STAC Update:
Big Workloads

Bishop Brock
Head of Research, STAC
bishop.brock@STACresearch.com

Peter Nabicht
President, STAC
peter.nabicht@STACresearch.com

Overview

- STAC-A3 (backtesting)
- STAC-ML (training)
- STAC-M3 (tick history / timeseries analysis)

STAC-A3

(backtesting)

STAC-A3

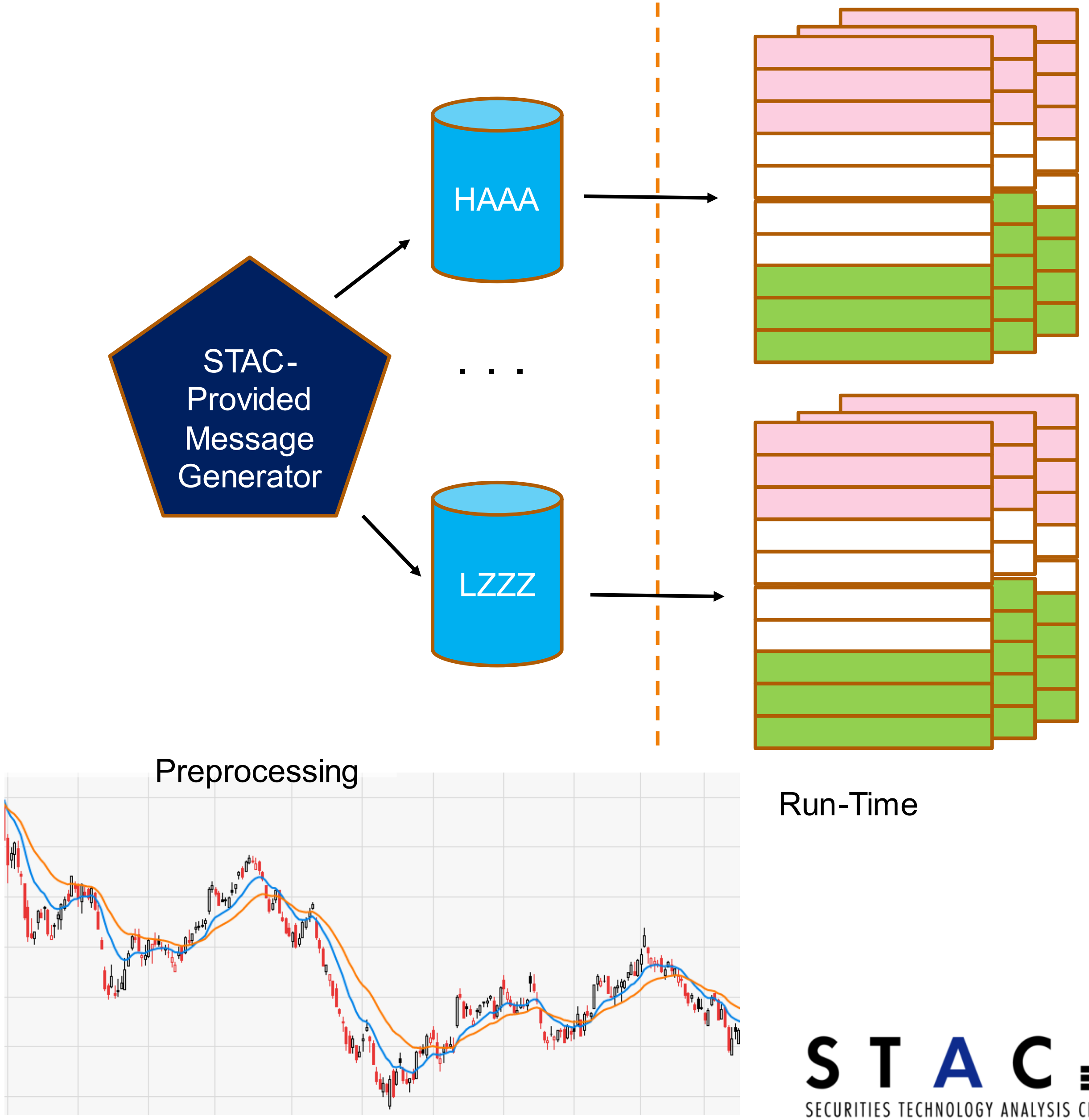
- Overview:
  - Workloads that emulate real-world backtesting jobs
  - Measures speed, throughput, scalability, efficiency of any architecture
- Architectures: scale-up and scale-out, cloud and bare metal
- Languages: Python, C++, Scala
- Hardware: CPU, GPU
- Different approaches to parallelization and optimizations are very informative
- *Currently looking to modify current / add new STAC-A3 benchmark suites*

www.stacresearch.com/A3

STAC® SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-A3: Backtesting Benchmark – Current State

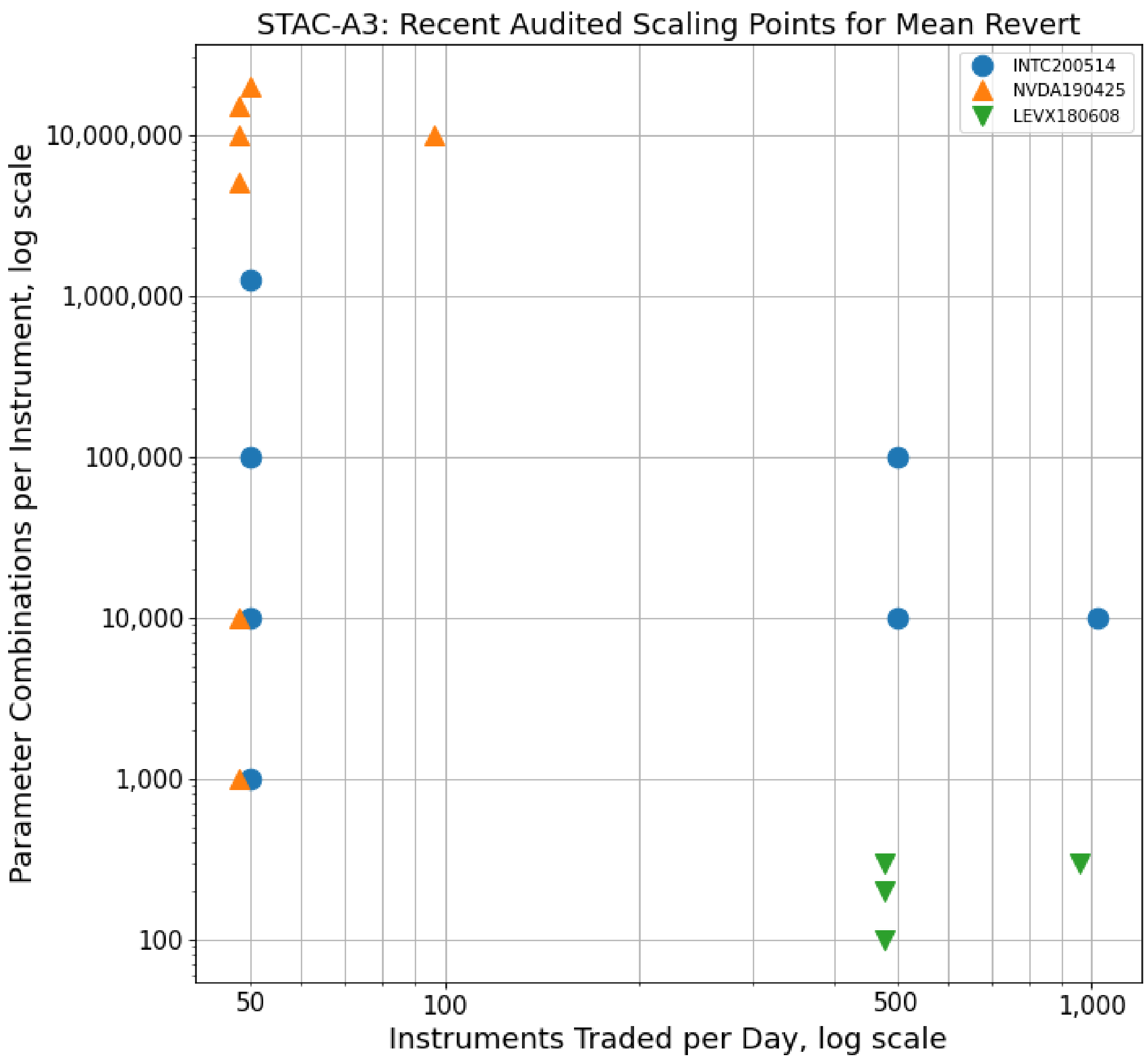
- STAC-A3 Mean Revert suite
  - parameterized technical strategy for intraday trading
  - portfolio of instruments
  - Simulated market data provided at the order level
    - Both high- and low-volume instruments are represented
  - SUT must build the order book
  - Strategy evaluated every second
  - Sweeps over strategy parameters
  - Low to moderate I/O
  - Branch-heavy computation



STAC SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-A3 Mean Revert: Scaling

- Scaling dimensions
  - Number of instruments traded each day
  - Number of parameter combinations per instrument
- Values currently chosen by testing vendor
- Looking to standardize (some) scaling points
  - What values are impactful to you?



STAC-A3 BLASH (Proposal); Options (Proposal)

- BLASH (Buy-Low/Sell-High) is an incomplete, never-implemented spec.
- BLASH was designed to extend the computational complexity of STAC-A3 by addressing a couple of aspects of Mean-Revert:
  - Mean-Revert is embarrassingly parallel by parameter set and symbol
  - Mean-Revert only looks at the top-of-book on either side
- However, analysis showed that BLASH implementations:
  - Would still admit very high or unrealistic parallelism
  - Could use “tricks” to short-circuit backtesting, and sometimes eliminate the order-book analysis
- A proposal was also floated to modify Mean-Revert for options:
  - Trade options and compute Greeks at the same Mean-Revert signal points

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

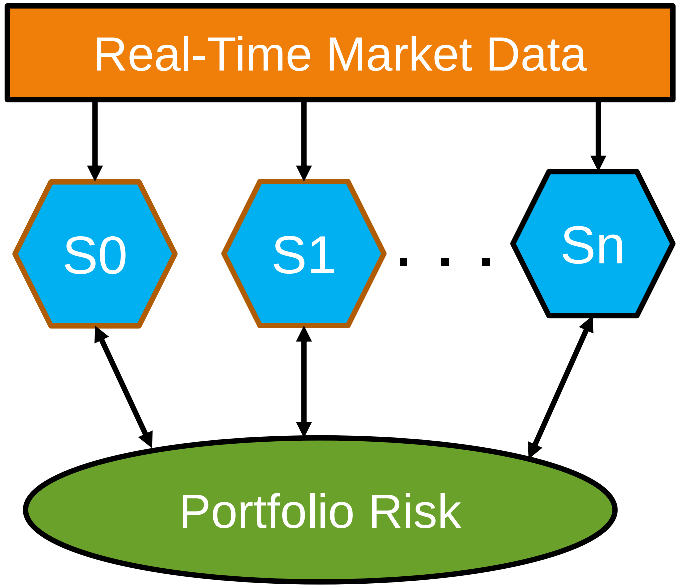
STAC-A3: Next Steps

- STAC has been working on a new “Portfolio Trading Suite” proposal
- Factors considered:

Mean-Revert suite	Portfolio trading suite proposal
Embarrassingly parallel	Fine grained interlock
Low # of instruments = low IO	Portfolio evaluation requires high IO for all instrument scales
Strategy / parameterization has low complexity (low compute)	Strategy elements are more complex
No position sizing	Portfolio optimization

STAC-A3 Portfolio Trading (Proposal): High Level View

- The proposal models a set of independent strategies S0, … Sn, each reacting to real-time market data, but coordinating trading activity through portfolio risk mechanism.
- Coordination could be daily (easier) or intraday (harder).
- Strategy instruments change daily on conditions TBD.
- Options calculations are still on the table



Working Group discussions underway - **Join us!**

www.STACresearch.com/A3

STAC
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-ML

(training)

STAC-ML Markets (Training) Benchmark : Underway

- Existing ML training benchmarks are not *specific* to Finance:
  - They typically focus on <u>categorical</u> decisions (e.g., most probable next word)
  - Finance often requires <u>quantitative</u> models (e.g., fair value of a derivative)
- Finance use cases may require training many, many models
  - Historical backtesting may involve models specific to points in time
  - This becomes a scale-out problem vs. scale-up (e.g., LLM training)
- Many use cases have been proposed and discussed, but may not satisfy all high-level requirements:
  - Is this an ongoing concern for many end-users?
  - Can performance and quality be reliably measured and compared?
  - Can we validate that the implementation conforms to the specifications?

Some ML Training Use Cases Being Considered

Model Type / Use case	Issues / Notes
Predict prices/returns/portfolio-weights from market data	• Obviously interesting use cases • Training / re-training very important ! Low signal-noise means models learn quickly and erratically – difficult to benchmark
Complex multi-dimensional functions (Derivative valuation, Model Calibration PDE solving)	• Also sees much current interest ! Not clear if training is the bottleneck for most use cases (train once and done?)
Synthetic market data generation	• Useful research and risk testing tool ! Quality evaluation may be difficult ! Again, not clear training is bottleneck
Reinforcement learning for (hedging, trading, …)	• Under investigation

Training: Tell us what *You* think

- STAC Benchmarks are defined by financial firms to reflect their needs
- What training workloads give you the insight you need?
- **Join us!** ***www.STACresearch.com/ML***

Or… Come to our table to chat about ML training!

STAC-M3

(tick history / timeseries analysis)

STAC-M3 Overview

- Performance benchmarks for enterprise tick analytics
  - Language/DBMS neutral
  - Developed by banks and hedge funds
- Workload:
  - Synthetic data modeled on NYSE TAQ
  - Simulates concurrent access with varying number of users
  - Mix of I/O- and compute-intensive operations
- Many years of comparison points on diverse architectures

www.STACresearch.com/m3

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

Object Store

- Recent interest in SUTs involving object store
  - Get / Push
  - SUTs can leverage the meta data for searching
  - Impact of tiering
- Changes to data?
  - Many want to take advantage of compressibility within the object store
  - Is compressibility of STAC-M3 data realistic?
- Changes to queries?
  - Feedback from financial firms welcome
- Applies to all existing suites

STAC-M3 Suites

Suite	Dataset	Purpose	Impediments to caching/ pre-loading	Storage I/O	Network I/O*	Compute burden	Concurrent users
Baseline (Antuco)	Historical (~4TB)	Using a limited dataset size for convenience, simulate performance that would be obtained with a larger real-world dataset residing mostly on non-volatile media. Study a broad range of read and write operations.	Yes	Mostly high intensity reads	Neglible	Low to moderate	Varies
Small db in-memory (Shasta)	Historical (~4TB)	Study a broad range of operations for datasets that are relatively small in the real world. (While the dataset tested is the same size as in Antuco, there is no attempt to simulate the storage-access pattern of a larger dataset.)	No	Mostly high intensity reads	Neglible	Low to moderate	Varies
Scale (Kanaga)	Historical (theoretically unlimited TB)	Study a few operations on large datasets with large numbers of concurrent requests.	No	Mostly high intensity reads	Neglible	Low to moderate	Theoretically unlimited
Streaming (Jalua)	Streaming ingest & historical (~400GB)	Study ingest capacity, how long it takes ingested data to be available for querying, and query response times on both live and historical data.	tbd	Potentially high intensity writes, Low intensity random reads	Potentially high	Low to moderate	Theoretically unlimited

* Between SUT and test harness. Not necessarily within the SUT (e.g., a storage network)

STAC® SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 Data Generators

Share a data generator. Completely random from update to update

Suite	Dataset	Purpose
Baseline (Antuco)	Historical (~4TB)	Using a limited dataset size for convenience, simulate performance that would be obtained with a larger real-world dataset residing mostly on non-volatile media. Study a broad range of read and write operations.
Small db in-memory (Shasta)	Historical (~4TB)	Study a broad range of operations for datasets that are relatively small in the real world. (While the dataset tested is the same size as in Antuco, there is no attempt to simulate the storage-access pattern of a larger dataset.)
Scale (Kanaga)	Historical (theoretically unlimited TB)	Study a few operations on large datasets with large numbers of concurrent requests.
Streaming (Jalua)	Streaming ingest & historical (~400GB)	Study ingest capacity, how long it takes ingested data to be available for querying, and query response times on both live and historical data.

STAC® SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 Data Generators

Share a data generator. Completely random from update to update

Different data generator. Price changes less frequent and based on prev. price.

Suite	Dataset	Purpose
Baseline (Antuco)	Historical (~4TB)	Using a limited dataset size for convenience, simulate performance that would be obtained with a larger real-world dataset residing mostly on non-volatile media. Study a broad range of read and write operations.
Small db in-memory (Shasta)	Historical (~4TB)	Study a broad range of operations for datasets that are relatively small in the real world. (While the dataset tested is the same size as in Antuco, there is no attempt to simulate the storage-access pattern of a larger dataset.)
Scale (Kanaga)	Historical (theoretically unlimited TB)	Study a few operations on large datasets with large numbers of concurrent requests.
Streaming (Jalua)	Streaming ingest & historical (~400GB)	Study ingest capacity, how long it takes ingested data to be available for querying, and query response times on both live and historical data.

STAC® SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 Data Generators

Share a data generator. Completely random from update to update

Different data generator. Price changes less frequent and based on prev. price.

Suite	Dataset	Purpose
Baseline (Antuco)	Historical (~4TB)	Using a limited dataset size for convenience, simulate performance that would be obtained with a larger real-world dataset residing mostly on non-volatile media. Study a broad range of read and write operations.
Small db in-memory (Shasta)	Historical (~4TB)	Study a broad range of operations for datasets that are relatively small in the real world. (While the dataset tested is the same size as in Antuco, there is no attempt to simulate the storage-access pattern of a larger dataset.)
Scale (Kanaga)	Historical (theoretically unlimited TB)	Study a few operations on large datasets with large numbers of concurrent requests.
Streaming (Jalua)	Streaming ingest & historical (~400GB)	Study ingest capacity, how long it takes ingested data to be available for querying, and query response times on both live and historical data.

STAC®
SECURITIES TECHNOLOGY ANALYSIS CENTER

STAC-M3 with Object Store: Next Steps

- Modifying data generator for realistic compressibility
- Have a proof of concept with new data set for Working Group
- Working Group weigh in on the changes
- Modify suites as need be

www.STACresearch.com/m3

STAC SECURITIES TECHNOLOGY ANALYSIS CENTER