



Myrtle.ai

# VOLLO

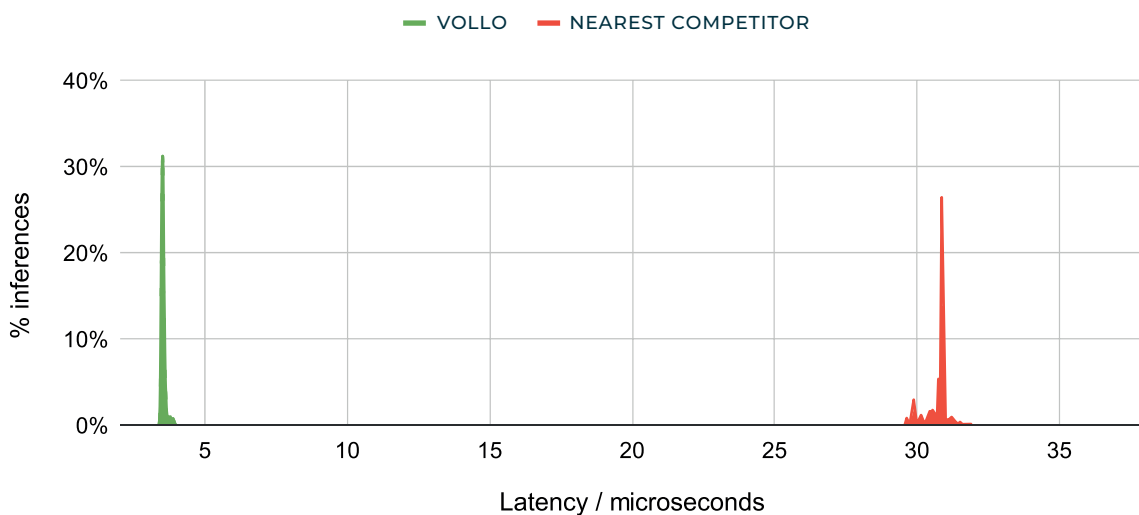
## LOWEST LATENCY ML INFERENCE FOR TRADING DECISIONS

**RUN YOUR AI  
MODELS IN  
UNDER 10 $\mu$ s**

- Unrivalled STAC-ML latency<sup>1</sup>
- Trusted by leading firms
- Simple to evaluate and deploy
- High throughput

Evaluate your future  
performance today  
using the tools at  
[vollo.myrtle.ai](https://vollo.myrtle.ai)

LATENCY COMPARISON WITH NEAREST COMPETITOR<sup>2</sup>



## Development environment:

- Work within your existing framework
- Train in PyTorch or TensorFlow, export in ONNX
- Change models in seconds



Quickly check how fast your models would run on Vollo  
Visit [vollo.myrtle.ai](https://vollo.myrtle.ai) | No new hardware required

## Deployment options:

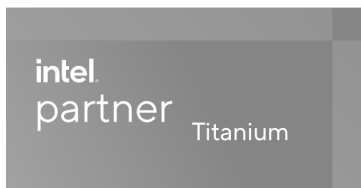
- Standard FHFL & HHHH PCIe cards
- FPGA netlist for integration into existing systems

### STAC-ML™ AUDITED PERFORMANCE

	SUMACO TEST SUITE	TACANA TEST SUITE
99th Percentile Latencies	<ul style="list-style-type: none"><li>• 24.1 <math>\mu</math>s for LSTM_A (smallest model tested)</li><li>• 64.8 <math>\mu</math>sec for LSTM_B</li><li>• 1.35 ms for LSTM_C (largest model tested)</li></ul> <small>STAC-ML.Markets.Inf.S.LSTM_A[, 1, 2, 3, 4].LAT.v1 STAC-ML.Markets.Inf.S.LSTM_B[, 1, 2, 3, 4].LAT.v1 STAC-ML.Markets.Inf.S.LSTM_C[, 1, 2, 3, 4].LAT.v1</small>	<ul style="list-style-type: none"><li>• 5.07 <math>\mu</math>s for LSTM_A (smallest mode tested)</li><li>• 6.77 <math>\mu</math>s for LSTM_B</li><li>• 31.0 <math>\mu</math>s for LSTM_C (largest mode tested)</li></ul> <small>STAC-ML.Markets.Inf.T.LSTM_A[, 1, 2].LAT.v1 STAC-ML.Markets.Inf.T.LSTM_B.2.LAT.v1 STAC-ML.Markets.Inf.T.LSTM_C.LAT.v1</small>
Throughput	Throughput exceeded 650K inf/sec for LSTM_A with 48 NMI <small>STAC-ML.Markets.Inf.S.LSTM_A.48.TPUT.v1</small>	Throughput exceeded 14 M inf/sec for LSTM_A with 24 NMI <small>STAC-ML.Markets.Inf.T.LSTM_A.24.TPUT.v1</small>
Space Efficiency	Space efficiency exceeded 646K inf/sec/cubic foot for LSTM_A with 48 NMI <small>STAC-ML.Markets.Inf.S.LSTM_A.48.SPACE_EFF.v1</small>	Space efficiency exceeded 1.4M inf/sec/cubic foot for LSTM_A with 24 NMI <small>STAC-ML.Markets.Inf.T.LSTM_A.24.SPACE_EFF.v1</small>
Energy Efficiency	Energy efficiency exceeded 1.18M inf/sec/kW for LSTM_A with 48 NMI <small>STAC-ML.Markets.Inf.S.LSTM_A.48.ENERG_EFF.v1</small>	Energy efficiency exceeded 2.32M inf/sec/kW for LSTM_A with 24 NMI <small>STAC-ML.Markets.Inf.T.LSTM_A.24.ENERG_EFF.v1</small>

<sup>1</sup> Audited by the STAC Research Council, the finance industry's own benchmarking organization

<sup>2</sup> Latency model testing by Myrtle.ai on proprietary models



"STAC" and all STAC™ related names are trademarks or registered trademarks of the Strategic Technology Analysis Center LLC. Only audited results with a STAC-ML code starting STAC-ML.Markets.Inf. are official. Full independent STAC-ML audit results are only available to STAC members. Other companies' STAC audit results may only be available to premium STAC finance subscribers. Comparisons in this document that are not between STAC-ML.Markets.Inf coded data will vary from system to system. All purchasing decisions should only be made after direct assessment of the product within the official Vollo Evaluation Program. Vollo Accelerator and the VOLLO logo are trademarks of Myrtle.ai. PyTorch, the PyTorch logo and any related marks are trademarks of Facebook, Inc. Other names and brands may be claimed as the property of others. Copyright © Myrtle.ai 2024. Rev 2024.4.12

Full product  
information at  
[myrtle.ai/vollo](https://myrtle.ai/vollo)



Myrtle.ai