- 解析タイプにパフォーマンス・スナップショットを追加  
HTML、CLI 向けのレポートも引き続き利用可能

## ■ 次の解析タイプの選択

パフォーマンス・スナップショットの結果から、次にどの解析タイプを選択すれば有益な情報が得られるか参照

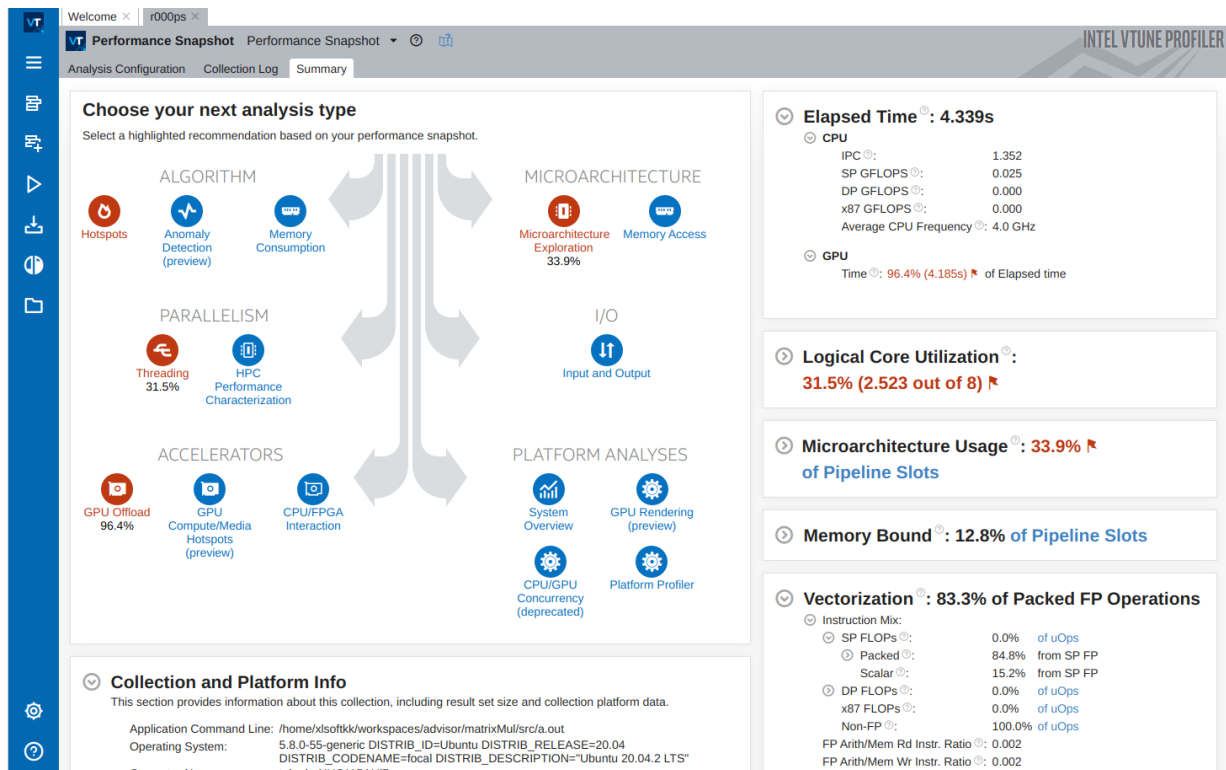
The image shows the Intel VTune Profiler interface. On the left, the 'Configure Analysis' window is open, showing the 'WHAT' section with 'Launch Application' selected. The application path is set to `/home/xlsoftk/intel/vtune/samples/matrix/matrix`. The 'HOW' section on the right displays a grid of analysis types, with 'Performance Snapshot' highlighted at the top. The analysis types are categorized into: ALGORITHM (Hotspots, Anomaly Detection (preview), Memory Consumption), MICROARCHITECTURE (Microarchitecture Exploration, Memory Access), PARALLELISM (Threading, HPC Performance Characterization), I/O (Input and Output), ACCELERATORS (GPU Offload, GPU Compute/Media Hotspots (preview), CPU/FPGA), and PLATFORM ANALYSES (System Overview, GPU Rendering (preview), CPU/GPU Concurrency (deprecated), Platform Profiler).

# パフォーマンス・スナップショットの解析結果

- パフォーマンスに重要な項目に限定した簡単なスナップショットを取得

Bottom-up 画面やソースコードとの関連付けは行わない

- より詳細な分析を行うためのステップ



# 得られる情報

ベクトル化された処理とスカラー処理の割合  
メモリアクセスに制限されているか  
GPU の利用状況と、動作効率の表示

## Vectorization <sup>Ⓞ</sup>: 0.1% <sup>🚩</sup> of Packed FP Operations

### Instruction Mix:

SP FLOPs <sup>Ⓞ</sup> :	0.0%	of uOps
Packed <sup>Ⓞ</sup> :	7.0%	from SP FP
128-bit <sup>Ⓞ</sup> :	6.4% <sup>🚩</sup>	from SP FP
256-bit <sup>Ⓞ</sup> :	0.6%	from SP FP
512-bit <sup>Ⓞ</sup> :	0.0%	from SP FP
Scalar <sup>Ⓞ</sup> :	93.0% <sup>🚩</sup>	from SP FP
DP FLOPs <sup>Ⓞ</sup> :	24.3%	of uOps
Packed <sup>Ⓞ</sup> :	0.1%	from DP FP
Scalar <sup>Ⓞ</sup> :	99.9% <sup>🚩</sup>	from DP FP
x87 FLOPs <sup>Ⓞ</sup> :	0.0%	of uOps
Non-FP <sup>Ⓞ</sup> :	75.7%	of uOps

FP Arith/Mem Rd Instr. Ratio <sup>Ⓞ</sup>: 0.959

FP Arith/Mem Wr Instr. Ratio <sup>Ⓞ</sup>: 1.906

## Memory Bound <sup>Ⓞ</sup>: 80.4% <sup>🚩</sup> of Pipeline Slots

L1 Bound <sup>Ⓞ</sup>: 4.0% of Clockticks

L2 Bound <sup>Ⓞ</sup>: 0.3% of Clockticks

L3 Bound <sup>Ⓞ</sup>: 18.5% <sup>🚩</sup> of Clockticks

DRAM Bound <sup>Ⓞ</sup>: 64.1% <sup>🚩</sup> of Clockticks

Store Bound <sup>Ⓞ</sup>: 0.0% of Clockticks

## GPU Active Time <sup>Ⓞ</sup>: 1.6% <sup>🚩</sup>

GPU Utilization when Busy <sup>Ⓞ</sup>: 16.8% <sup>🚩</sup>

### EU State <sup>Ⓞ</sup>:

Active <sup>Ⓞ</sup>: 16.8%

Stalled <sup>Ⓞ</sup>: 46.5% <sup>🚩</sup>

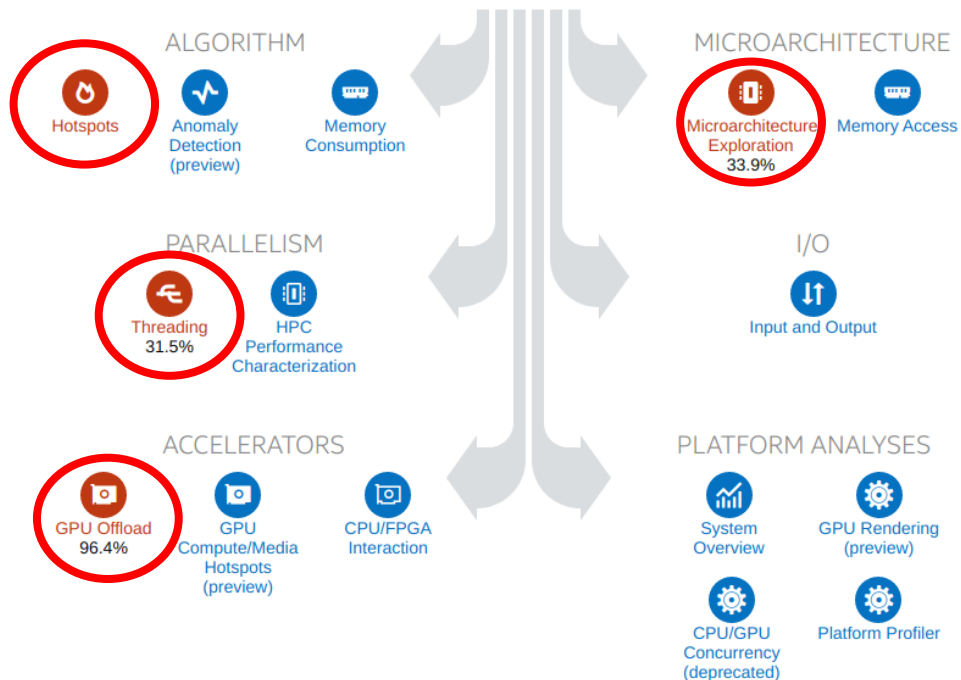
Idle <sup>Ⓞ</sup>: 36.7% <sup>🚩</sup>

Occupancy <sup>Ⓞ</sup>: 47.4% <sup>🚩</sup> of peak value

# 遅延の要因を特定する次のステップ

## Choose your next analysis type

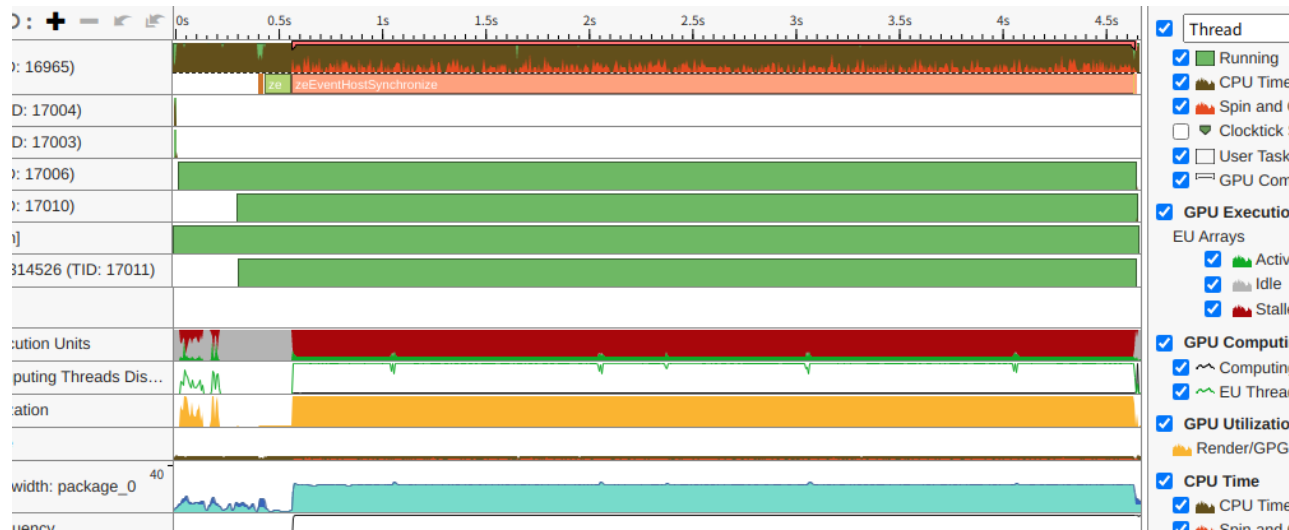
Select a highlighted recommendation based on your performance snapshot.



# DPC++ アプリケーション向け GPU 解析

## GPU Offload

- アプリケーションのパフォーマンスにとって重要な GPU 内のカーネルを特定
- メモリ確保、データ転送、実行、同期にかかる時間を区分
- パフォーマンスを妨げる不要なメモリー転送を把握

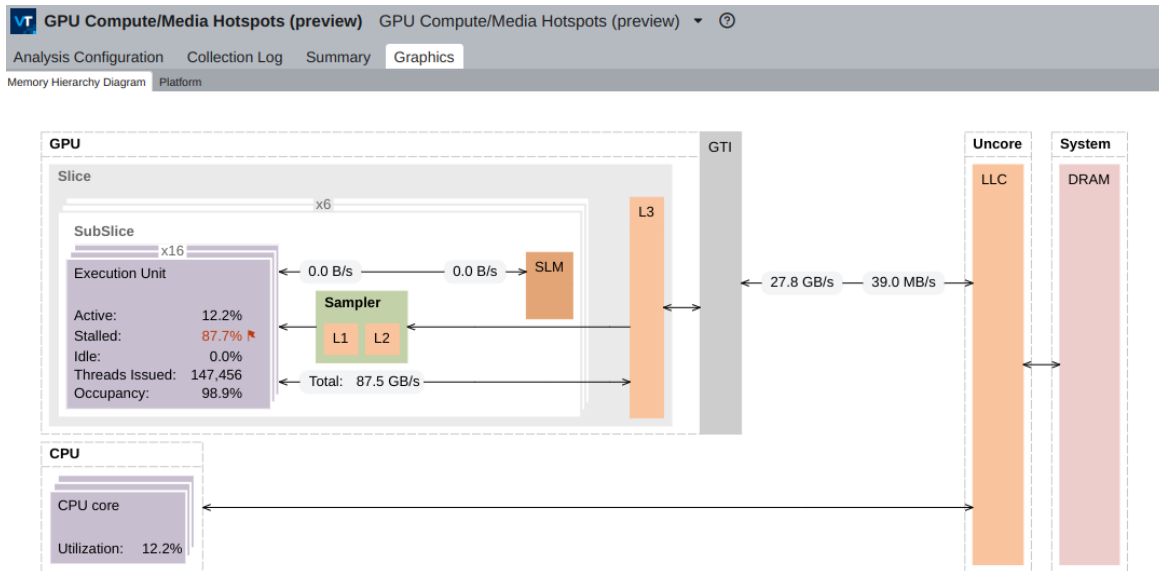


Computing Task	Total Time by Device Operation Type					Instance Count	Host-to-De
	Allocation	Host-to-Device Transfer	Execution	Device-to-Host Transfer	Synchronization		
Matrix1<float>	0.016s	0s	4.063s	0.010s	0s	1	
zeCommandListAppendBarrier	0s	0s	0s	0s	0.000s	1	
[Outside any task]						0	

# 計算タスクの分析 - メモリー階層図

## GPU Compute/Media Hotspots (preview)

- 計算タスクのメモリー階層図を表示
- DRAM、L3 のメモリーユニットと実行ユニット間のデータアクセスの状況を可視化  
データサイズ、帯域幅、GPU ユニット間の最大帯域幅
- タスクごとのワークサイズと実行ユニットの動作状態などの情報をメモリー階層図と紐付け

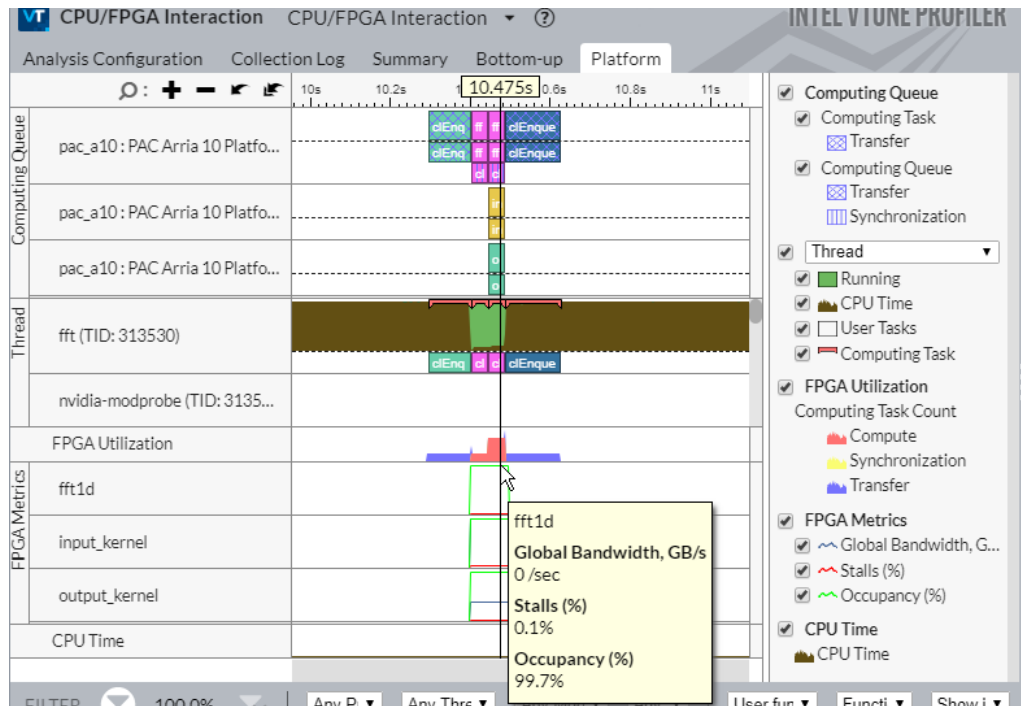


Computing Task	Work Size		EU Array			Computing Ta		
	Global	Local	Active	Stalled	Idle	Total Time ▼	Average Time	Instan
▶ Matrix1<float>	3072 x 3072	512 x 1	12.2%	87.7%	0.0%	4.087s	4.087s	
▶ zeCommandListAppendMemoryCopyRegion			31.5%	65.7%	2.7%	0.003s	0.003s	
▶ zeCommandListAppendBarrier			0.0%	0.0%	100.0%	0.000s	0.000s	
▶ [Outside capture]			0.0%	0.0%	0.0%	0.000s	0.000s	

# FPGA を利用した DPC++ アプリケーション解析

## CPU/FPGA Interaction

- FPGA の上位の計算タスク  
および、CPU の上位タスクと  
ホットスポットを解析
- カーネルのメモリー転送、CPU  
コンテキスト・スイッチ、FPU の  
利用、および CPU スレッド  
各カーネルの合計および平均実行  
時間の表示  
CPU と FPGA 間のメモリー転送の  
全体的な時間  
DPC++ または OpenCL\* カーネ  
ルのメモリー転送にかかった時間



# Input and output 解析の強化

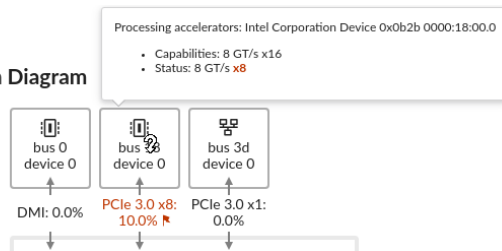
## プラットフォーム・ダイアグラム

### ■ NUMA システムにおける各階層ごとのデータ転送を可視化

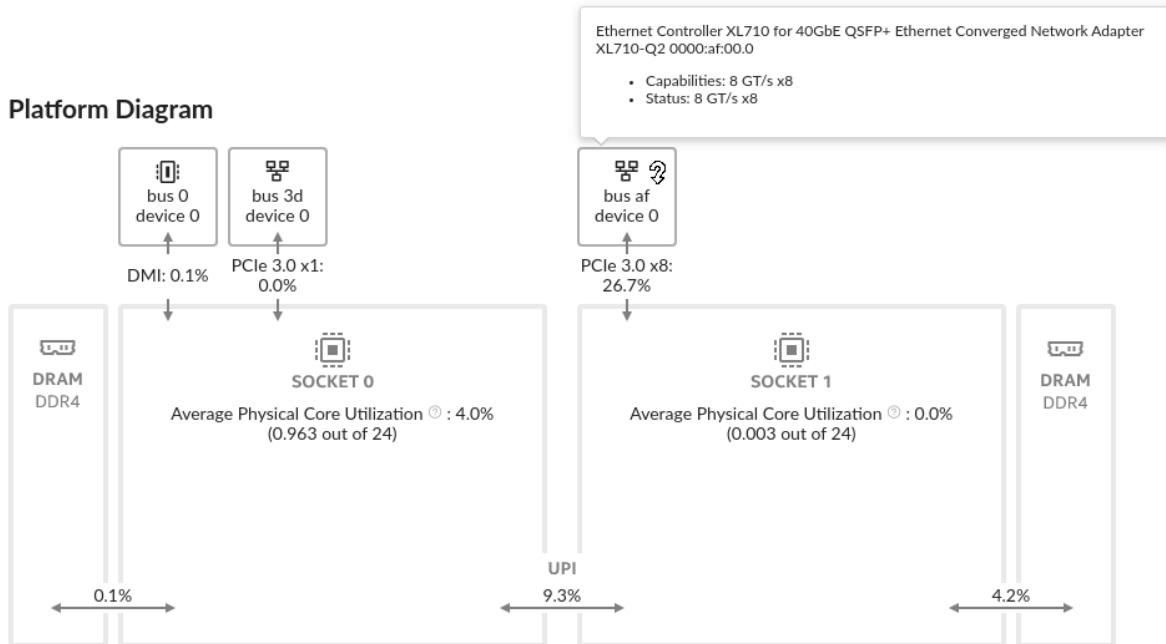
PCIe\* および インテル® UPI  
リンク、DRAM

物理コアのシステムトポロ  
ジーと使用率のメトリック

#### Platform Diagram



#### Platform Diagram



最大 4 ソケット構成の第 1 世代および第 2 世代インテル® Xeon® スケーラブル・プロセッサに対応

# インテル® Advisor

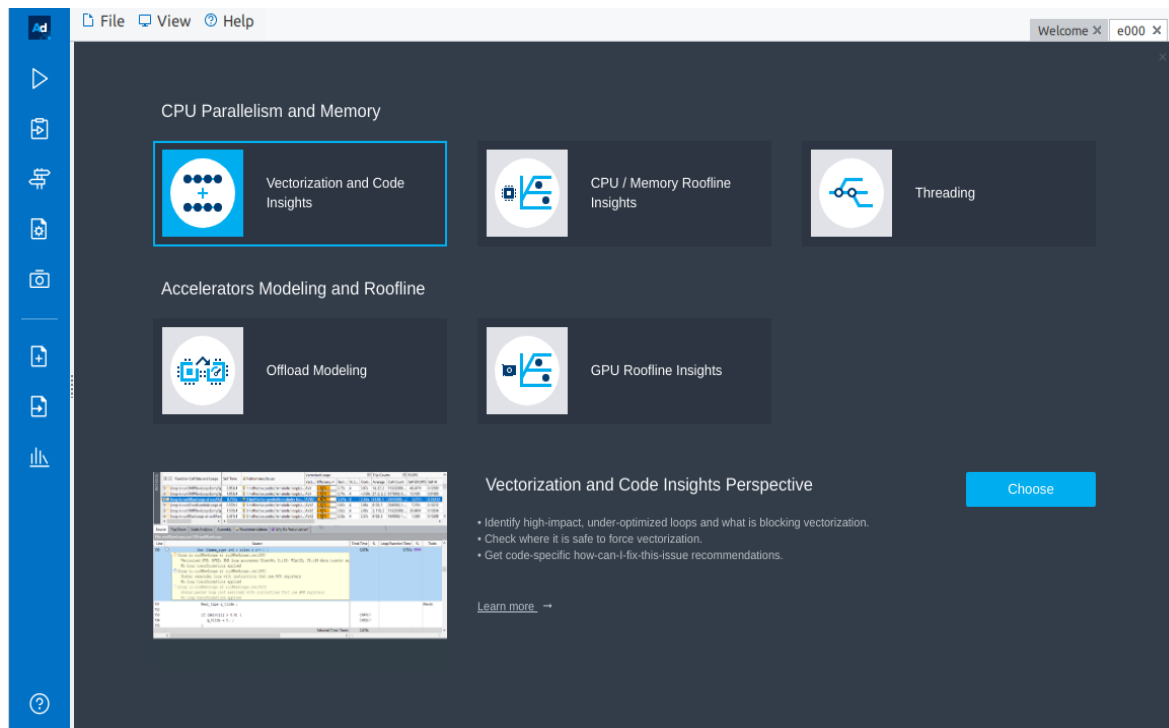
# インテル® Advisor

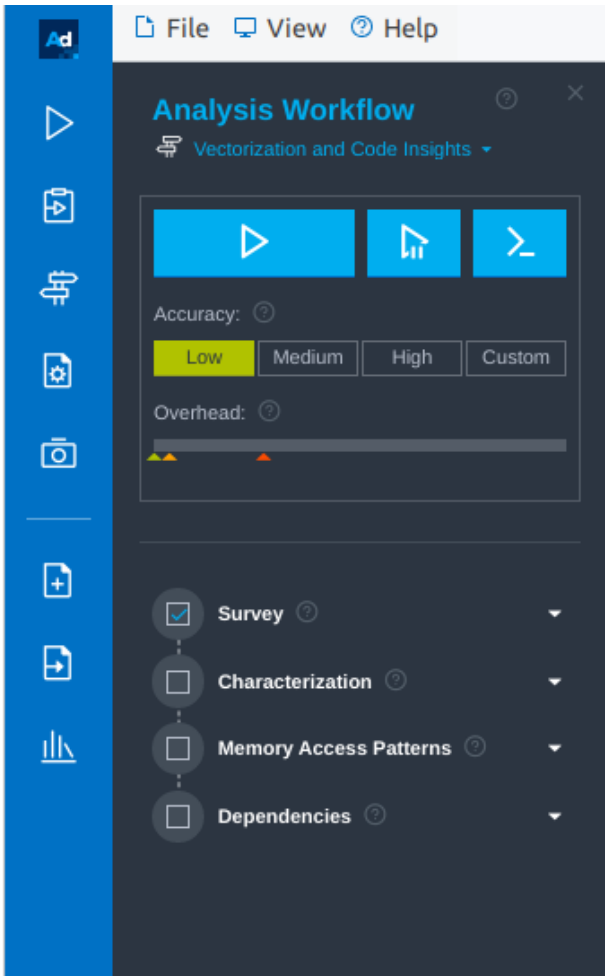
## 最新ハードウェア向けの設計を支援

- オフロード・アドバイザー  
アクセラレーターにオフロードした際のパフォーマンスを推定
- ルーフライン解析  
CPU/GPU コードの計算とメモリーを最適化
- ベクトル化アドバイザー  
ベクトル化とその最適化
- スレッド化アドバイザー  
スレッド化されていないアプリケーションに効率的なスレッド化を追加
- フローグラフ・アナライザー  
効率的なフローグラフを作成して解析

# 各機能へのアクセス

- 既存の CPU 向け最適化機能に加えて GPU 向けに 2 つの機能を追加
  - Offload Modeling
  - GPU Roofline Insights
- 各機能はパースペクティブで分けられ、サイドメニューから切り替え

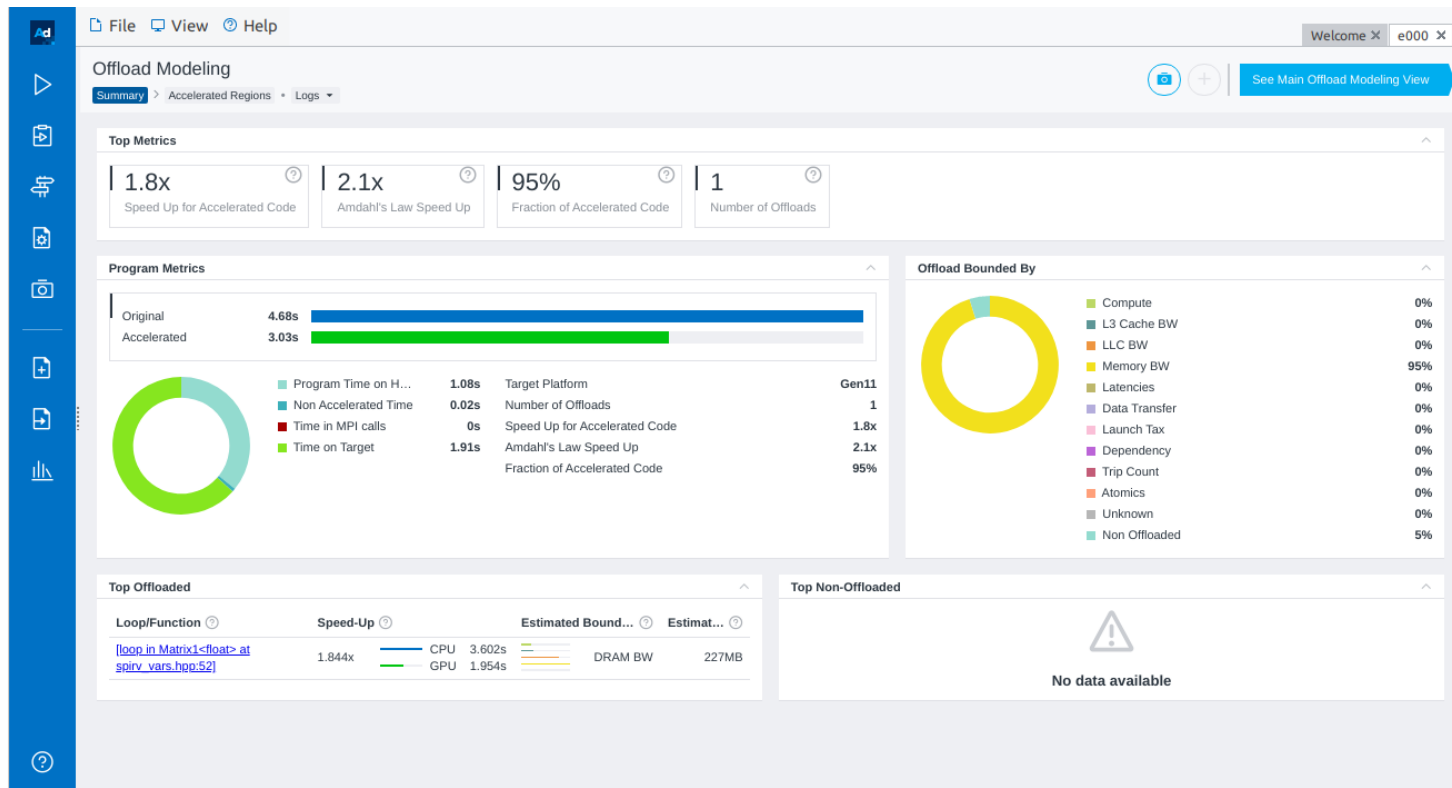




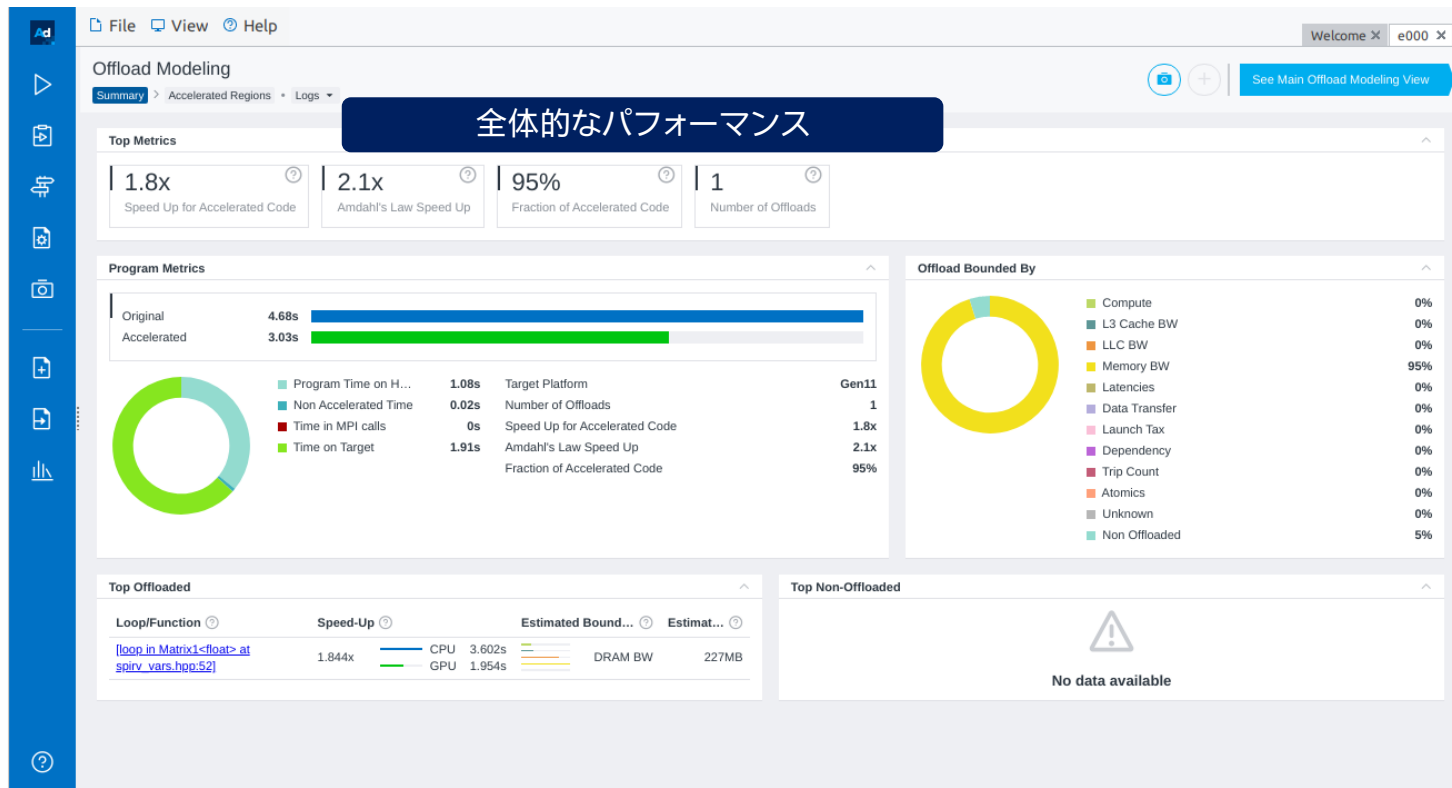
# ワークフローの変更

- 各情報を目的の情報を確認しやすいように変更

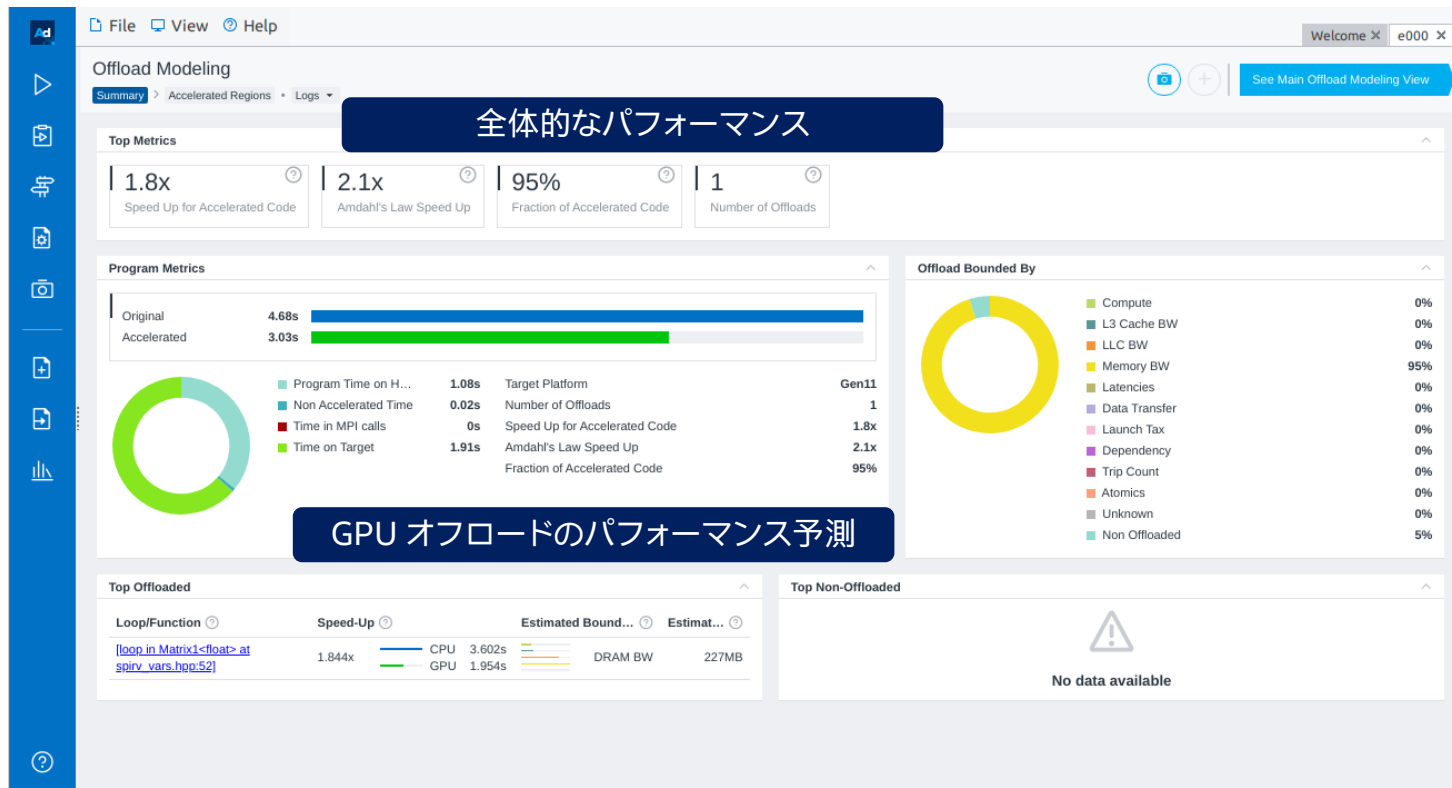
# Offload Modeling – オフロード・アドバイザー



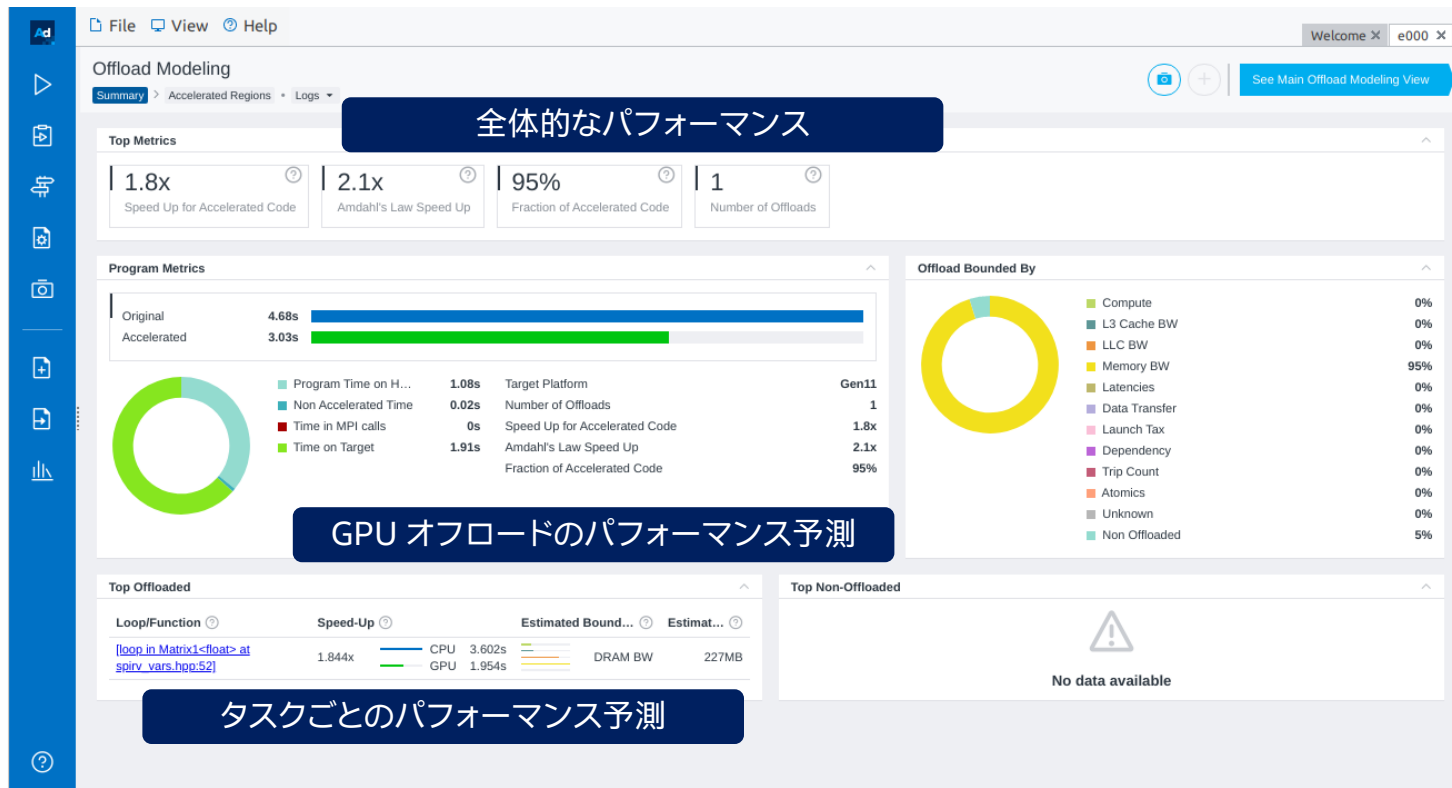
# Offload Modeling – オフロード・アドバイザー



# Offload Modeling – オフロード・アドバイザー



# Offload Modeling – オフロード・アドバイザー

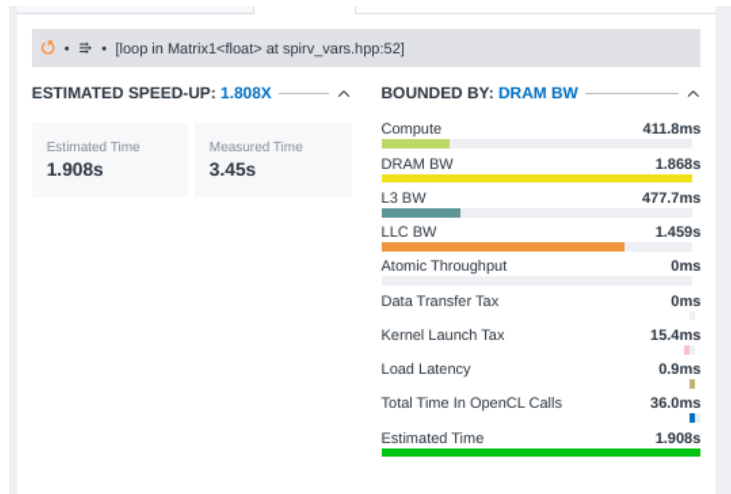
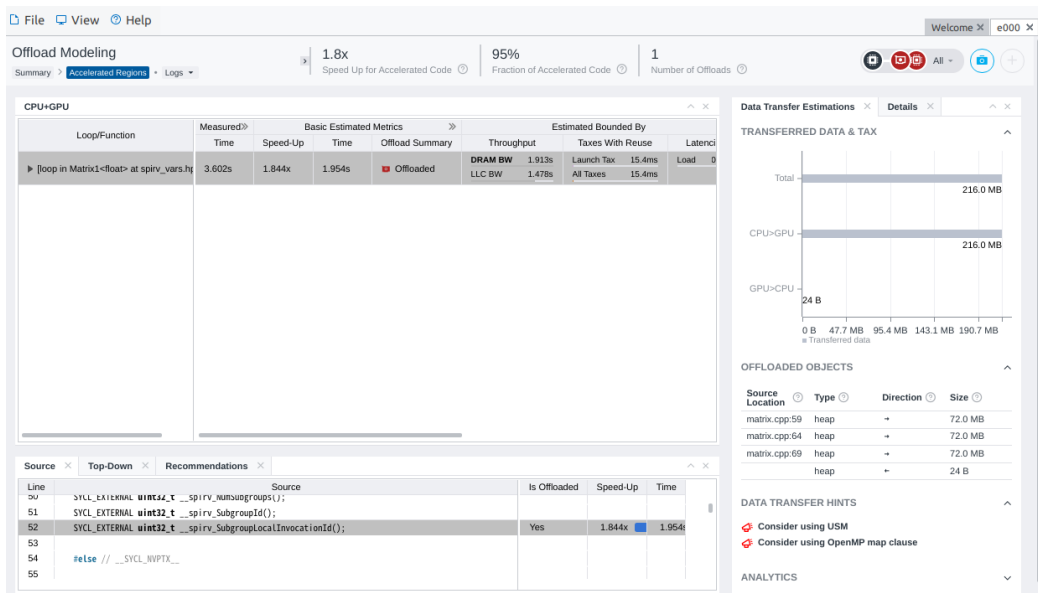


# Offload Modeling – オフロード・アドバイザー



# オフロードターゲットの詳細

各コード領域の実行を GPU 向けにモデル化した結果を表示  
予測実行時間と、計算依存かメモリー依存かを確認



# GPU Roofline Insights

- 対象システムが持つ最大パフォーマンスのベンチマークを取り、GPU カーネルとシステムのルーフライン・グラフを生成
- 以下の疑問の解決に役立ちます
  - ハードウェア・リソースで達成可能な最大のパフォーマンスは何か？
  - アプリケーションは最適に動作しているか？
  - 最適化されていない場合、最適化を行うために最良の方法は何か？
  - 各最適化候補のパフォーマンスを制限しているのは、メモリー帯域幅か計算性能か？
- DPC++、C++/Fortran (OpenMP\* プラグマ付き)、GPU 上で実行可能な OpenCL\* アプリケーション

## GPU Roofline Insights

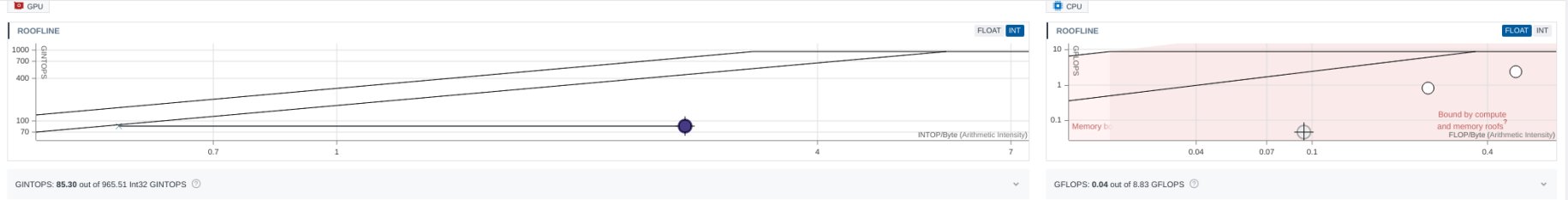
Summary > GPU Roofline Regions > Logs

See Main GPU Roofline View

### Program Metrics

<b>2.87s</b> Program Elapsed Time	<b>2.235s</b> GPU Time	<b>0.003s</b> Data Transfer Time	<b>0.63s</b> CPU Time
<b>91.0%</b> EU Threading Occupancy			<b>1</b> Thread Count
GINTOPS: <b>85.30</b>			GFLOPS: <b>0.04</b>

### OPIS and Bandwidth



### Top Hotspots

Compute Task	Elapsed Time	GFLOPS	GINTOPS	Work Size/Local	Active/Stalled/Idle, %	Function Call Sites and Loops	Self Elapsed Time	Self GFLOPS	Self GINTOPS
<a href="#">Matrix-ctransp</a>	2.23s	25.974	85.411	3072 x 3072/512 x 1	21.5/76.3/2.1	<a href="#">loop_in_GLOBAL_sub_I...</a>	0.02s	0	0
						<a href="#">loop_in_main</a>	0.01s	0.786	0.885
						<a href="#">loop_in_intel_avx_res_me...</a>	0.01s	0	0.177
						<a href="#">loop_in_main</a>	<0.01s	2.355	1.325
						<a href="#">loop_in_ZZSt9call_onceIZN...</a>	0s	0	0

Platform Information | Collection Information

### Performance Characteristics



### GPU Roofline Insights

Summary > GPU Roofline Regions > Logs

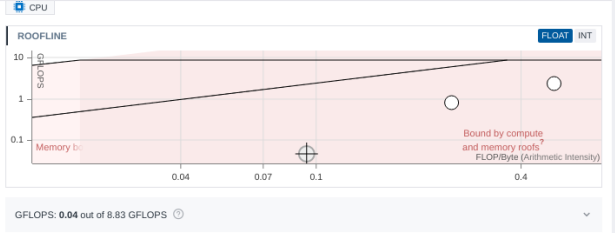
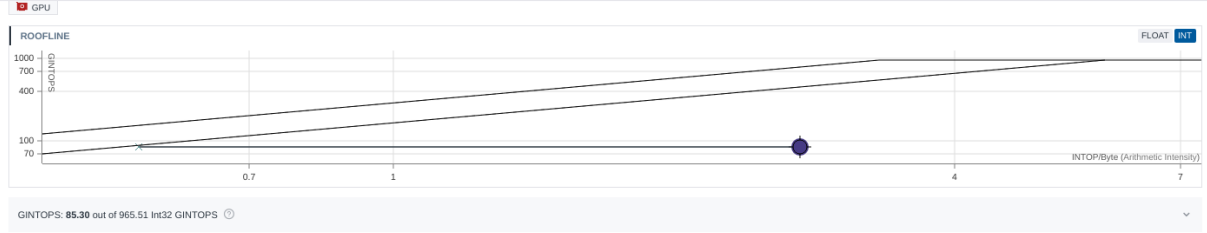
# 全体的なパフォーマンス

See Main GPU Roofline View

#### Program Metrics

2.87s Program Elapsed Time	2.235s GPU Time	0.003s Data Transfer Time	0.63s CPU Time
91.0% EU Threading Occupancy			1 Thread Count
GINTOPS: 85.30			GFLOPS: 0.04

#### OPIS and Bandwidth



#### Top Hotspots

Compute Task	Elapsed Time	GFLOPS	GINTOPS	Work Size/Local	Active/Stalled/Idle, %	Function Call Sites and Lo...	Self Elapsed Time	Self GFLOPS	Self GINTOPS
<a href="#">Matrix-ctrfloat</a>	2.23s	25.974	85.411	3072 x 3072/512 x 1	21.5/76.3/2.1	<a href="#">loop_in_GLOBAL_sub_1...</a>	0.02s	0	0
						<a href="#">loop_in_main</a>	0.01s	0.786	0.885
						<a href="#">loop_in_intel_avx_res_me...</a>	0.01s	0	0.177
						<a href="#">loop_in_main</a>	<0.01s	2.355	1.325
						<a href="#">loop_in_ZZSt9call_once1Z...</a>	0s	0	0

Platform Information < Collection Information

#### Performance Characteristics



### GPU Roofline Insights

Summary > GPU Roofline Regions > Logs

# 全体的なパフォーマンス

See Main GPU Roofline View

#### Program Metrics

2.87s Program Elapsed Time | 2.235s GPU Time | 0.003s Data Transfer Time | 0.63s CPU Time

GPU

91.0% EU Threading Occupancy

GINTOPS: 85.30

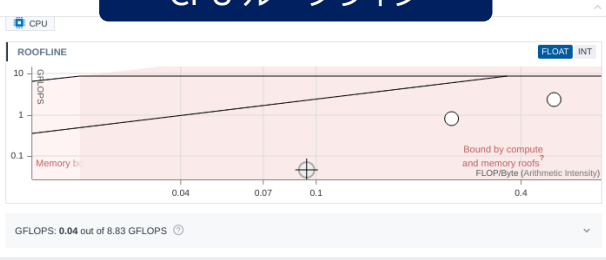
CPU

1 Thread Count

GFLOPS: 0.04

## GPU ルーフライン

## CPU ルーフライン



#### Top Hotspots

Compute Task	Elapsed Time	GFLOPS	GINTOPS	Work Size/Local	Active/Stalled/Idle, %	Function Call Sites and Lo...	Self Elapsed Time	Self GFLOPS	Self GINTOPS
<a href="#">Matrix-ctransp</a>	2.23s	25.974	85.411	3072 x 3072/512 x 1	21.5/76.3/2.1	<a href="#">loop_in_GLOBAL_sub_1...</a>	0.02s	0	0
						<a href="#">loop_in_main</a>	0.01s	0.786	0.885
						<a href="#">loop_in_intel_avx_res_me...</a>	0.01s	0	0.177
						<a href="#">loop_in_main</a>	<0.01s	2.355	1.325
						<a href="#">loop_in_ZZSt9call_once1Z...</a>	0s	0	0

Platform Information | Collection Information



### GPU Roofline Insights

Summary > GPU Roofline Regions > Logs

See Main GPU Roofline View

# 全体的なパフォーマンス

**Program Metrics**

2.87s Program Elapsed Time	2.235s GPU Time	0.003s Data Transfer Time	0.63s CPU Time
-------------------------------	--------------------	------------------------------	-------------------

**GPU**

91.0%  
EU Threading Occupancy

GINTOPS: 85.30

**CPU**

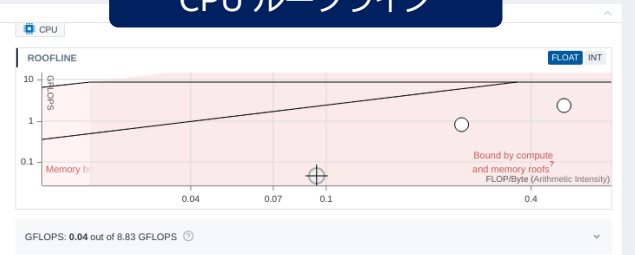
1  
Thread Count

GFLOPS: 0.04

## GPU ルーフライン



## CPU ルーフライン



## 各 GPU カーネルの時間、GFLOPS

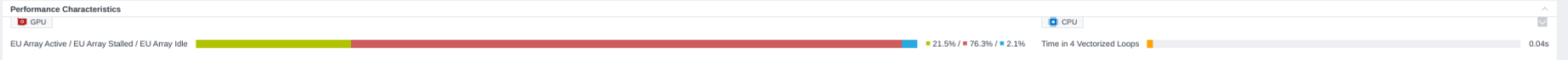
**Top Hotspots**

Compute Task	Elapsed Time	GFLOPS	GINTOPS	Work Size/Local	Active/Stalled/Idle, %
<a href="#">Matrix-ctrmat</a>	2.23s	25.974	85.411	3072 x 3072/512 x 1	21.5/76.3/2.1

**CPU**

Function Call Sites and L...	Self Elapsed Time	Self GFLOPS	Self GINTOPS
<a href="#">[loop_in_GLOBAL_sub_1...</a>	0.02s	0	0
<a href="#">[loop_in_main]</a>	0.01s	0.786	0.885
<a href="#">[loop_in_intel_avx_res_me...</a>	0.01s	0	0.177
<a href="#">[loop_in_main]</a>	<0.01s	2.355	1.325
<a href="#">[loop_in_ZZSt9call_onceLN...</a>	0s	0	0

Platform Information > Collection Information



### GPU Roofline Insights

Summary > GPU Roofline Regions > Logs

See Main GPU Roofline View

# 全体的なパフォーマンス

**Program Metrics**

2.87s Program Elapsed Time	2.235s GPU Time	0.003s Data Transfer Time	0.63s CPU Time
-------------------------------	--------------------	------------------------------	-------------------

**GPU**

91.0%  
EU Threading Occupancy

GINTOPS: 85.30

**CPU**

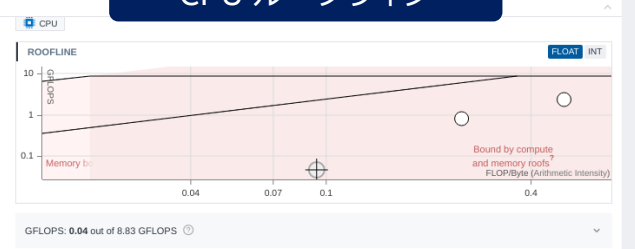
1  
Thread Count

GFLOPS: 0.04

## GPU ルーフライン



## CPU ルーフライン



## 各 GPU カーネルの時間、GFLOPS

**Top Hotspots**

Compute Task	Elapsed Time	GFLOPS	GINTOPS	Work Size/Local	Active/Stalled/Idle, %
<a href="#">Matrix-ctrmat</a>	2.23s	25.974	85.411	3072 x 3072/512 x 1	21.5/76.3/2.1

Function Call Sites and L...	Self Elapsed Time	Self GFLOPS	Self GINTOPS
<a href="#">[loop_in_GLOBAL_sub_1...</a>	0.02s	0	0
<a href="#">[loop_in_main]</a>	0.01s	0.786	0.885
<a href="#">[loop_in_intel_avx_res_me...</a>	0.01s	0	0.177
<a href="#">[loop_in_main]</a>	<0.01s	2.355	1.325
<a href="#">[loop_in_ZZ59call_once1Z...</a>	0s	0	0

## EU の状態

**Platform Information**

**Performance Characteristics**

EU Array Active / EU Array Stalled / EU Array Idle

Time in 4 Vectorized Loops

# 参考情報

インテル® VTune™ プロファイラー・パフォーマンス・クックブック

<https://software.intel.com/content/www/us/en/develop/documentation/vtune-cookbook/top/configuration-recipes/profiling-openmp-offload-application.html> (英語)

インテル® VTune™ プロファイラー・ユーザーガイド

<https://software.intel.com/content/www/us/en/develop/documentation/vtune-help/top.html> (英語)

インテル® Advisor ユーザーガイド

<https://software.intel.com/content/www/us/en/develop/documentation/advisor-user-guide/top.html> (英語)

# 参考情報

iSUS - IA Software User Society

<https://www.isus.jp/>

GPU 上で実行する OpenMP\* オフロード・アプリケーションのプロファイル

<https://www.isus.jp/products/vtune/vtune-cookbook-profiling-openmp-offload-application/>

パフォーマンス異常を見つけるアプリケーションのプロファイル

<https://www.isus.jp/products/vtune/vtune-cookbook-profiling-performance-anomalies/>

Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。

\*その他の社名、製品名などは、一般に各社の商標または登録商標です。

製品および性能に関する情報: 性能は、使用状況、構成、その他の要因によって異なります。詳細については、<http://www.intel.com/PerformanceIndex/> (英語) を参照してください。

注意事項の改訂 #20201201

© 2021 Intel Corporation. 無断での引用、転載を禁じます。

XLsoft のロゴ、XLsoft は XLsoft Corporation の商標です。Copyright © 2021 XLsoft Corporation.