

www.phoronix-test-suite.com

10980XE onednn onnx

Intel Core i9-10980XE testing with a ASRock X299 Steel Legend (P1.30 BIOS) and NVIDIA GeForce GTX 1080 Ti 11GB on Ubuntu 22.04 via the Phoronix Test Suite.

Automated Executive Summary

C had the most wins, coming in first place for 39% of the tests.

Based on the geometric mean of all complete results, the fastest (C) was 1.154x the speed of the slowest (D). A was 0.876x the speed of C, B was 0.998x the speed of A, D was 0.991x the speed of B.

The results with the greatest spread from best to worst included:

oneDNN (Harness: Deconvolution Batch shapes_1d - Data Type: u8s8f32 - Engine: CPU) at 69.719x
oneDNN (Harness: Deconvolution Batch shapes_3d - Data Type: u8s8f32 - Engine: CPU) at 5.773x
speedtest-cli (Internet Latency) at 1.858x
ONNX Runtime (Model: super-resolution-10 - Device: CPU - Executor: Standard) at 1.663x
fast-cli (Internet Latency) at 1.625x
perf-bench (Benchmark: Sched Pipe) at 1.456x
oneDNN (Harness: Matrix Multiply Batch Shapes Transformer - Data Type: f32 - Engine: CPU) at 1.347x
ONNX Runtime (Model: ArcFace ResNet-100 - Device: CPU - Executor: Standard) at 1.335x
oneDNN (Harness: Recurrent Neural Network Training - Data Type: f32 - Engine: CPU) at 1.288x

10980XE onednn onnx

ONNX Runtime (Model: fcn-resnet101-11 - Device: CPU - Executor: Standard) at 1.244x.

Test Systems:

A

B

C

D

Processor: Intel Core i9-10980XE @ 4.80GHz (18 Cores / 36 Threads), Motherboard: ASRock X299 Steel Legend (P1.30 BIOS), Chipset: Intel Sky Lake-E DMI3 Registers, Memory: 32GB, Disk: Samsung SSD 970 PRO 512GB, Graphics: NVIDIA GeForce GTX 1080 Ti 11GB, Audio: Realtek ALC1220, Monitor: ASUS VP28U, Network: Intel I219-V + Intel I211

OS: Ubuntu 22.04, Kernel: 5.15.0-17-generic (x86_64), Desktop: GNOME Shell 40.5, Display Server: X Server 1.20.13, Display Driver: NVIDIA 495.46, OpenGL: 4.6.0, OpenCL: OpenCL 3.0 CUDA 11.5.103, Vulkan: 1.2.186, Compiler: GCC 11.2.0, File-System: ext4, Screen Resolution: 3840x2160

Kernel Notes: Transparent Huge Pages: madvise
Compiler Notes: --build=x86_64-linux-gnu --disable-vtable-verify --disable-werror --enable-bootstrap --enable-cet --enable-checking=release --enable-clocale=gnu --enable-default-pie --enable-gnu-unique-object --enable-languages=c,ada,c++,go,brig,d,fortran,objc,obj-c++,m2 --enable-libphobos-checking=release --enable-libstdcxx-debug --enable-libstdcxx-time=yes --enable-link-serialization=2 --enable-multiarch --enable-multilib --enable-nls --enable-objc-gc=auto --enable-offload-targets=nvptx-none=/build/gcc-11-pWTZs6/gcc-11-11.2.0/debian/tmp-nvptx/usr,amdgcn-amdhsa=/build/gcc-11-pWTZs6/gcc-11-11.2.0/debian/tmp-gcn/usr --enable-plugin --enable-shared --enable-threads=posix --host=x86_64-linux-gnu --program-prefix=x86_64-linux-gnu- --target=x86_64-linux-gnu --with-abi=m64 --with-arch-32=i686 --with-build-config=bootstrap-lto-lean --with-default-libstdcxx-abi=new --with-gcc-major-version-only --with-multilib-list=m32,m64,mx32 --with-target-system-zlib=auto --with-tune=generic --without-cuda-driver -v
Processor Notes: Scaling Governor: intel_cpufreq schedutil - CPU Microcode: 0x5003102
Java Notes: OpenJDK Runtime Environment (build 11.0.13+8-Ubuntu-0ubuntu1)
Python Notes: Python 3.9.9
Security Notes: itlb_multihit: KVM: Mitigation of VMX disabled + l1tf: Not affected + mds: Not affected + meltdown: Not affected + spec_store_bypass: Mitigation of SSB disabled via prctl and seccomp + spectre_v1: Mitigation of usercopy/swapgs barriers and __user pointer sanitization + spectre_v2: Mitigation of Enhanced IBRS IBPB: conditional RSB filling + srbds: Not affected + tsx_async_abort: Mitigation of TSX disabled

	A	B	C	D
fast-cli - I.D.S (Mbit/s)	370	370	**390**	**360**
Normalized	94.87%	94.87%	100%	92.31%
fast-cli - I.U.S (Mbit/s)	**6.7**	6.9	**7.9**	6.8
Normalized	84.81%	87.34%	100%	86.08%
fast-cli - Internet Latency (ms)	**8**	**8**	**13**	10
Normalized	100%	100%	61.54%	80%
fast-cli - I.L.L.B (ms)	**73**	**64**	**64**	70
Normalized	87.67%	100%	100%	91.43%
speedtest-cli - I.D.S (Mbit/s)	324.65	321.74	**280.18**	**327.12**
Normalized	99.24%	98.36%	85.65%	100%

10980XE onednn onnx

speedtest-cli - I.U.S (Mbit/s)	**9.47**	9.22	**8.4**	8.95
Normalized	100%	97.36%	88.7%	94.51%
speedtest-cli - Internet Latency (ms)	23.87	**15.487**	**28.782**	22.677
Normalized	64.88%	100%	53.81%	68.29%
perf-bench - Epoll Wait (ops/sec)	31157	**30615**	**35556**	35205
Normalized	87.63%	86.1%	100%	99.01%
perf-bench - Futex Hash (ops/sec)	4494394	**4493415**	**4513944**	4499081
Normalized	99.57%	99.55%	100%	99.67%
perf-bench - Memcpy 1MB (GB/sec)	17.382564	**16.766498**	**17.779377**	17.277106
Normalized	97.77%	94.3%	100%	97.17%
perf-bench - Memset 1MB (GB/sec)	**66.057732**	68.321136	**69.500836**	68.902923
Normalized	95.05%	98.3%	100%	99.14%
perf-bench - Sched Pipe (ops/sec)	86982	78301	**111014**	**76237**
Normalized	78.35%	70.53%	100%	68.67%
perf-bench - Futex Lock-Pi (ops/sec)	**236**	**215**	222	231
Normalized	100%	91.1%	94.07%	97.88%
perf-bench - Syscall Basic (ops/sec)	**17140545**	17221544	17145832	**17504146**
Normalized	97.92%	98.39%	97.95%	100%
oneDNN - IP Shapes 1D - f32 - CPU (ms)	87.0693	89.0246	**87.0626**	**90.3226**
Normalized	99.99%	97.8%	100%	96.39%
oneDNN - IP Shapes 3D - f32 - CPU (ms)	58.0758	58.2895	**57.2458**	**58.4361**
Normalized	98.57%	98.21%	100%	97.96%
oneDNN - IP Shapes 1D - u8s8f32 - CPU (ms)	**83.8984**	88.3586	**89.5153**	89.0099
Normalized	100%	94.95%	93.73%	94.26%
oneDNN - IP Shapes 3D - u8s8f32 - CPU (ms)	44.4304	**44.2168**	**45.8347**	44.3999
Normalized	99.52%	100%	96.47%	99.59%
oneDNN - IP Shapes 1D - bf16bf16bf16 - CPU (ms)	93.3073	**93.432**	93.2753	**91.696**
Normalized	98.27%	98.14%	98.31%	100%
oneDNN - IP Shapes 3D - bf16bf16bf16 - CPU (ms)	55.9293	56.0613	**56.7312**	**55.5827**
Normalized	99.38%	99.15%	97.98%	100%
oneDNN - C.B.S.A - f32 - CPU (ms)	**24.0564**	24.384	**24.4218**	24.3468
Normalized	100%	98.66%	98.5%	98.81%
oneDNN - D.B.s - f32 - CPU (ms)	**99.18**	109.808	107.714	**112.623**
Normalized	100%	90.32%	92.08%	88.06%
oneDNN - D.B.s - f32 - CPU (ms)	19.0856	**18.6987**	**19.6687**	18.9448
Normalized	97.97%	100%	95.07%	98.7%
oneDNN - C.B.S.A - u8s8f32 - CPU	**23.7444**	23.7723	**24.0619**	23.7889
Normalized	100%	99.88%	98.68%	99.81%
oneDNN - D.B.s - u8s8f32 - CPU (ms)	50.2996	50.1553	**0.774919**	**54.0262**
Normalized	1.54%	1.55%	100%	1.43%
oneDNN - D.B.s - u8s8f32 - CPU (ms)	9.33345	**9.40424**	**1.62907**	9.35235
Normalized	17.45%	17.32%	100%	17.42%
oneDNN - R.N.N.T - f32 - CPU (ms)	**23584**	23367	**18305**	23372
Normalized	77.62%	78.34%	100%	78.32%
oneDNN - R.N.N.I - f32 - CPU (ms)	26070	26549	**26034**	**26873**
Normalized	99.86%	98.06%	100%	96.88%

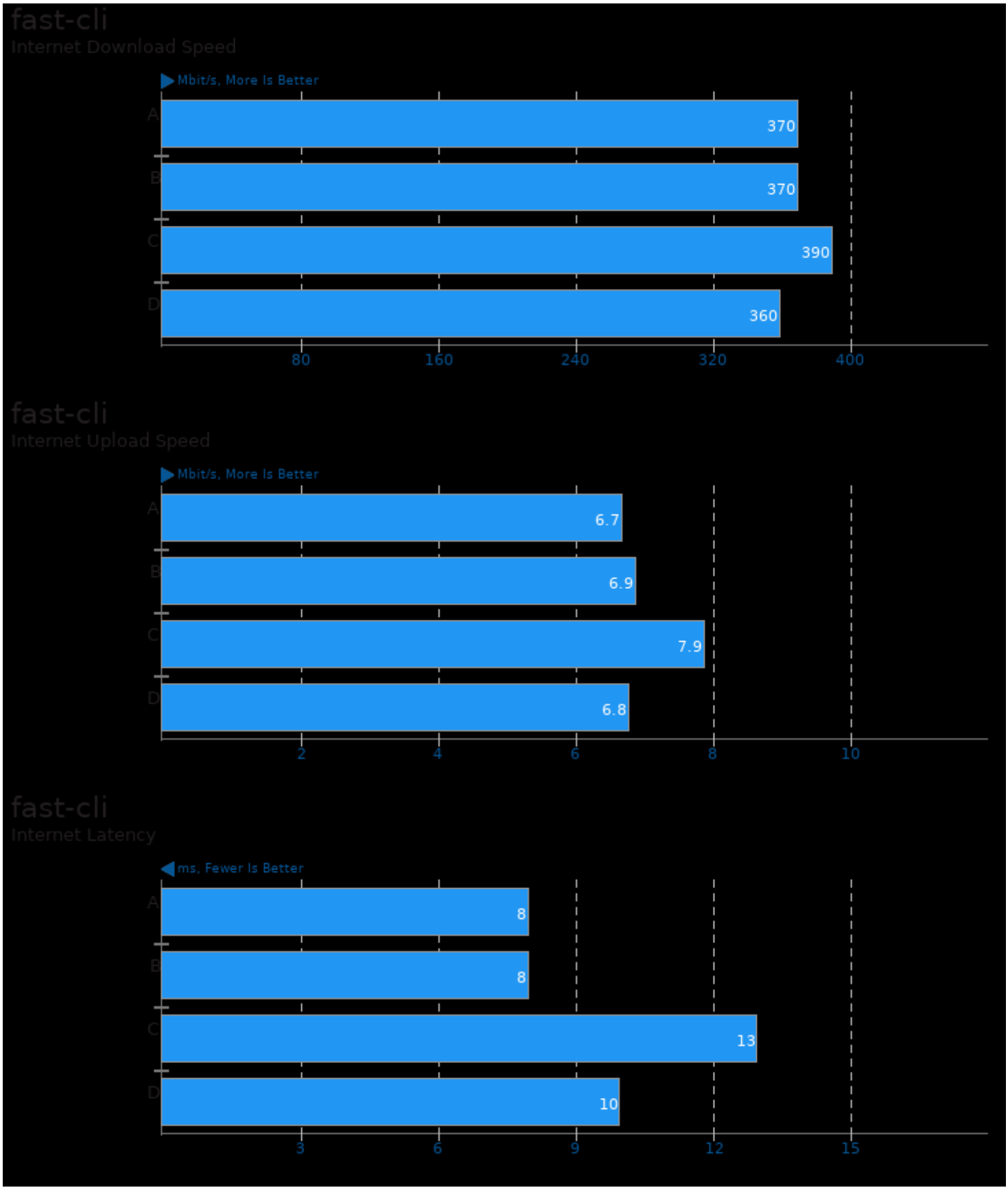
10980XE onednn onnx

oneDNN - R.N.N.T - u8s8f32 - CPU	23516	22791	**22789**	**23844**
Normalized	96.91%	99.99%	100%	95.57%
oneDNN - C.B.S.A - bf16bf16bf16 - CPU (ms)	**27.0016**	**25.8871**	26.1804	25.9415
Normalized	95.87%	100%	98.88%	99.79%
oneDNN - D.B.s - bf16bf16bf16 - CPU (ms)	**116.514**	**135.473**	130.958	118.154
Normalized	100%	86.01%	88.97%	98.61%
oneDNN - D.B.s - bf16bf16bf16 - CPU (ms)	**22.3996**	**23.9021**	22.7645	23.0195
Normalized	100%	93.71%	98.4%	97.31%
oneDNN - R.N.N.I - u8s8f32 - CPU (ms)	28357	**27248**	28624	**28759**
Normalized	96.09%	100%	95.19%	94.74%
oneDNN - M.M.B.S.T - f32 - CPU (ms)	35.2021	**36.3562**	**26.991**	31.2678
Normalized	76.67%	74.24%	100%	86.32%
oneDNN - R.N.N.T - bf16bf16bf16 - CPU (ms)	22990	22727	**22605**	**23320**
Normalized	98.32%	99.47%	100%	96.94%
oneDNN - R.N.N.I - bf16bf16bf16 - CPU (ms)	**27155**	27354	**28514**	28429
Normalized	100%	99.27%	95.24%	95.52%
oneDNN - M.M.B.S.T - u8s8f32 - CPU (ms)	**30.7784**	**34.5221**	30.8774	33.3777
Normalized	100%	89.16%	99.68%	92.21%
oneDNN - M.M.B.S.T - bf16bf16bf16 - CPU (ms)	37.7937	**39.0485**	37.7823	**37.5531**
Normalized	99.36%	96.17%	99.39%	100%
Java JMH - Throughput (Ops/s)	33246015809	**33247767441**	**33236796630**	33246121586
Normalized	99.99%	100%	99.97%	100%
ONNX Runtime - GPT-2 - CPU - Parallel (Inferences/min)	5799	**5795**	**6114**	6072
Normalized	94.85%	94.78%	100%	99.31%
ONNX Runtime - GPT-2 - CPU - Standard (Inferences/min)	**8749**	8812	**9595**	9591
Normalized	91.18%	91.84%	100%	99.96%
ONNX Runtime - yolov4 - CPU - Parallel (Inferences/min)	**444**	448	447	**449**
Normalized	98.89%	99.78%	99.55%	100%
ONNX Runtime - yolov4 - CPU - Standard (Inferences/min)	665	663	**680**	**569**
Normalized	97.79%	97.5%	100%	83.68%
ONNX Runtime - bertsquad-12 - CPU - Parallel (Inferences/min)	782	**774**	786	**796**
Normalized	98.24%	97.24%	98.74%	100%
ONNX Runtime - bertsquad-12 - CPU - Standard (Inferences/min)	**897**	911	**923**	**897**
Normalized	97.18%	98.7%	100%	97.18%
ONNX Runtime - fcn-resnet101-11 - CPU - Parallel (Inferences/min)	**101**	**102**	**101**	**101**
Normalized	99.02%	100%	99.02%	99.02%

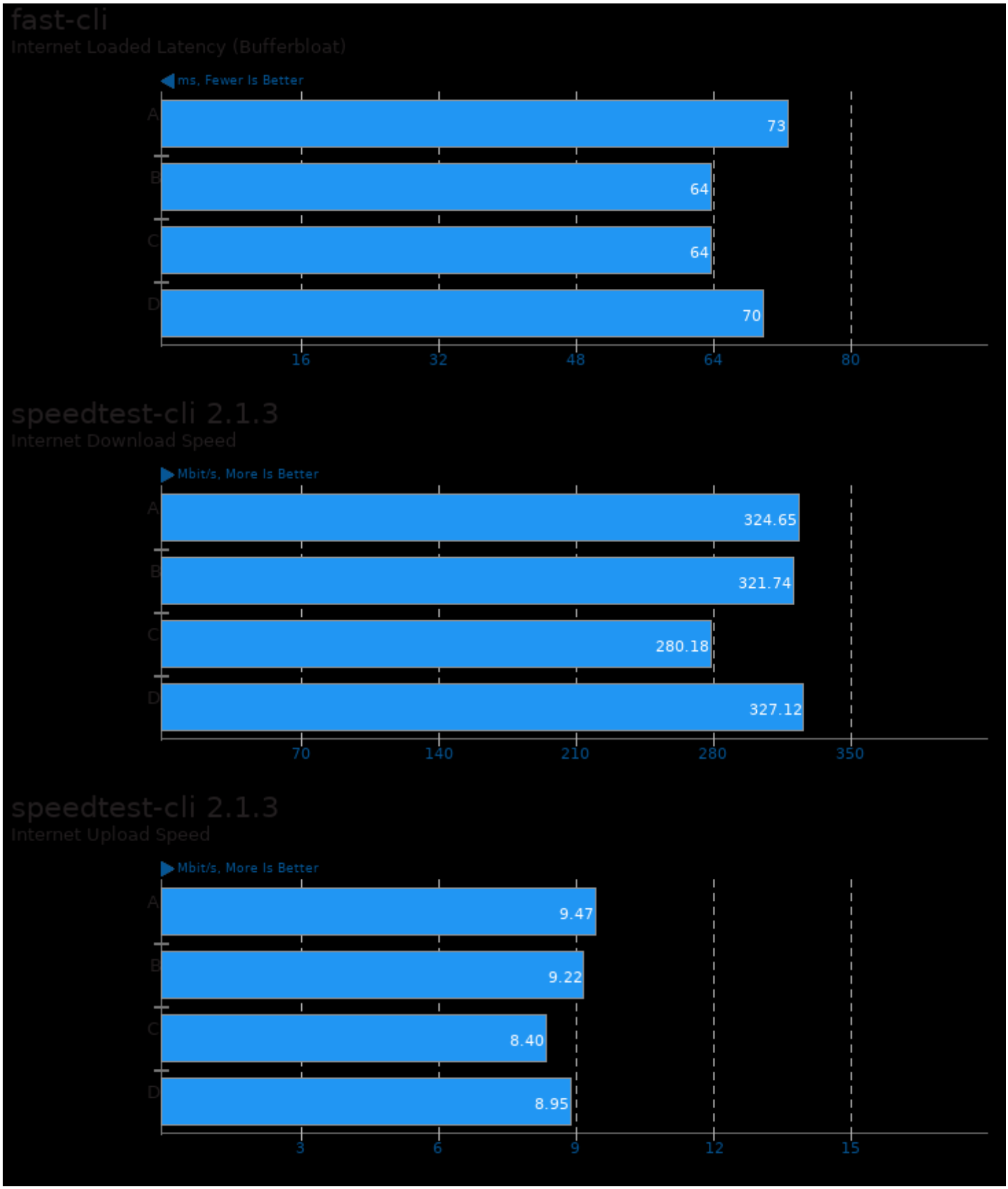
10980XE onednn onnx

ONNX Runtime - fcn-resnet101-11 - CPU - Standard (Inferences/min)	**148**	**119**	**148**
Normalized	100%	80.41%	100%
ONNX Runtime - ArcFace ResNet-100 - CPU - Parallel (Inferences/min)	**1464**	1474	**1491**
Normalized	98.19%	98.86%	100%
ONNX Runtime - ArcFace ResNet-100 - CPU - Standard (Inferences/min)	**2053**	**1538**	2014
Normalized	100%	74.91%	98.1%
ONNX Runtime - super-resolution-10 - CPU - Parallel (Inferences/min)	**5890**	**5975**	5969
Normalized	98.58%	100%	99.9%
ONNX Runtime - super-resolution-10 - CPU - Standard (Inferences/min)	**6168**	9884	**10255**
Normalized	60.15%	96.38%	100%

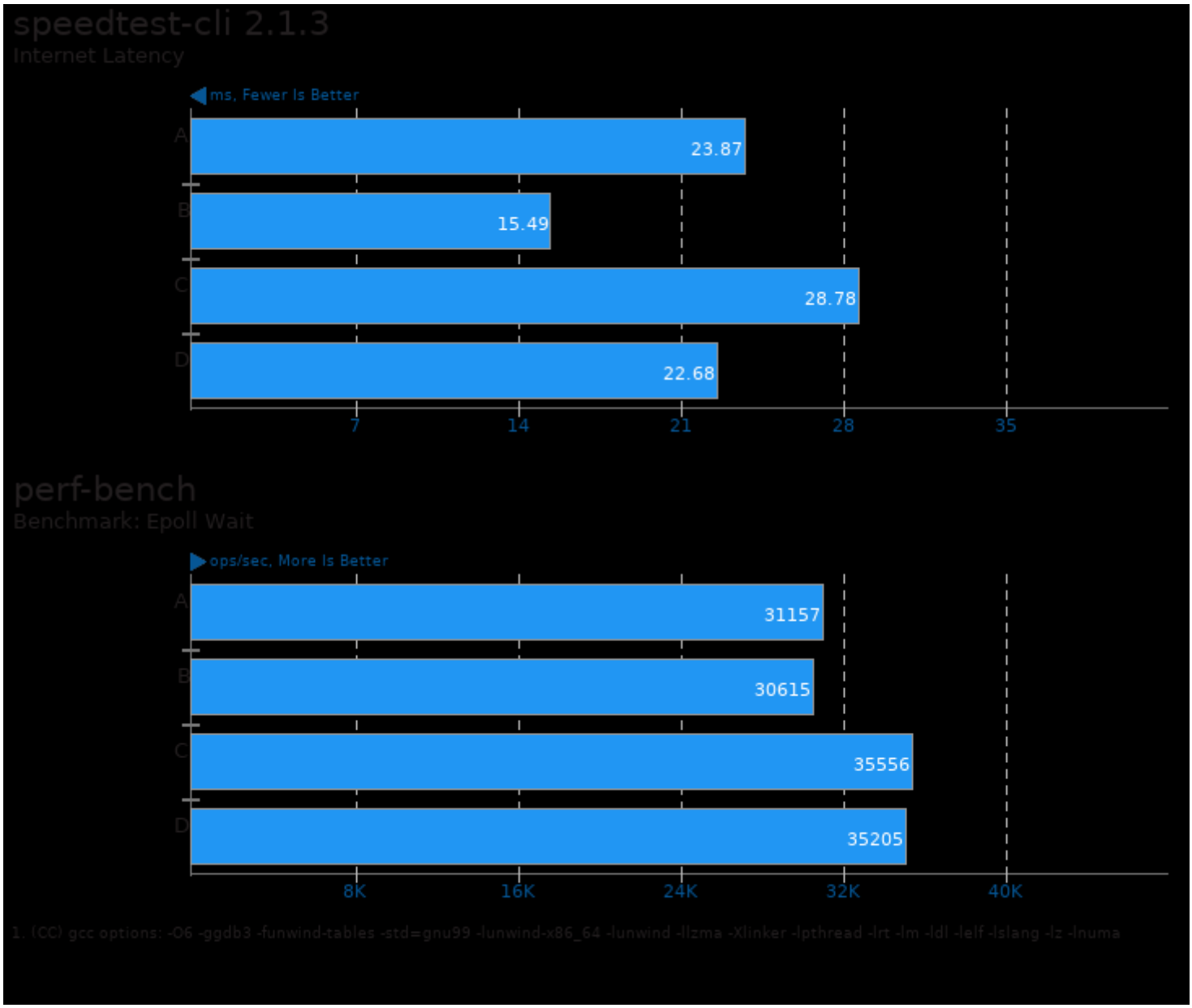
10980XE onednn onnx



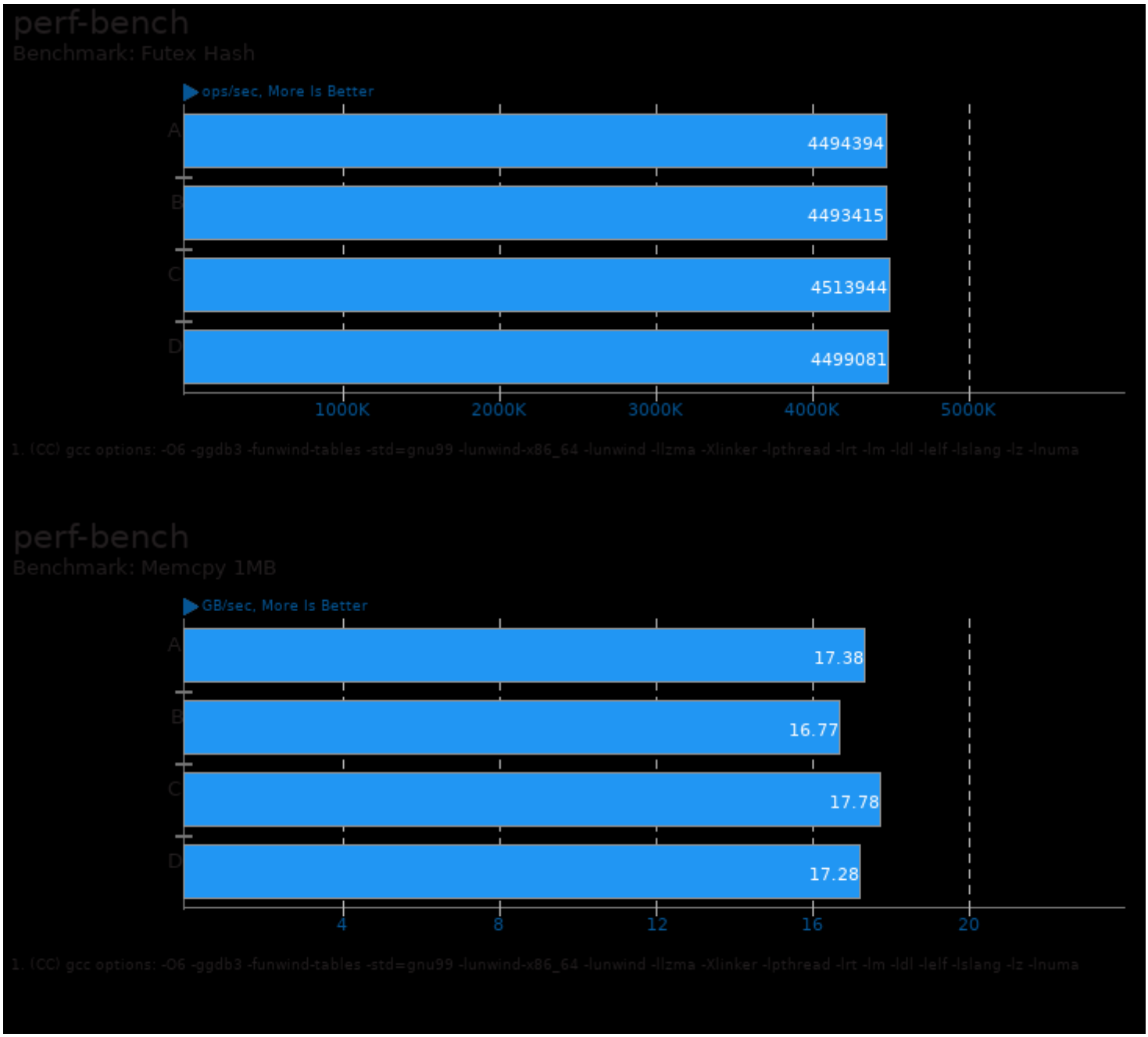
10980XE onednn onnx



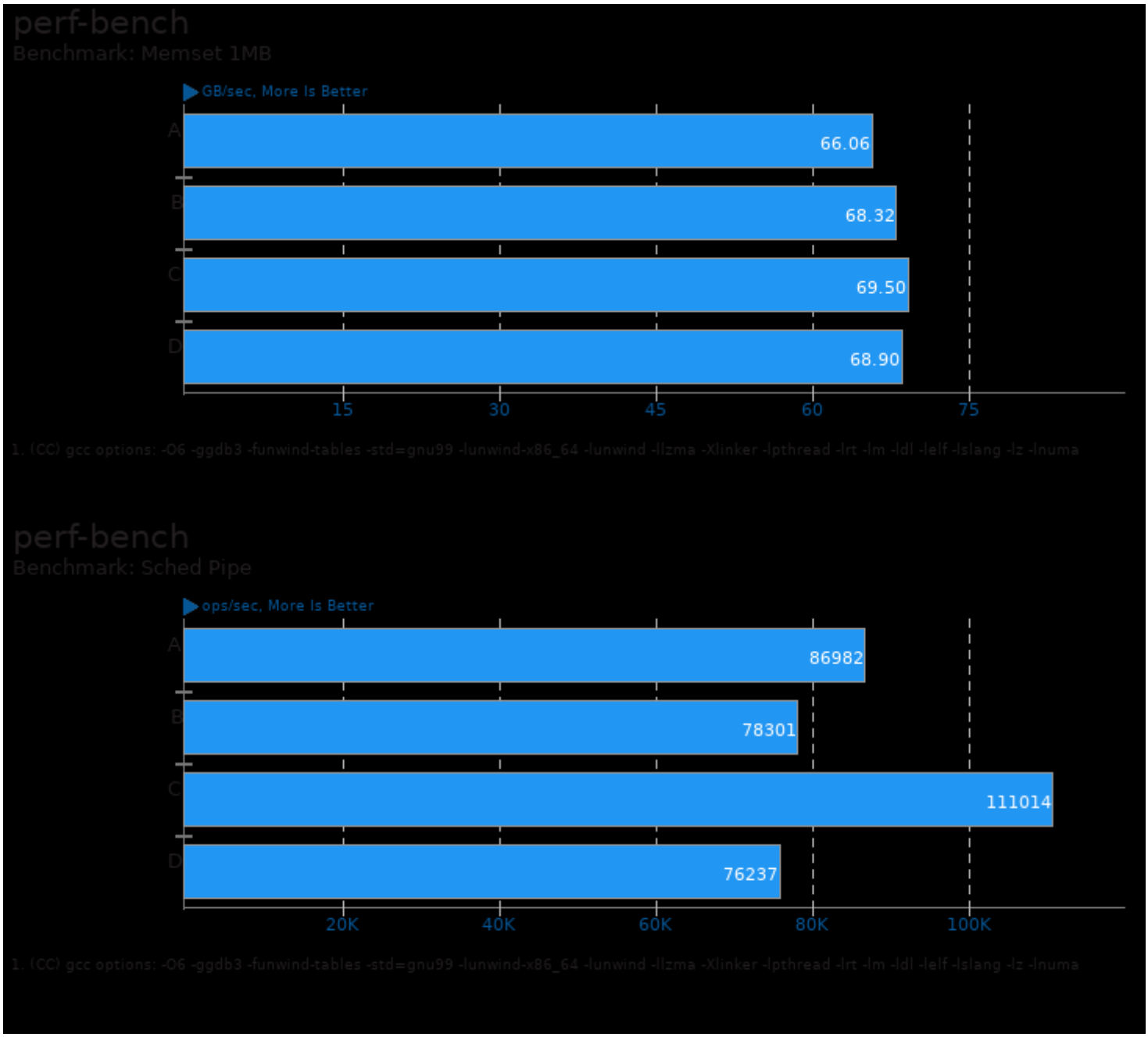
10980XE onednn onnx



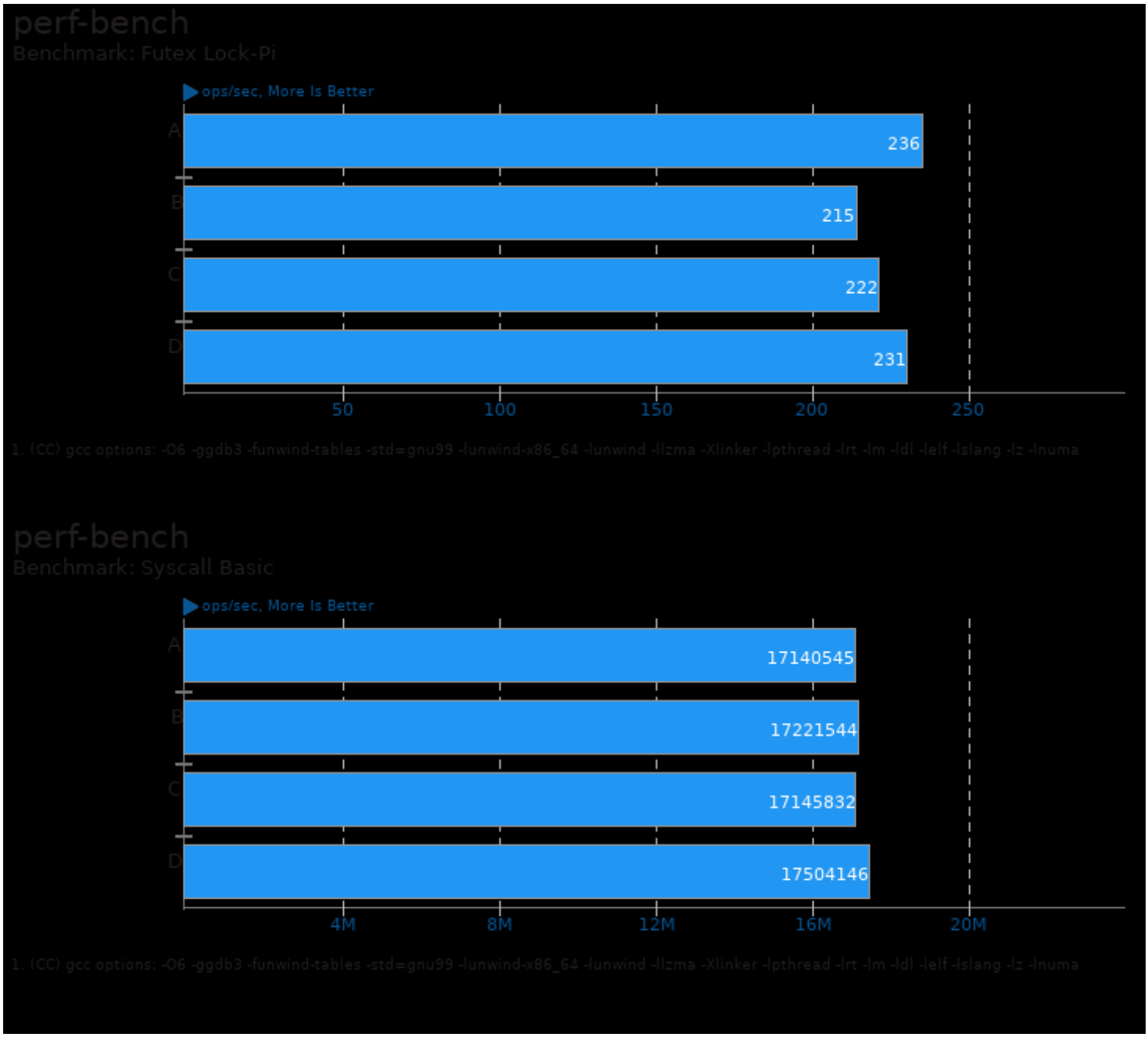
10980XE onednn onnx



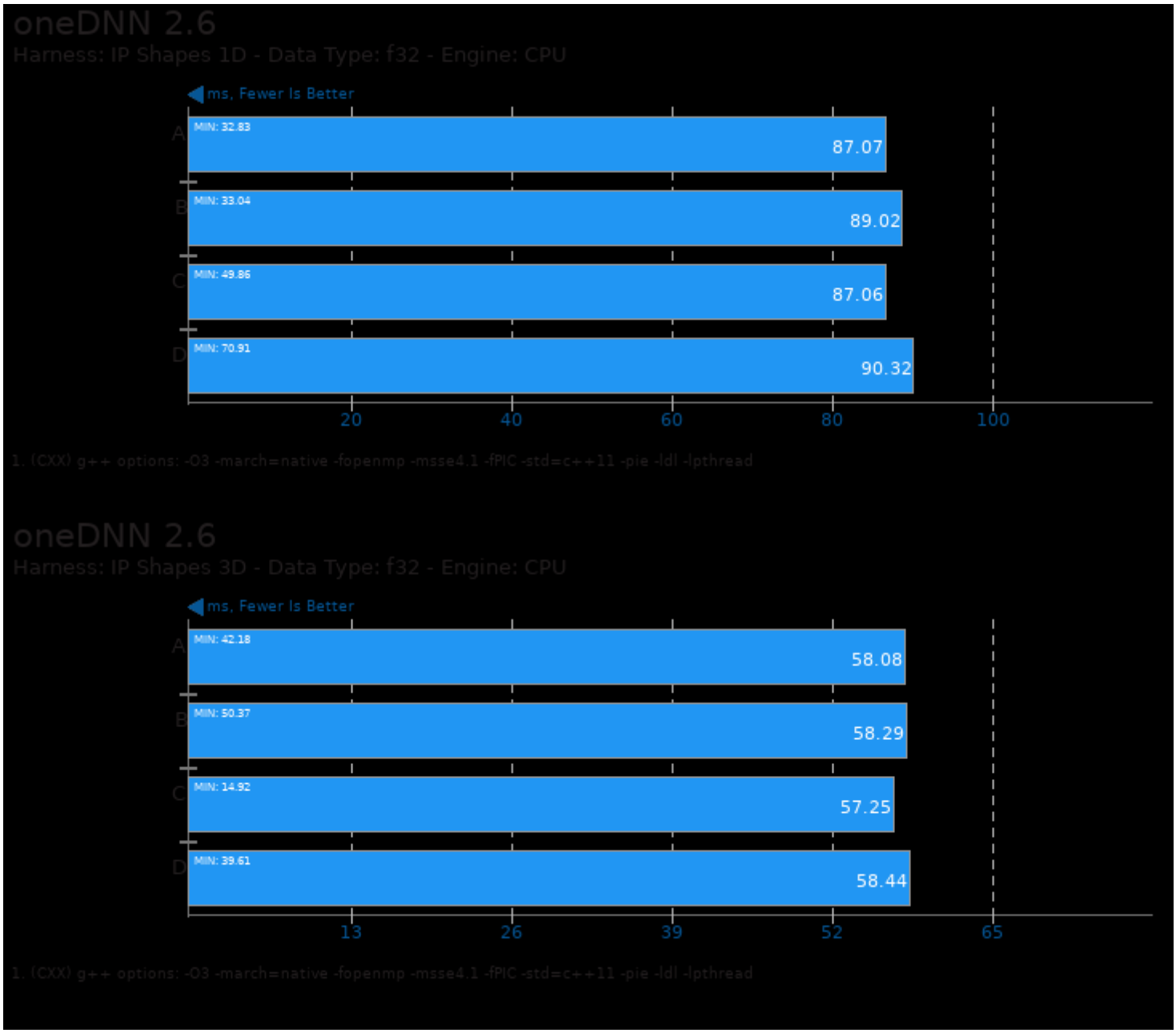
10980XE onednn onnx



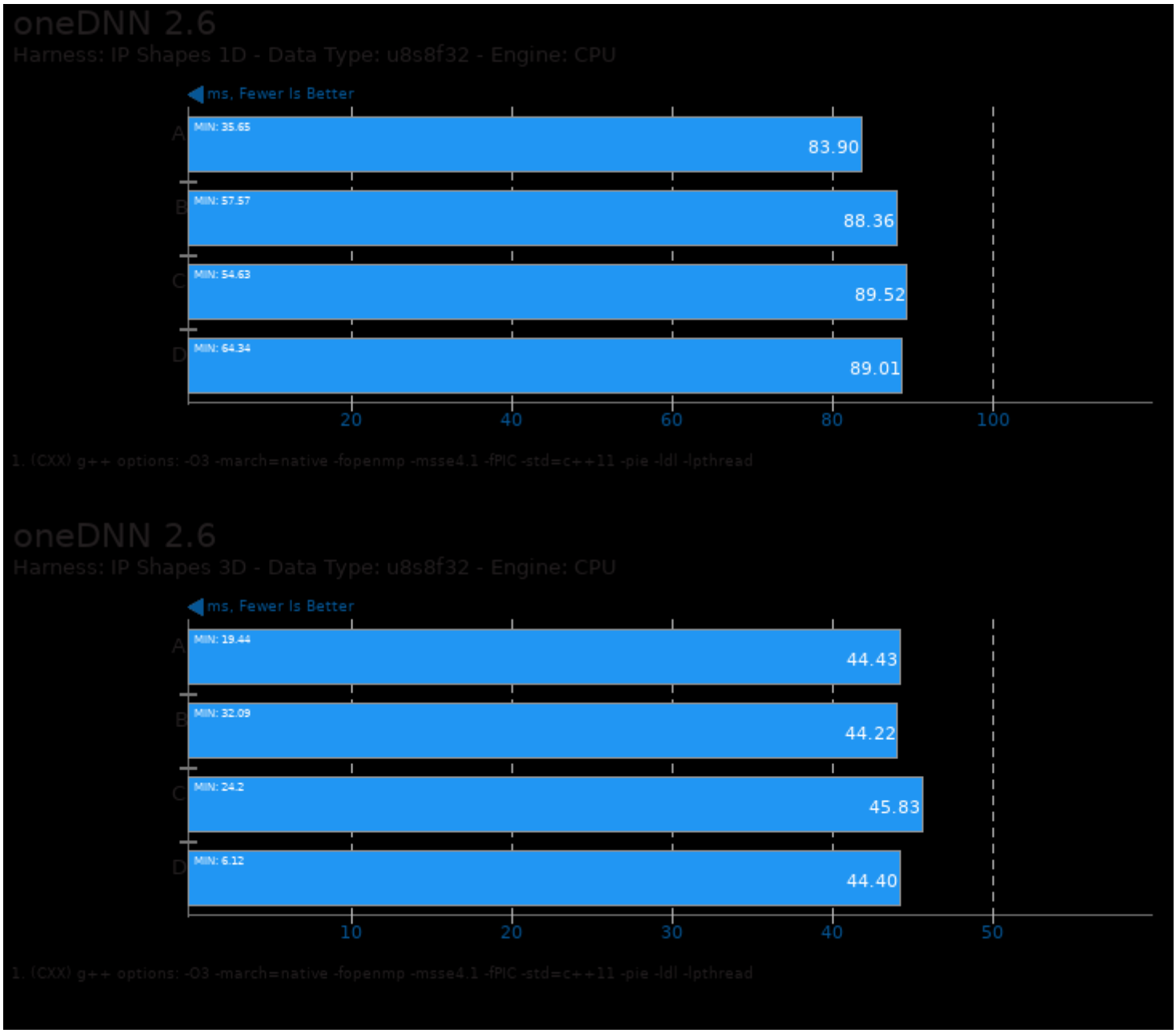
10980XE onednn onnx



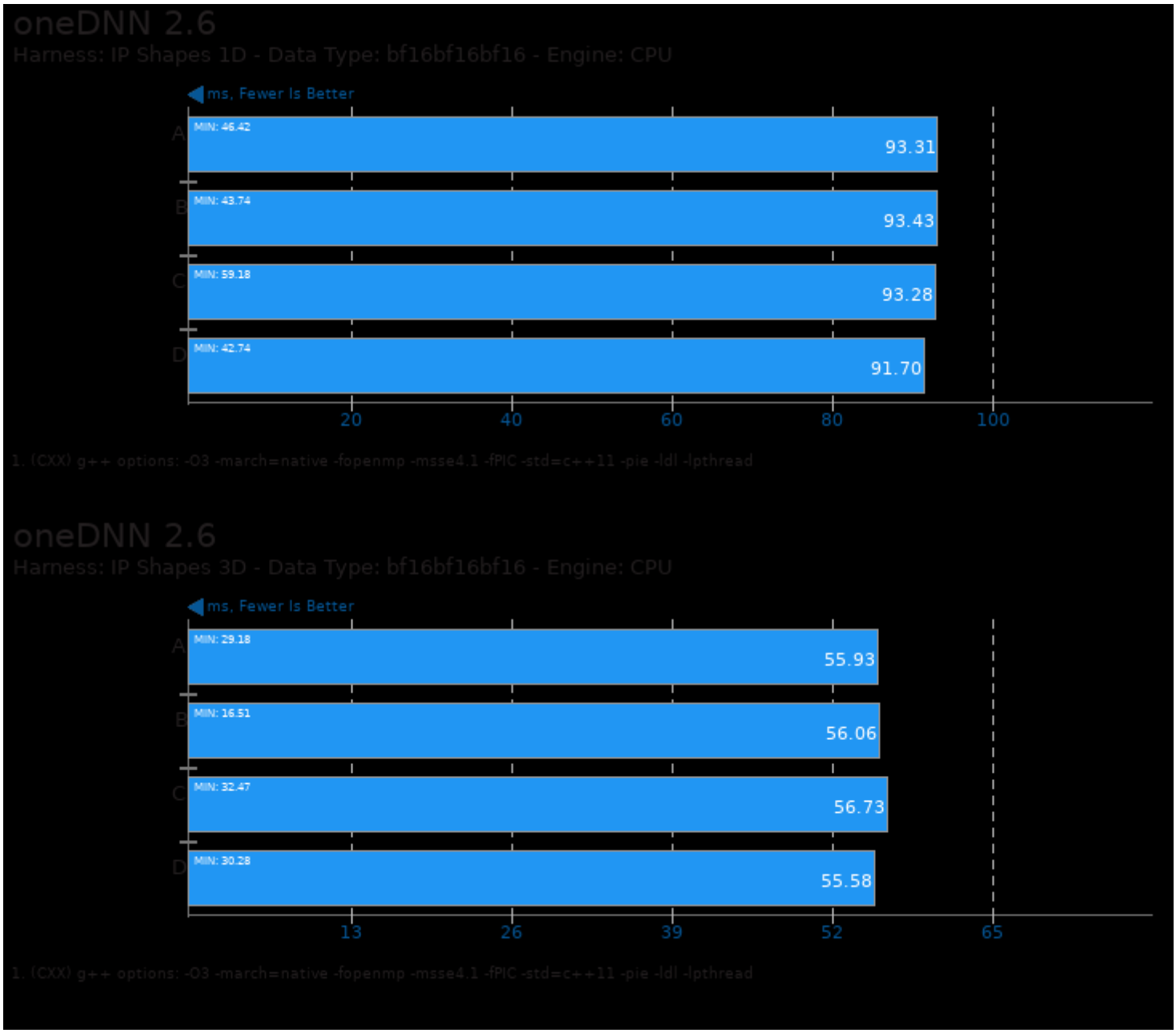
10980XE onednn onnx



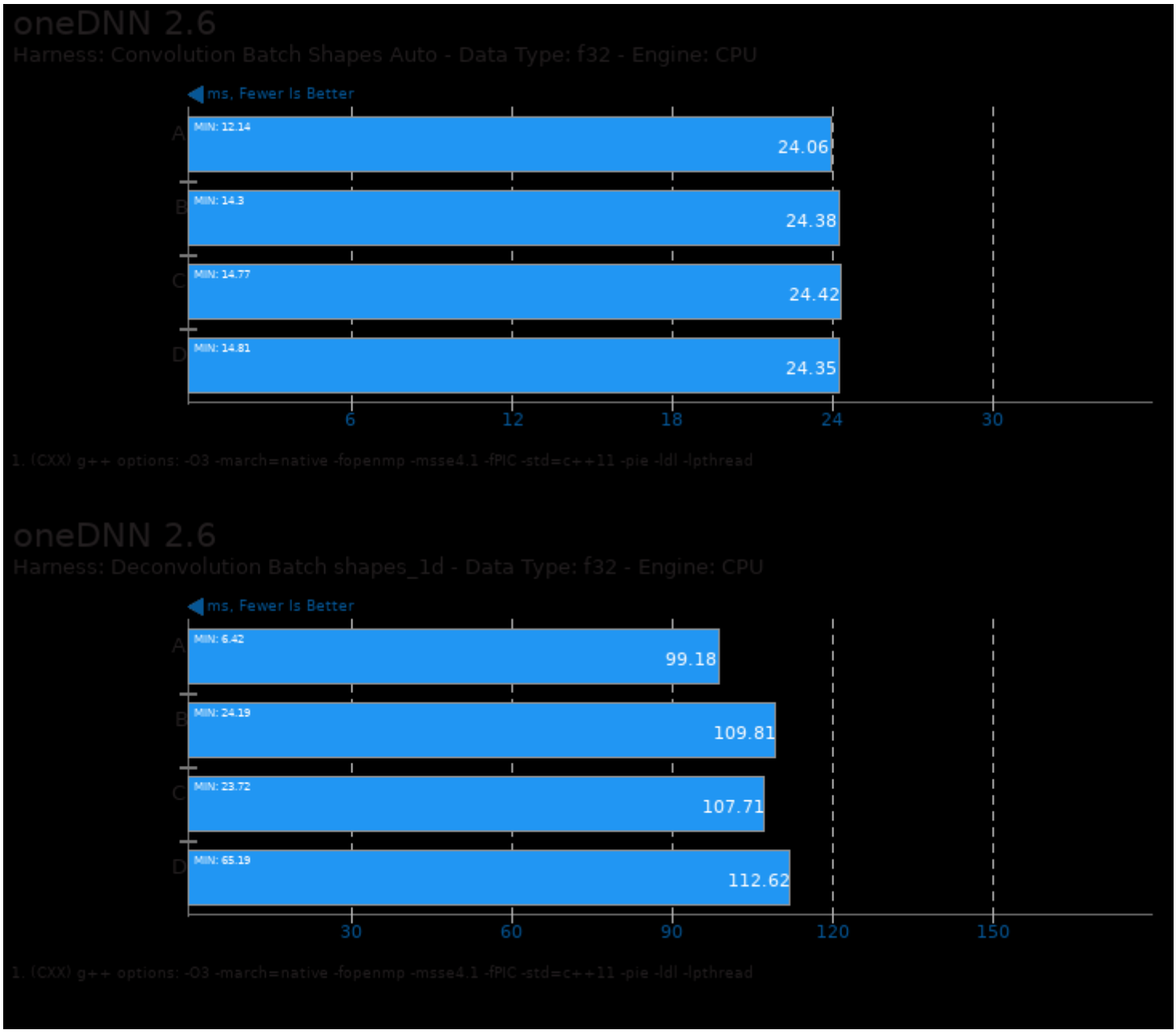
10980XE onednn onnx



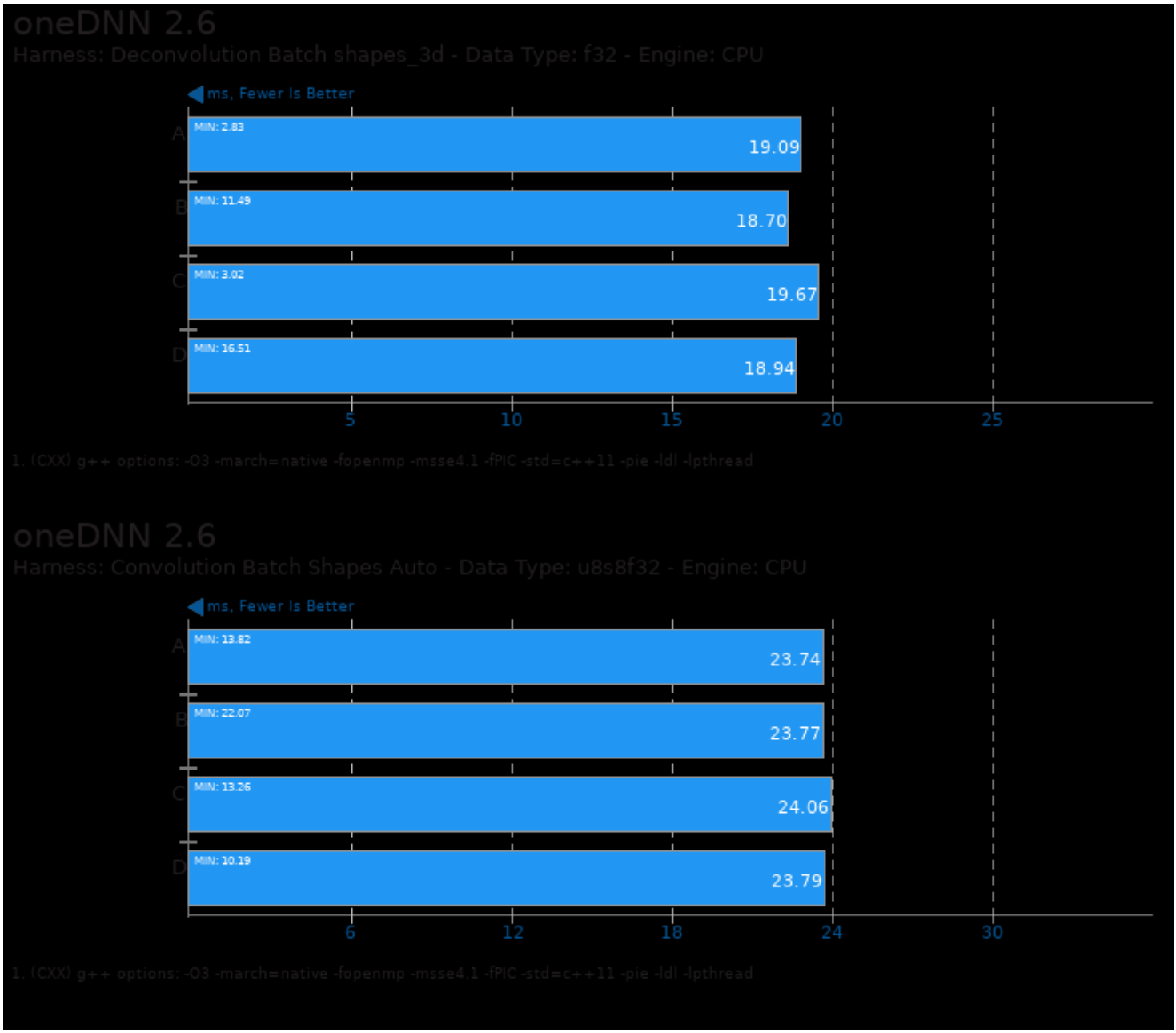
10980XE onednn onnx



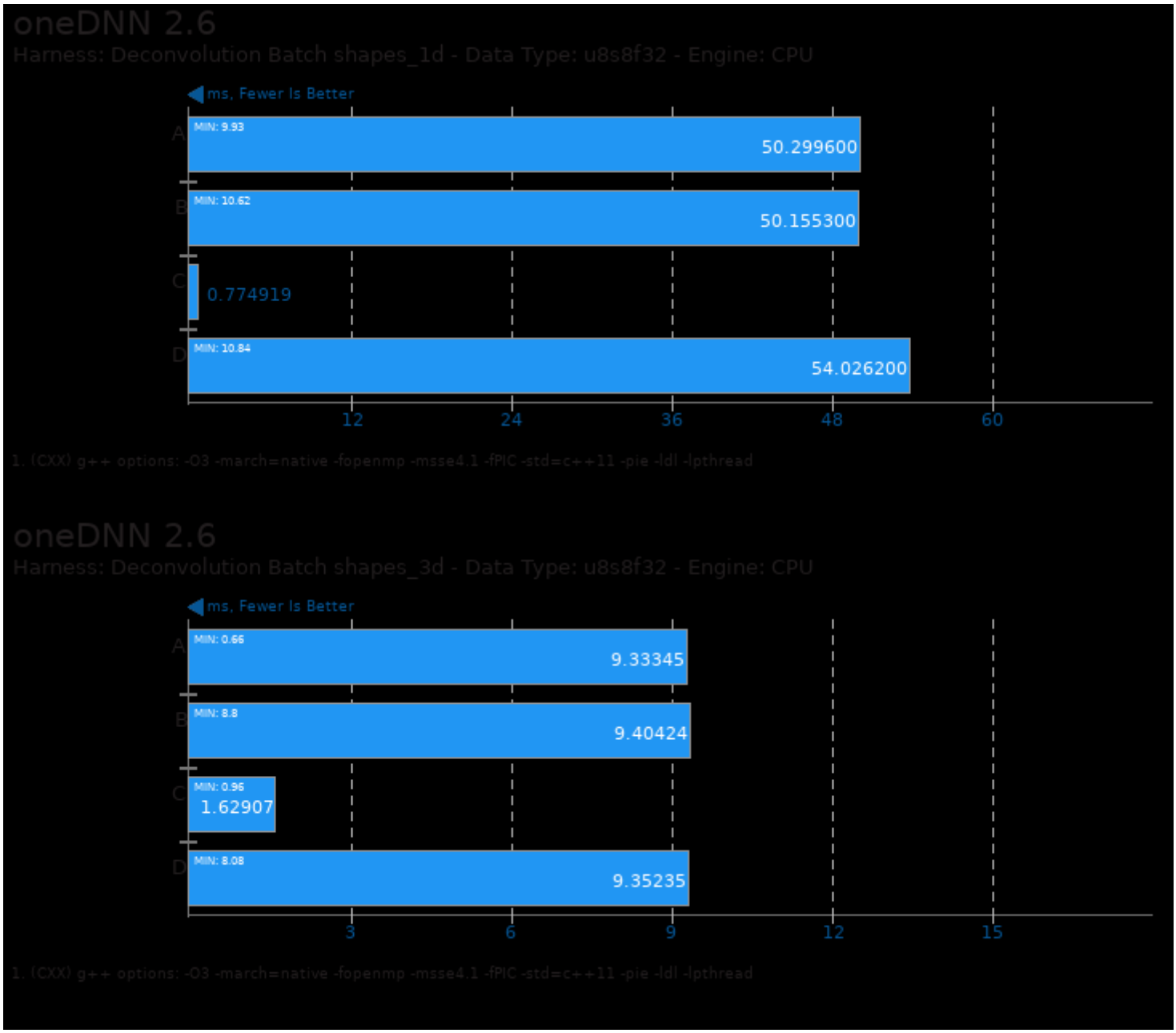
10980XE onednn onnx



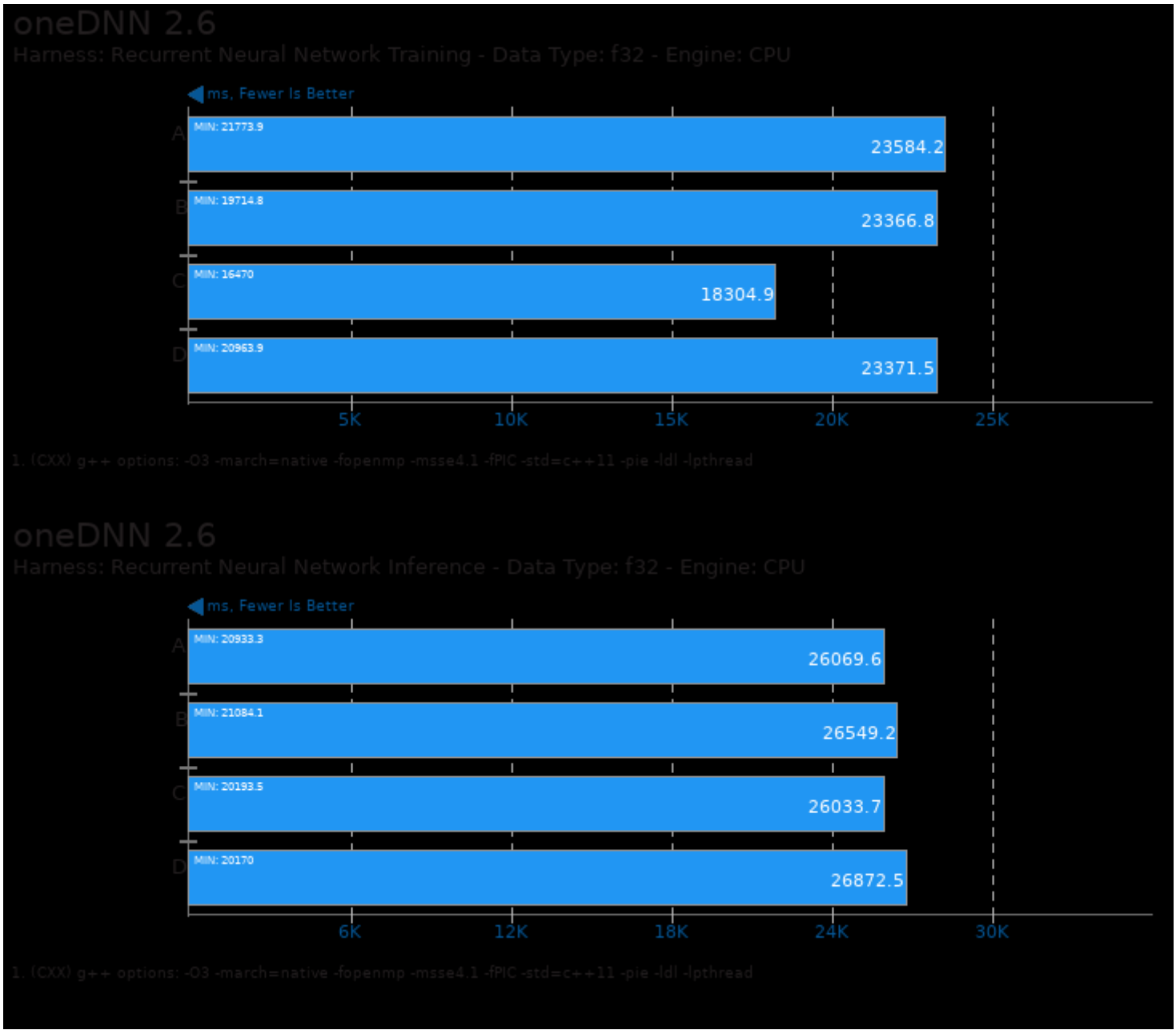
10980XE onednn onnx



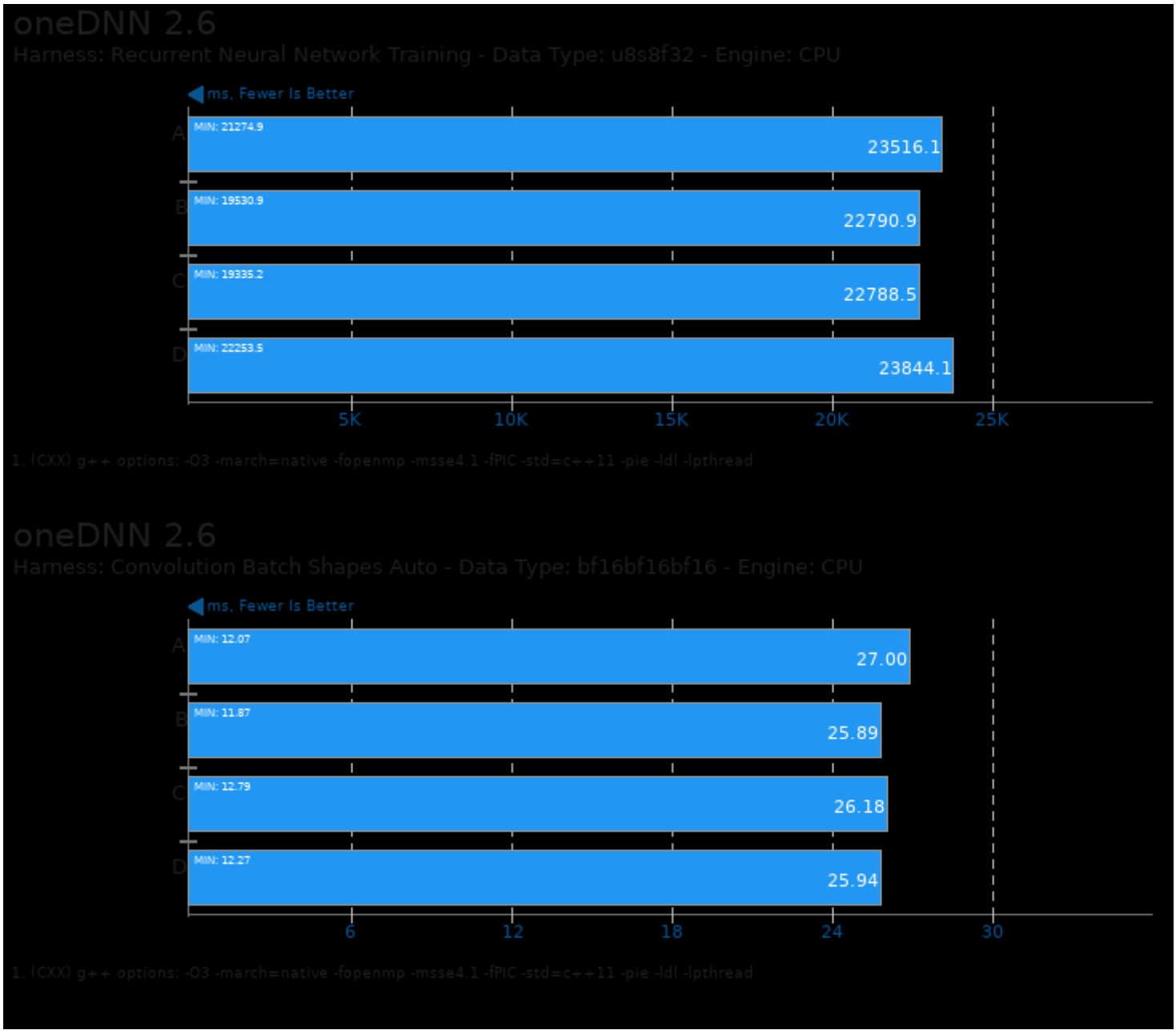
10980XE onednn onnx



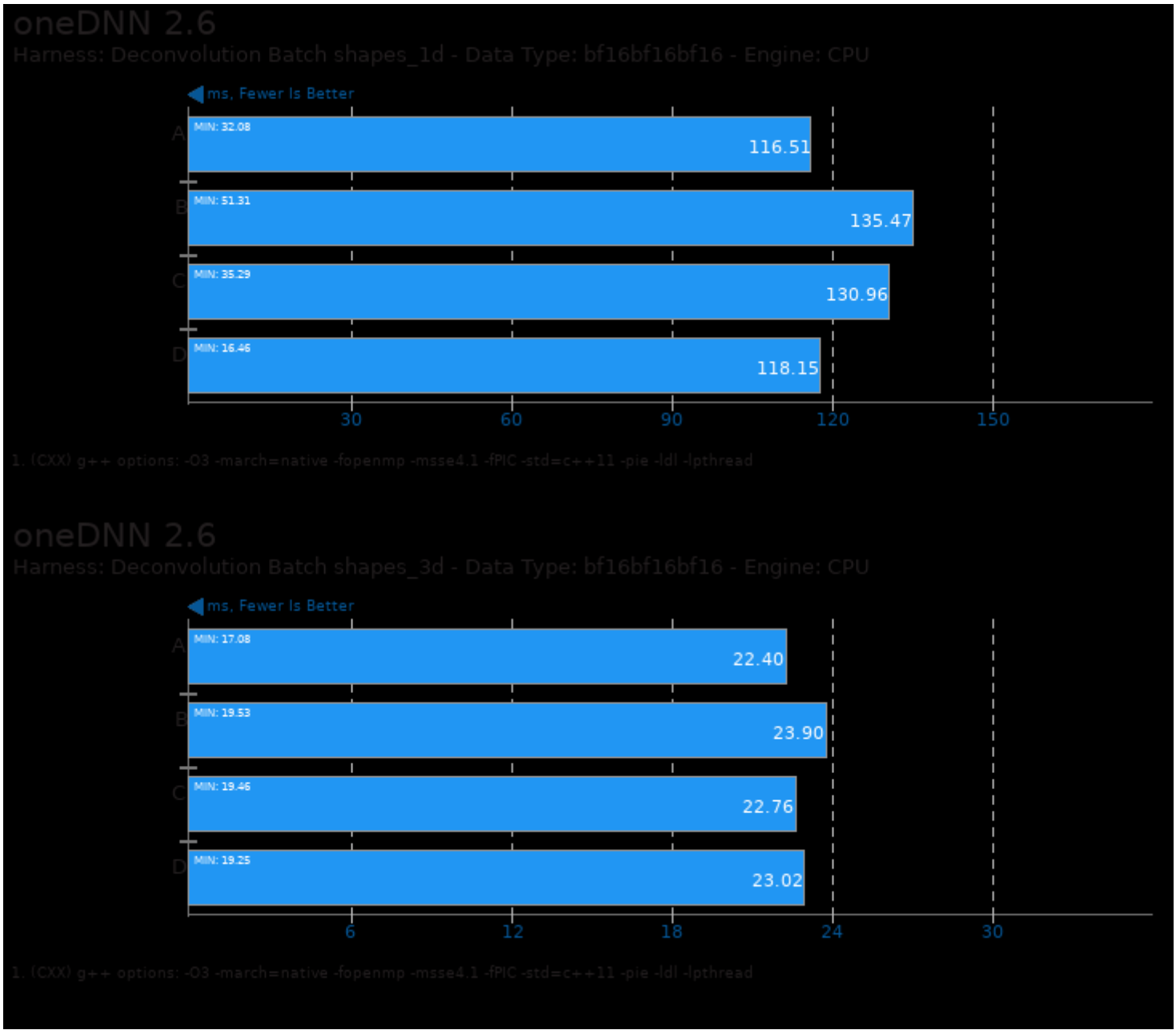
10980XE onednn onnx



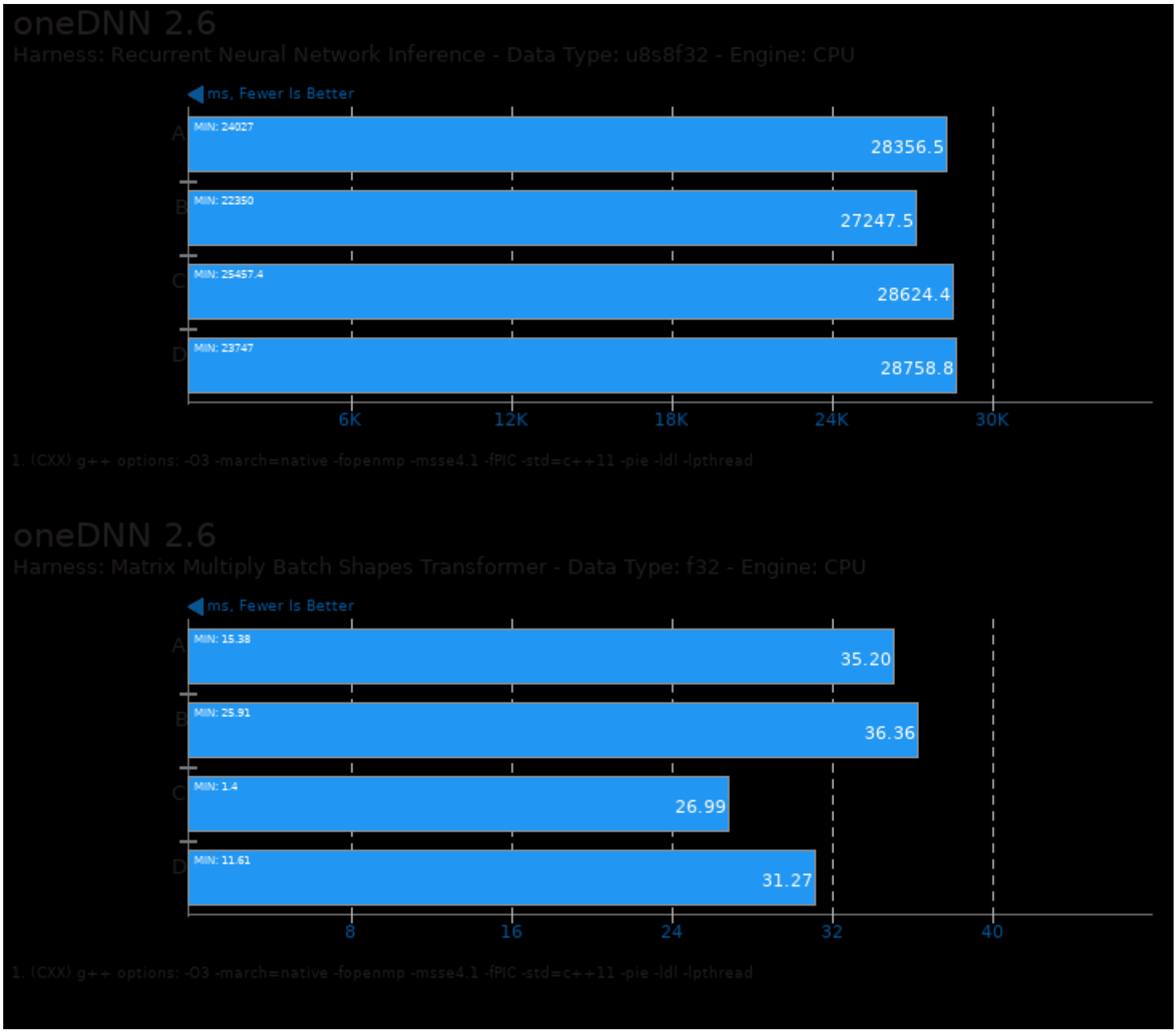
10980XE onednn onnx



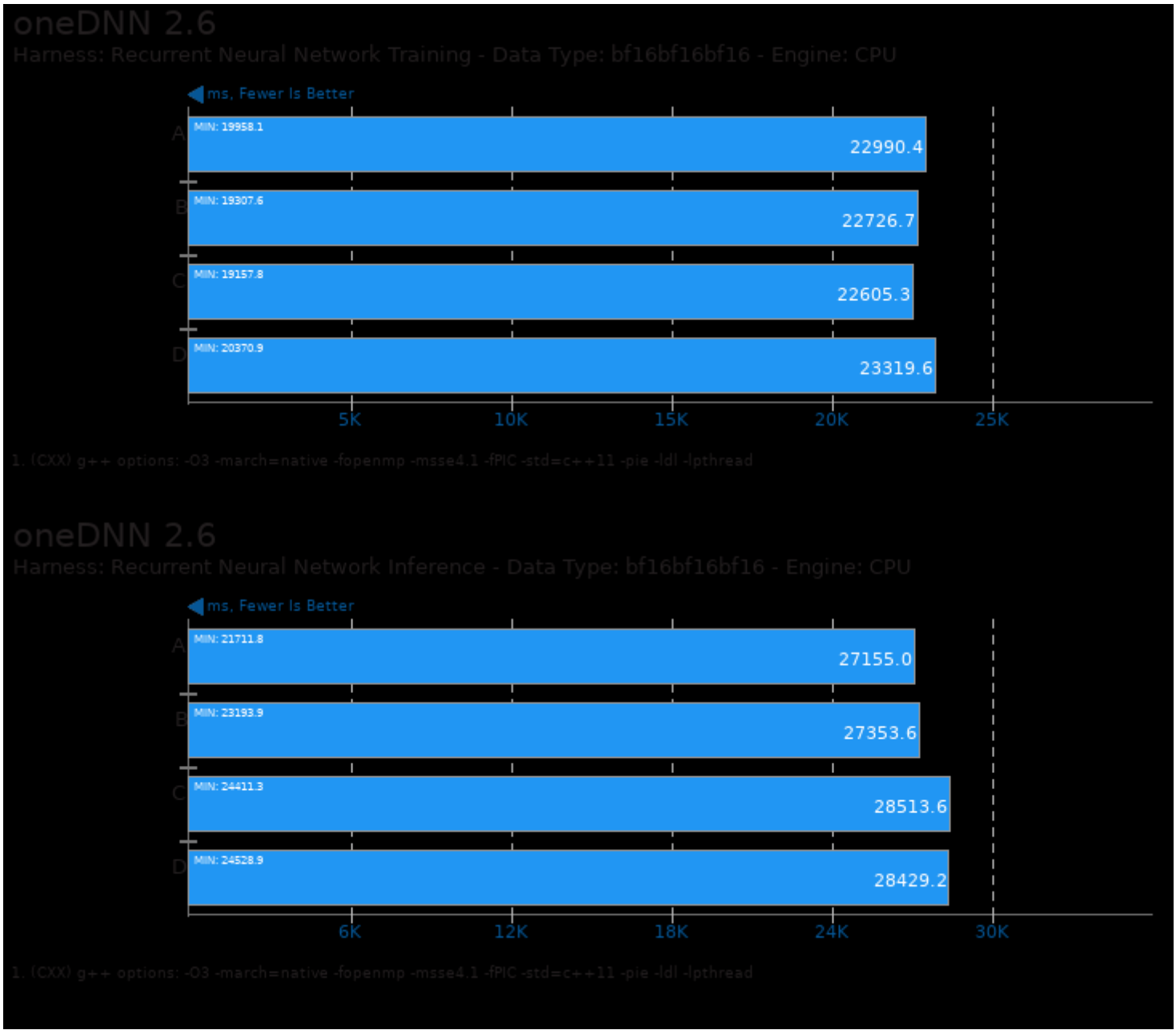
10980XE onednn onnx



10980XE onednn onnx

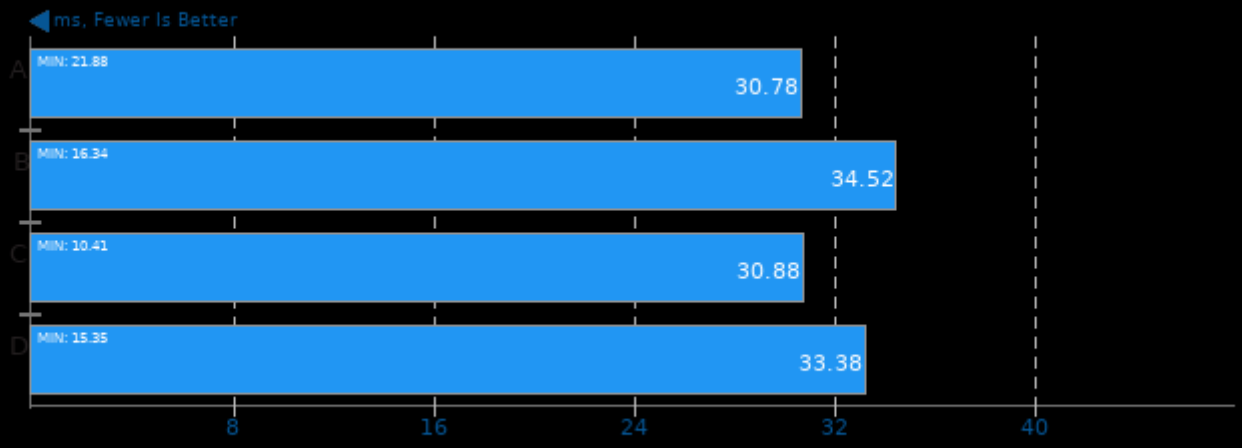


10980XE onednn onnx



oneDNN 2.6

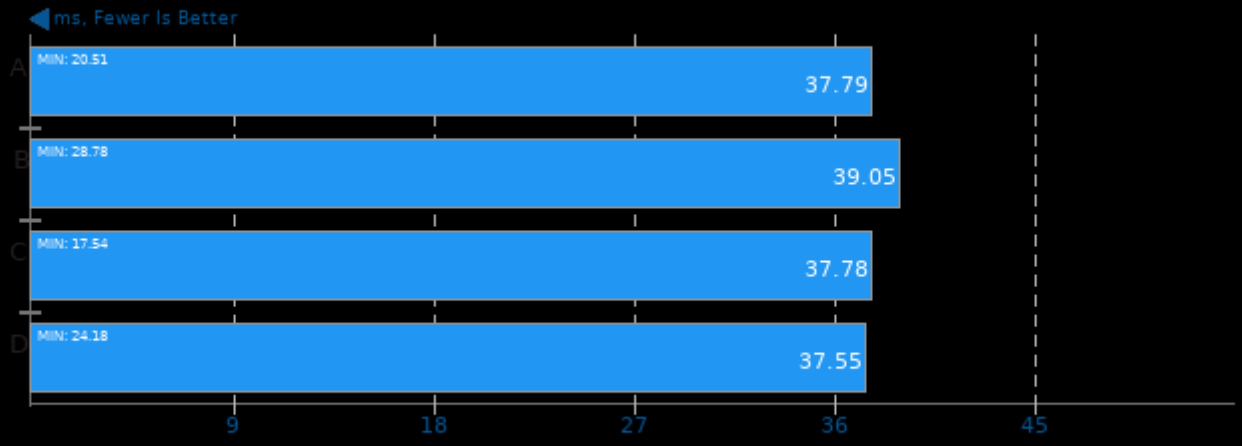
Harness: Matrix Multiply Batch Shapes Transformer - Data Type: u8s8f32 - Engine: CPU



1. (CXX) g++ options: -O3 -march=native -fopenmp -msse4.1 -fPIC -std=c++11 -pie -ldl -lpthread

oneDNN 2.6

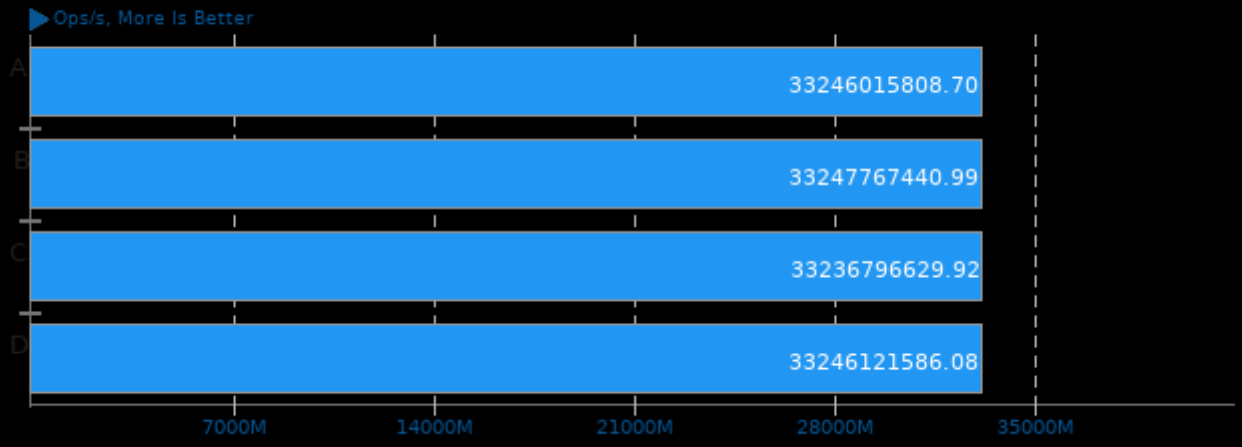
Harness: Matrix Multiply Batch Shapes Transformer - Data Type: bf16bf16bf16 - Engine: CPU



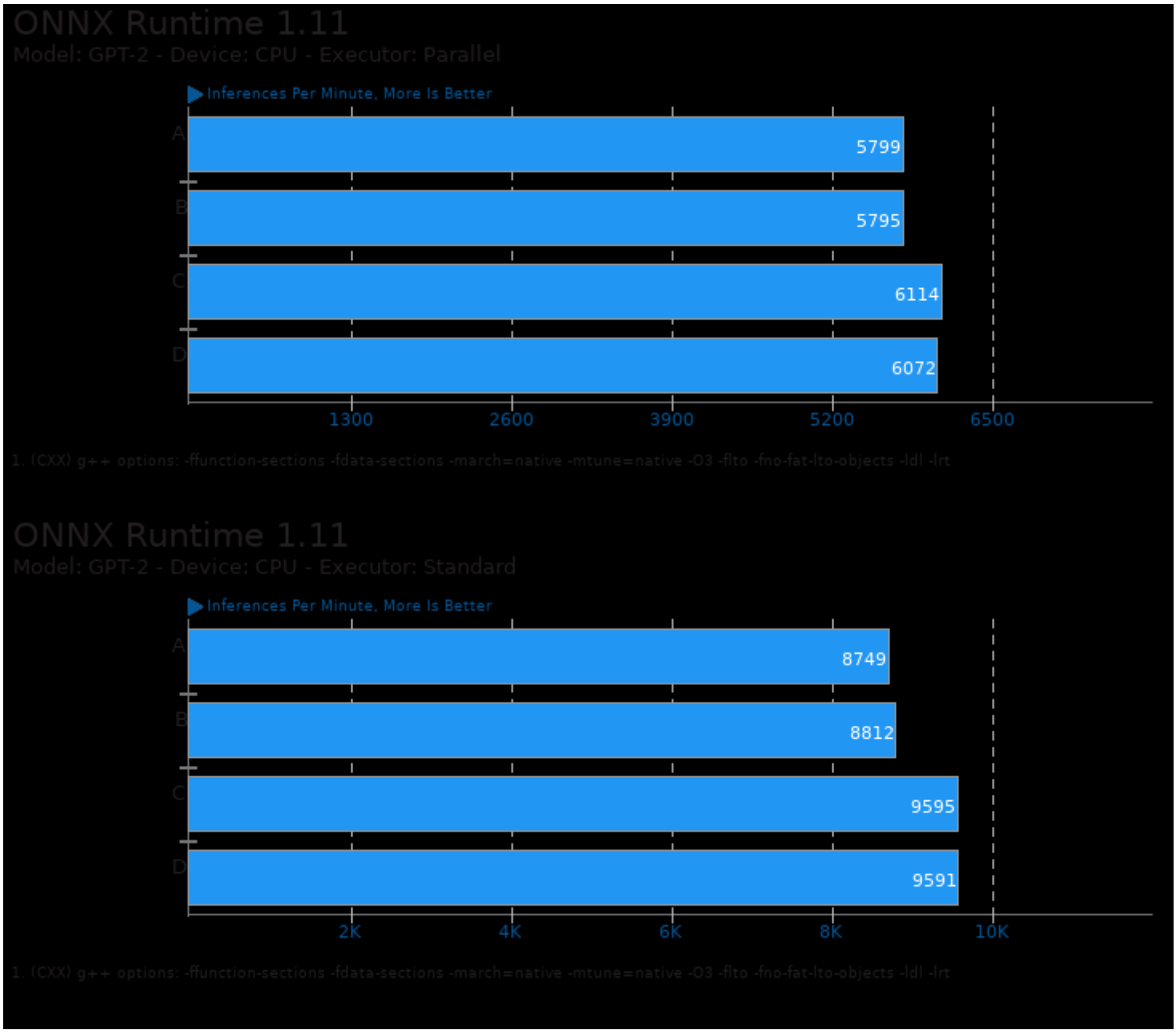
1. (CXX) g++ options: -O3 -march=native -fopenmp -msse4.1 -fPIC -std=c++11 -pie -ldl -lpthread

Java JMH

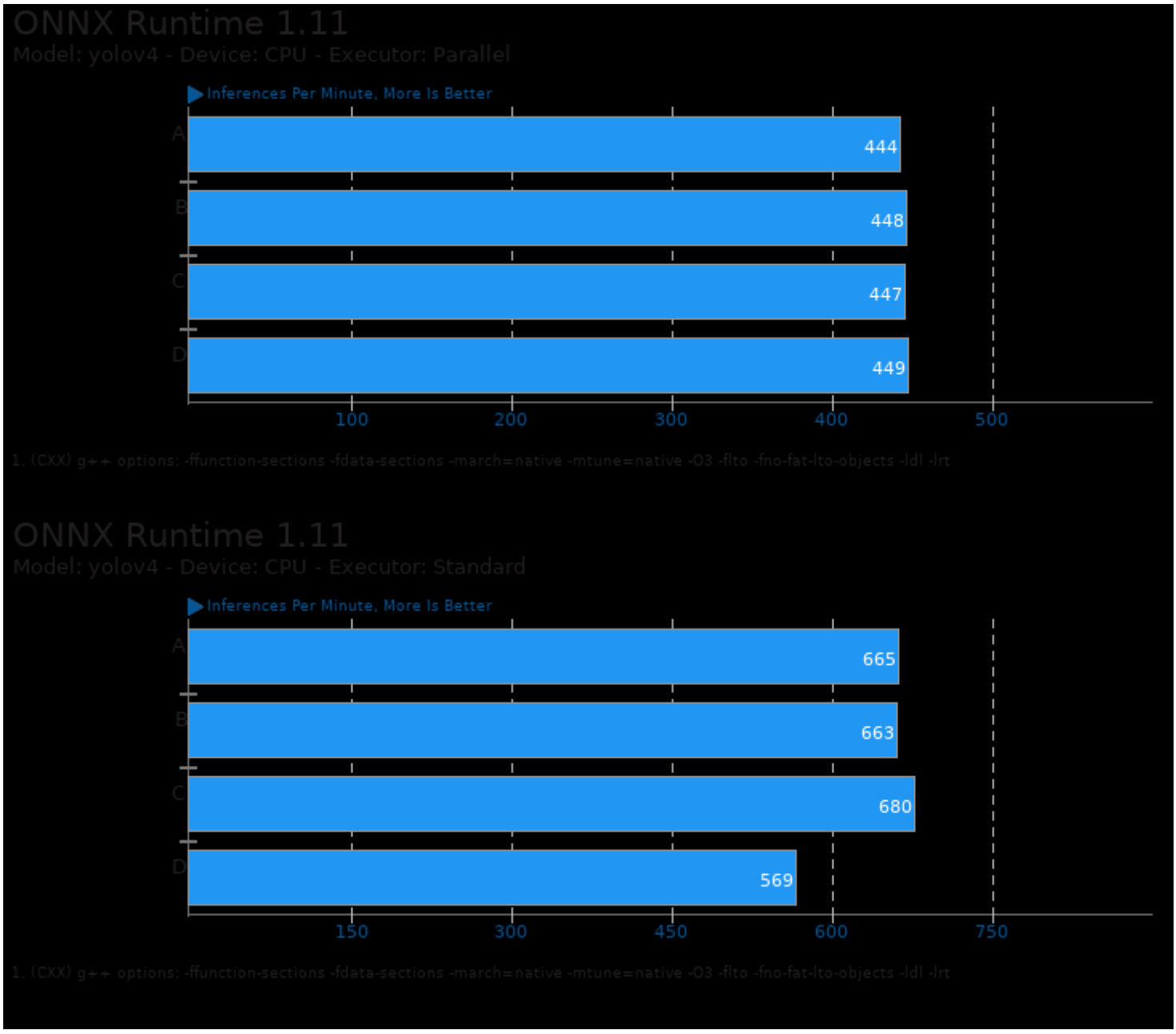
Throughput



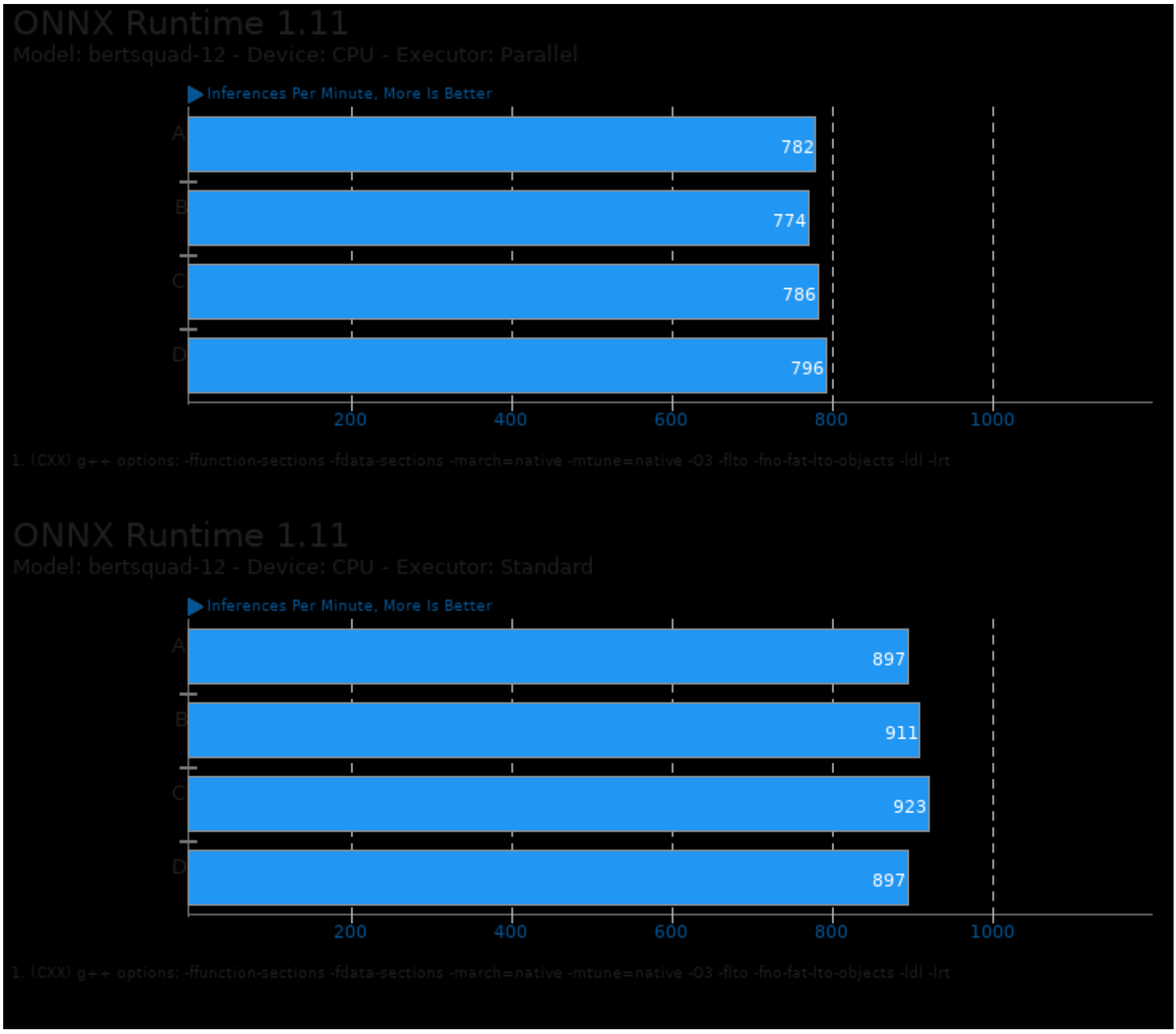
10980XE onednn onnx



10980XE onednn onnx



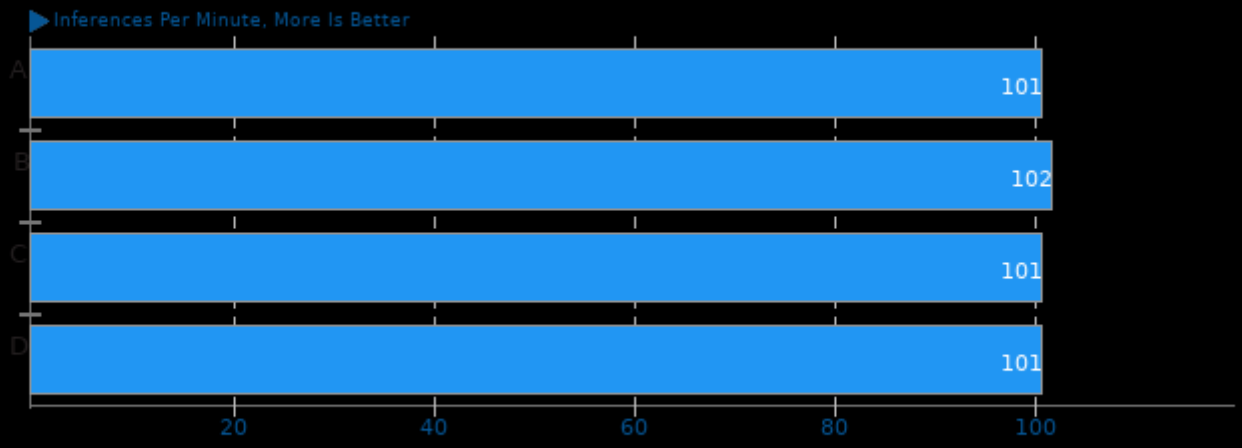
10980XE onednn onnx



10980XE onednn onnx

ONNX Runtime 1.11

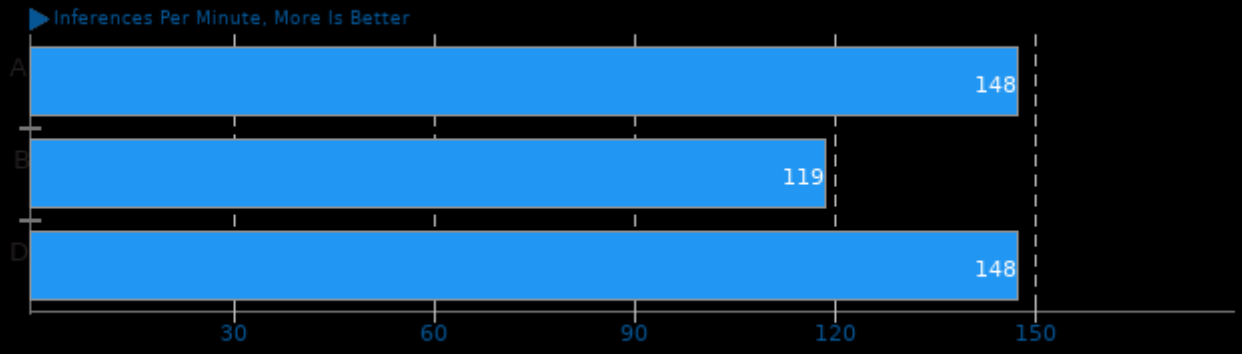
Model: fcn-resnet101-11 - Device: CPU - Executor: Parallel



1. (CXX) g++ options: -ffunction-sections -fdata-sections -march=native -mtune=native -O3 -flto -fno-fat-lto-objects -ldl -lrt

ONNX Runtime 1.11

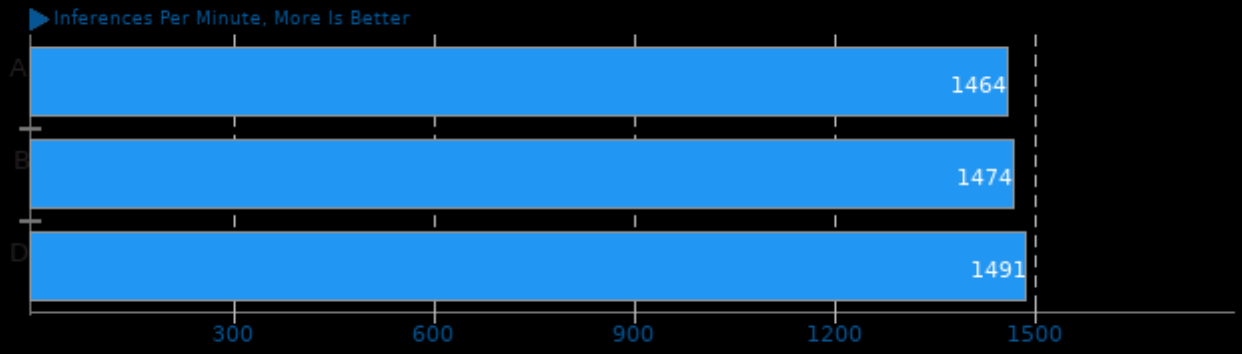
Model: fcn-resnet101-11 - Device: CPU - Executor: Standard



1. (CXX) g++ options: -ffunction-sections -fdata-sections -march=native -mtune=native -O3 -flto -fno-fat-lto-objects -ldl -lrt

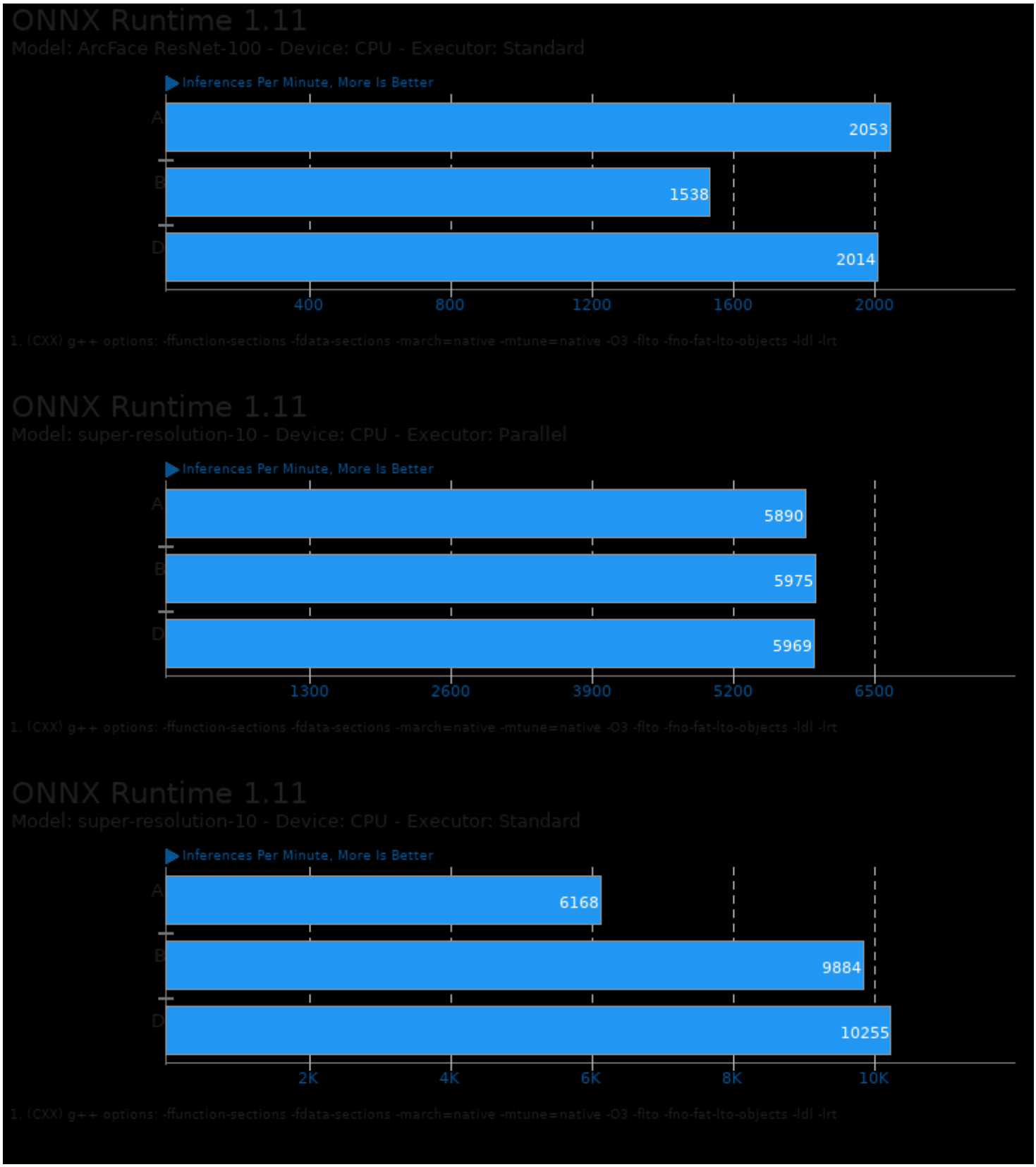
ONNX Runtime 1.11

Model: ArcFace ResNet-100 - Device: CPU - Executor: Parallel



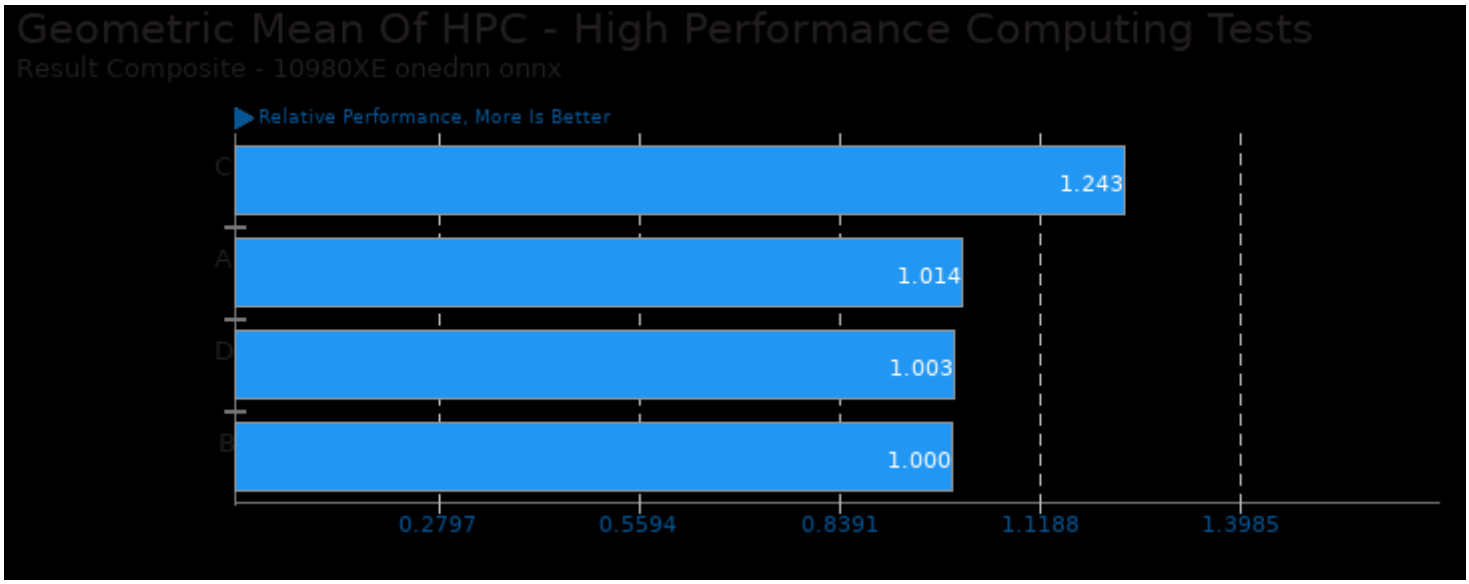
1. (CXX) g++ options: -ffunction-sections -fdata-sections -march=native -mtune=native -O3 -flto -fno-fat-lto-objects -ldl -lrt

10980XE onednn onnx

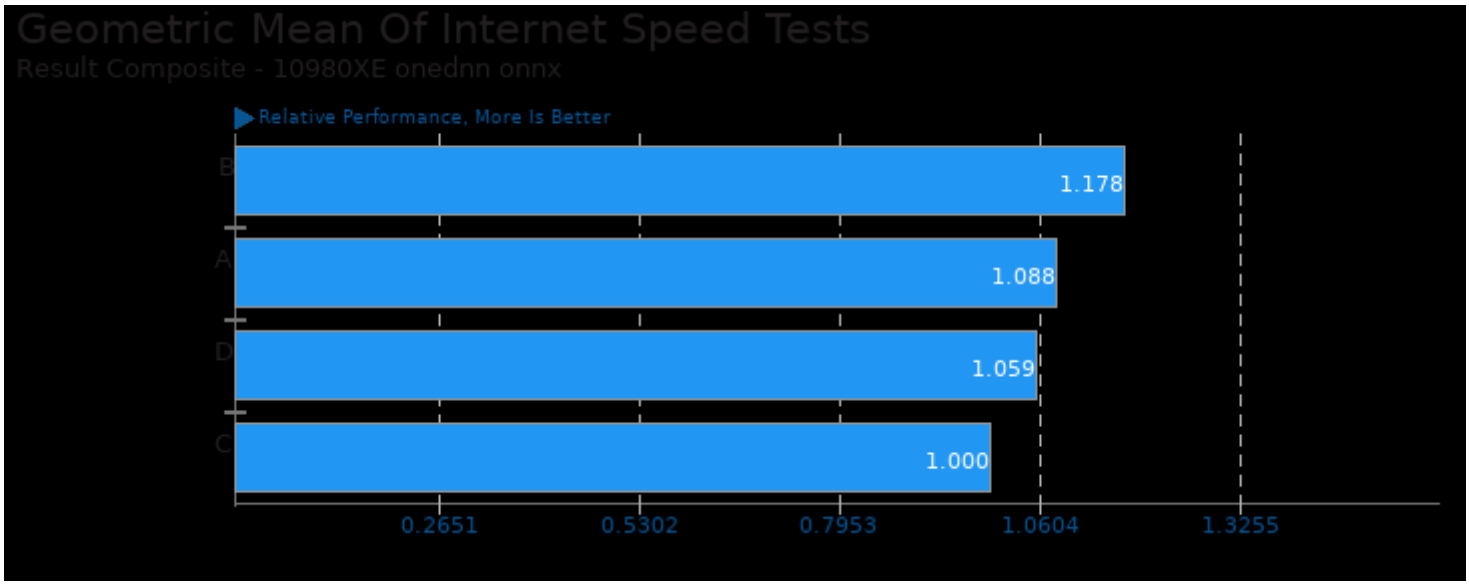


10980XE onednn onnx

These geometric means are based upon test groupings / test suites for this result file.

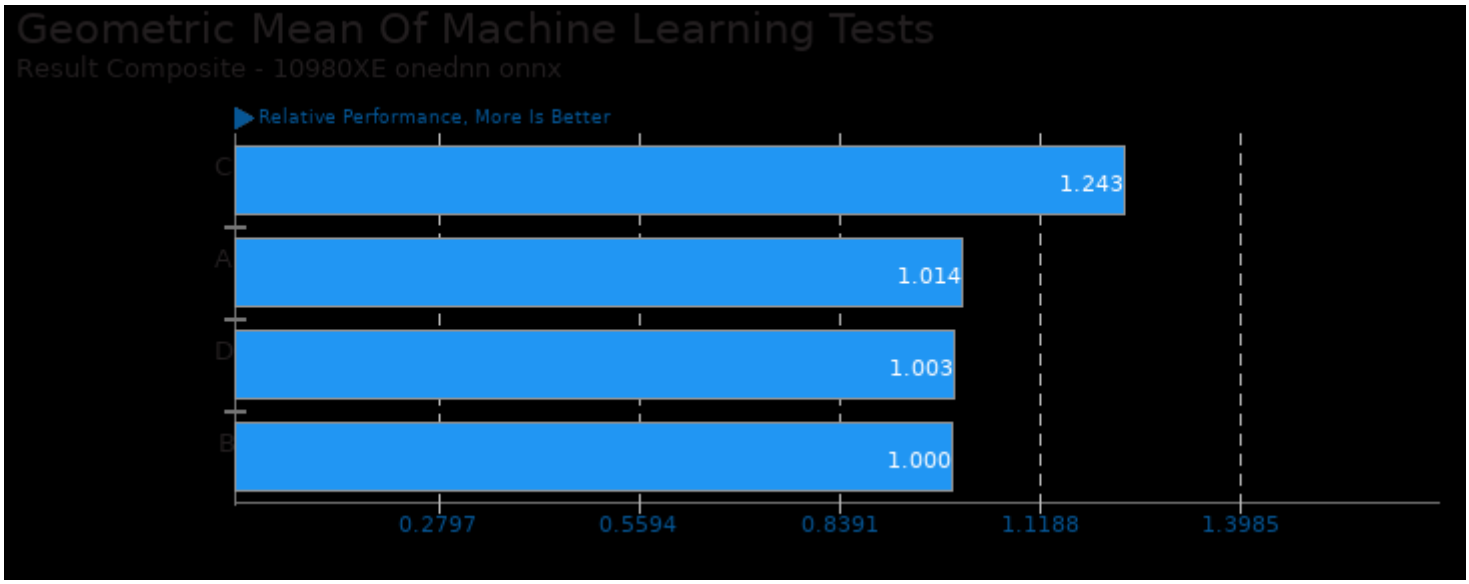


Geometric mean based upon tests: pts/onednn and pts/onnx

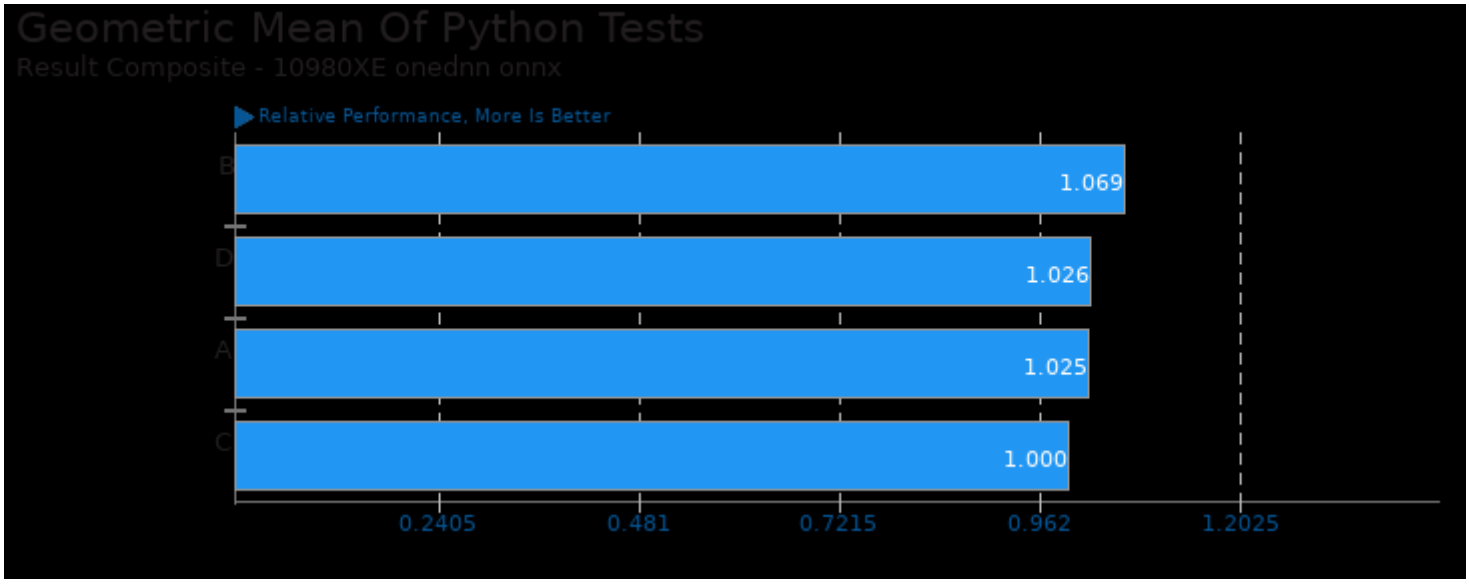


Geometric mean based upon tests: pts/fast-cli and pts/speedtest-cli

10980XE onednn onnx



Geometric mean based upon tests: pts/onednn and pts/onnx



Geometric mean based upon tests: pts/speedtest-cli and pts/onnx

This file was automatically generated via the Phoronix Test Suite benchmarking software on Thursday, 28 March 2024 12:50.