# Insights into Trading System Dynamics
## Deutsche Börse's T7®

May 2023

DEUTSCHE BÖRSE
GROUP

# Content

# 3
# Introduction

# T7® Technology Roadmap

Deutsche Börse is pursuing its technology roadmap to deliver innovative and superior trading technology.

**Recent developments**

- 21 February 2022: Partition 12 added for Eurex, Partition 60 added for Xetra
- 01 July 2022: Introduction of generic per port limit for CoLo 2.0 10 Gbit/s MD connections
- July 2022: Re-shuffling of products for Börse Frankfurt
- October/November 2022: Performance tuning of LF Gateways
- November 2022: Optional TLS encryption for ETI LF
- 21 November 2022: T7 Release 11.0
- March/April 2023: EOBI performance optimization on all partitions
- 06 March 2023: Introduction of combined PS Gateway/Matching Engine process for EEX
- 13 March 2023: Introduction of additional Eurex market data and order entry switches
- 17 April 2023: Latency optimization on Eurex futures partitions
- 08 May 2023: Mandatory TLS encryption of FIX LF

**Outlook**

- 22 May 2023: T7 Release 11.1
- 23 October 2023: Mandatory TLS encryption of ETI LF sessions
- 23 October 2023: HF Sessions only allowed from Equinix FR2 facility
- 20 November 2023: T7 Release 12.0
- 11 December 2023: Mandatory password encryption for HF sessions
- Q3/Q4 2023: Activation of PTRL for Eurex options.
- Q1 2024 – Consolidation of Matching Engine & EMDI process

# T7® Technology Roadmap

## Timeline of updates

**2023**
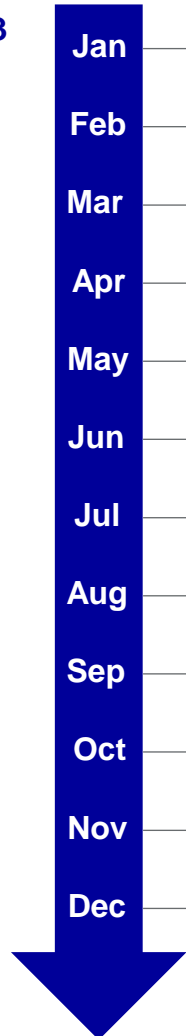
| | |
|---|---|
| **Jan** | |
| **Feb** | |
| **Mar** | EOBI performance improvement on all partitions<br>Introduction of consolidated PS gateway/Matching Engine for EEX |
| **Apr** | Additional Eurex CoLo 2.0 order entry and market data switches<br>Latency optimization on all Eurex futures partitions |
| **May** | Mandatory TLS encryption for FIX LF |
| **Jun** | T7® Release 11.1 |
| **Jul** | |
| **Aug** | |
| **Sep** | |
| **Oct** | Mandatory TLS encryption for ETI LF<br>HF sessions only allowed from Equinix FR2 Facility |
| **Nov** | Activation of PTRL for Eurex options<br>T7® Release 12.0 |
| **Dec** | Mandatory password encryption for HF sessions |

# Processed Transactions and Response Times

## T7 request – response round-trip times



T7 Request-Response Latency and Volume

- Deutsche Börse continuously invests in its trading system and is holding up transparency while providing a low latency trading venue.

- We have continued to add functionality while at the same time tuning our system further.

- Introduction of additional partitions for Eurex and Xetra increased system capacity (February 2022).

- After performance for HF sessions have been significantly improved with the combined PS Gateway/Matching Engine process, we also boosted performance for LF sessions in the 4th quarter of 2022.

- Pre-Trade-Risk-Limit (PTRL) functionality was introduced in T7 Release 11.0 for Xetra and enhanced for Eurex. This led to some performance impact and will be optimized in the next T7 Release 11.1.

- For Eurex options, the PTRL will be activated in Q3/Q4 2023.

# 7
## Recent Developments

# Introduction of TLS and Password Encryption

Deutsche Börse has implemented mandatory changes regarding data transmission via the T7 trading interfaces. These changes have been made to improve the security of information transmitted via communication networks to further reduce the risk of unauthorized interception, manipulation, or misuse.

**Overview:**

| Session Source Location | Session type | Encryption type | Mandatory change in Simulation | Mandatory change in Production |
|---|---|---|---|---|
| In Equinix FR2 co-lo facility | ETI (LF) | Payload or Password* | 24.11.2023 | 11.12.2023 |
| | ETI (HF) | Password | 24.11.2023 | 11.12.2023 |
| | ETI (HF) | Payload | Not supported | Not supported |
| Remote | ETI (LF) | Payload | 04.08.2023 | 23.10.2023 |
| | ETI (LF) | Password | Not applicable | Not applicable |
| | ETI (HF) | Not applicable | Not applicable | 23.10.2023** |

\* LF sessions connecting from Equinix FR2 co-lo facility must use either payload or password description.
\*\* The use of HF sessions in the production environment outside of the Equinix FR2 co-location facility will no longer be available.

Detailed information can be found in Eurex circular 44/2023 and Xetra circular 28/2023.

# EOBI DSCP Field and Discard IP Range
## Speculative Triggering

**What is speculative triggering?**

Initiation of a sending a request by a participant at a time when not all required information are available.

There are certain triggers in a market data packet that reveal information:

1. Ethernet preamble of market data (earliest time = 0 ns)
2. Destination MAC (ethernet header) = identification of the multicast stream
3. IP total length: indicates the type of message contained, e.g. ExecutionSummary or single order add
4. Payload (up to 100 ns later): Quantity and Price.

**Why is it used?**

Speculative triggering is incentivized by the deterministic technical network architecture of the T7 trading system. The 10 Gbit/s access layer switches operate in a 'cut-through' mode and the first bytes of an ethernet frame reserve priority on the competitive network path for the uplink to the next switch. This incentivizes latency sensitive trading participants to send technical transactions purely to reserve switch priority creating a high load on the T7 system. To avoid unnecessary technical transactions on the T7 a technical solution has been implemented.

# EOBI DSCP Field and Discard IP Range
## Confining the effects of speculative triggering

### Daily Discard IP usage during main trading hours



— % Eurex Discard    — % Xetra Discard
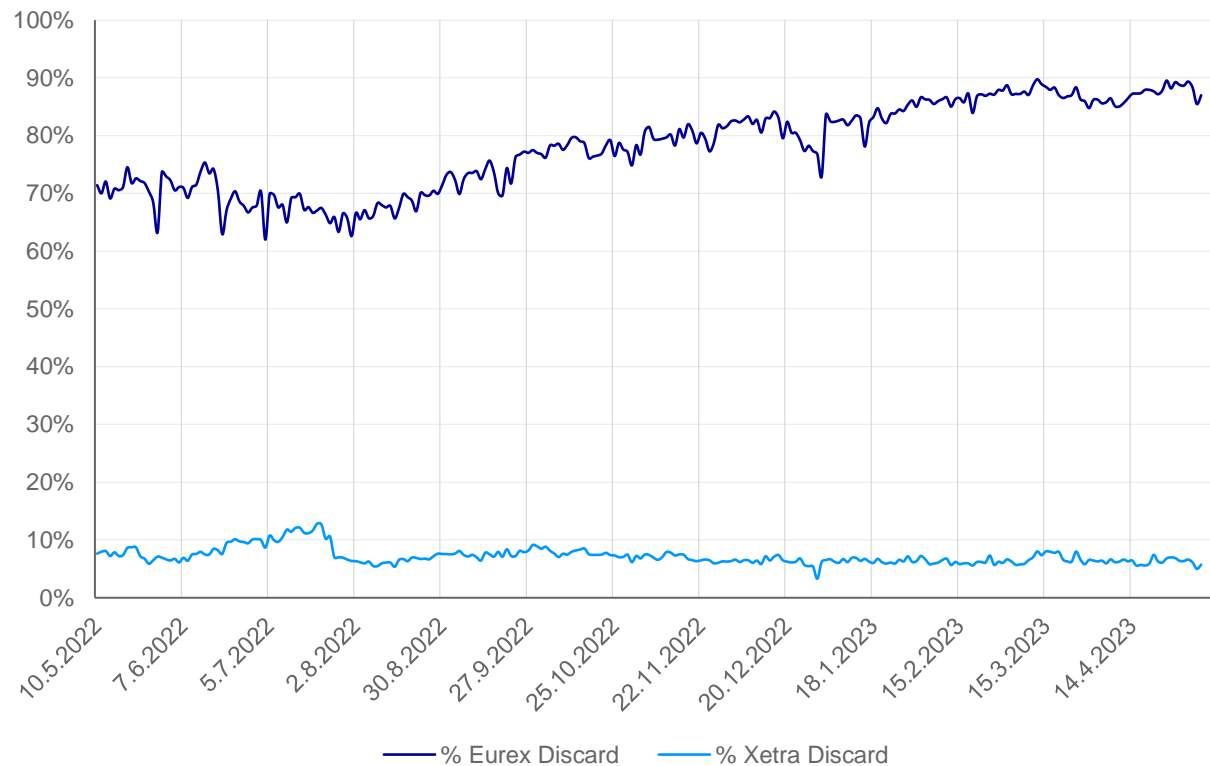
In July 2020, Eurex introduced a Discard IP address range 172.16.0.0/16 on the 10 Gbit/s order entry networks. Xetra followed in May 2021. Trading participants may send falsely 'speculative' triggered packets to the discard IP range, instead of sending it to the exchange. These packets will be discarded at the access layer switch port and no other participant is influenced. Packets sent to the discard IP address are not considered to be orders and are not forwarded to the exchange.

To enable market participants to effectively use this discard IP address, the DSCP field of the IPv4 headers in EOBI market data packets is used. Four different bits indicate the most common 'interesting' market situations.

The amount of packets reaching the Matching Engines decreased significantly after introduction of this measure. Since then, the amount of Discard IP packets have constantly increased.

The graph on the left shows the percentage of Discard IP packets compared to total number of packets reaching the access layer switches between 9:00 and 18:00 CET per day.

# EOBI DSCP Field and Discard IP Range

## Confining the effects of speculative triggering (continued)

**A technical solution to the speculative triggering problem:**

- Mark potential triggers early in the IP header of market data packet with the help of DSCP (Differentiated Services Code Point) flags.

- Offer a non-competitive Discard IP destination address to enable packets to be discarded right on the Access Layer switch.

DSCP flags indicate Execution summaries and/or widening or narrowing of the bid/ask spread from orders (not quotes).

Examining these flags allows participants to still change the destination IP address of an in-flight outgoing message to the Discard IP address for non interesting packets. These packets will be sent on the exchange network but will not reach the trading system as they will be rejected on the access layer switch port facing the participant sending the transaction.
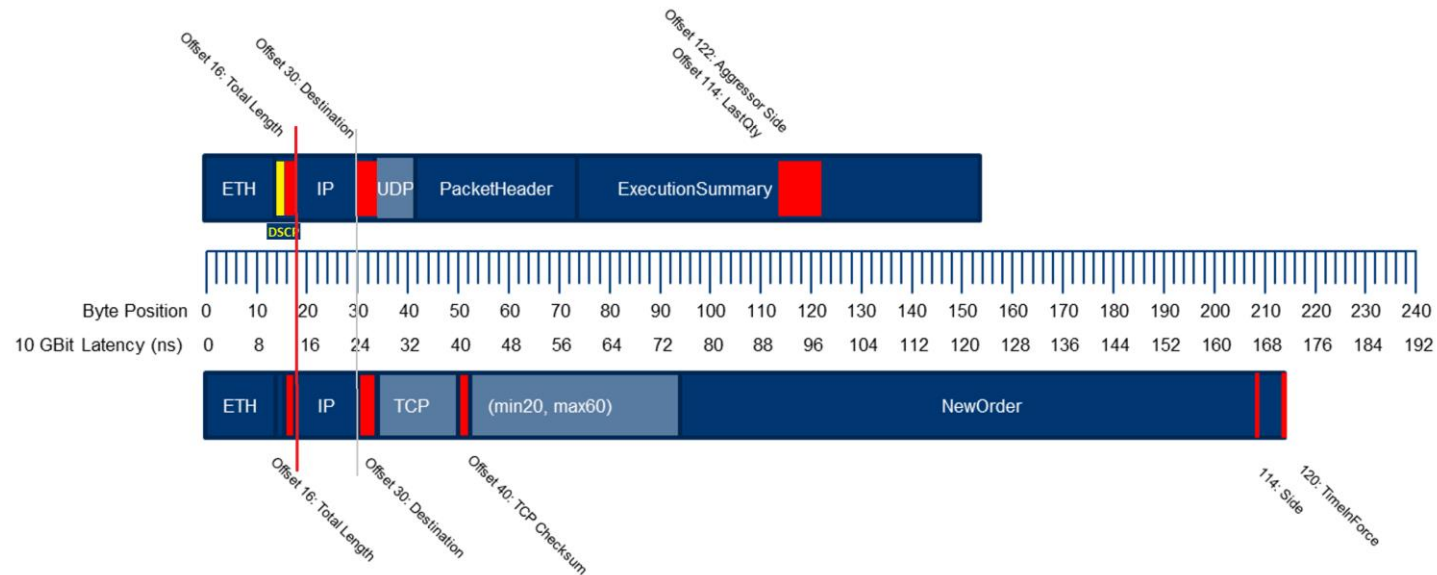


The left figure shows at which byte position different information is available in the message.

A market data packet is received at time 0.

The fastest participants may react and send a response as early as the first bit of market data has been received, dynamically reading the market data packet while already streaming out the response..

The response packet may be modified in-flight, after reading e.g. the DSCP flags, total length etc.

Even with a reaction time of 0 the outgoing destination IP address can still be modified after evaluating the DSCP flags of the incoming market data packet.

# Participant Reaction Time

## Insights from the High Precision Timestamp file

We define reaction time as the latency between a 'trigger' market data packet and a request that leads to an execution.

We use measurement point t_9d for the market data packet and t_3a for the request. We measure 2722 ns as the minimum latency between t_9d and t_3a determined by the T7 infrastructure.*

The below charts show the distribution of the reaction times for all Eurex products (blue) and FESX to FDAX (grey) from 13 March 2023 (right) compared to the last figures from 8 June 2021 (left).

We observe a high level of competition (there are around 10 trading participants with reaction times < 2770 ns for most active products). The fastest participants have moved closer to each other.
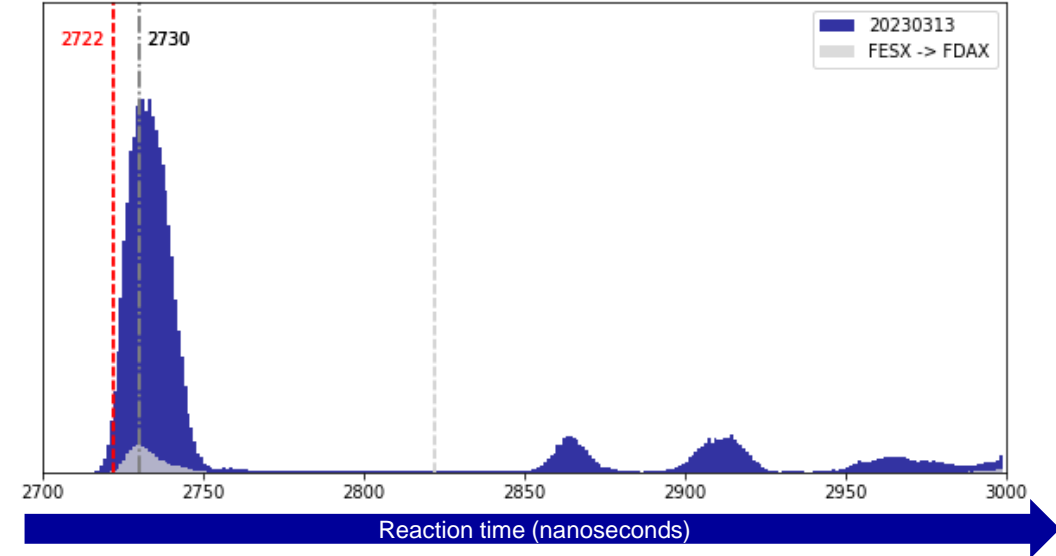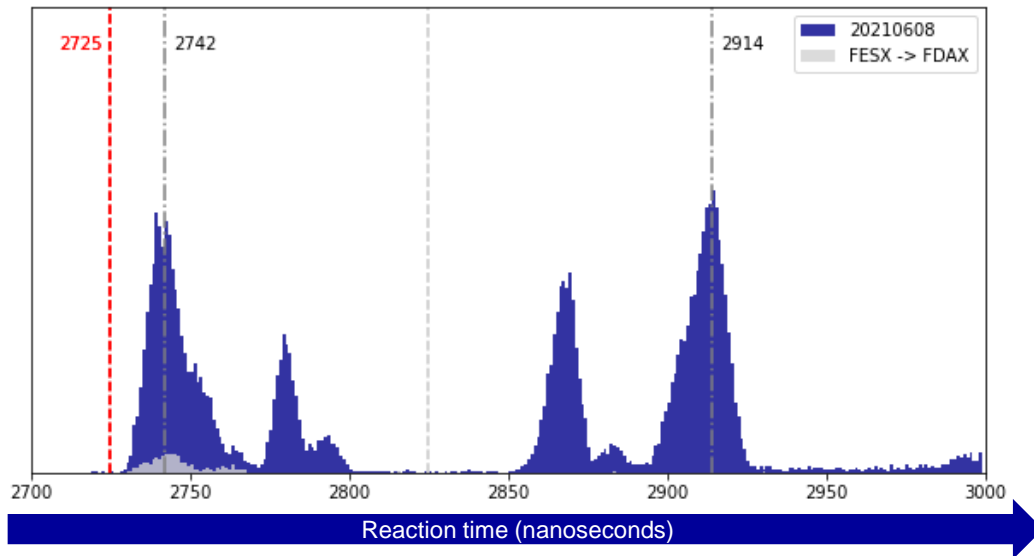


*Minimum latency reduced due to optimization of CoLo 2.0 connections

# 13
**Latency Analysis**

# T7® Topology
## At a glance

The below shows the Eurex B side as a schematic example of the topology of the T7 system.

Note that Xetra has only two access layer switches per side.
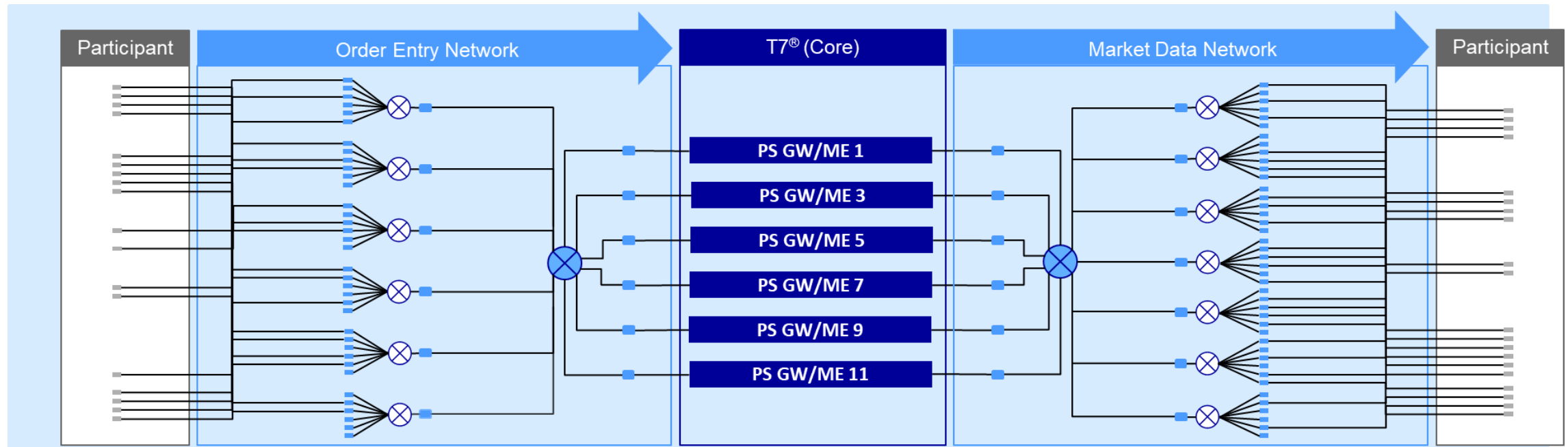


Legend:
- Cisco Nexus 3548-X Distribution Layer Switch
- Cisco Nexus 3548-X Access Layer Switch
- Optical TAP (capture and time measurement)

# T7® Latency
## Composition

The T7 trading system provides utmost transparency about its latency characteristics. Most of the timestamps taken at the various measurement points within T7 Core are included in each ETI response and EOBI market data. With the high precision timestamp file we also make three network timestamps available for each EOBI market data packet (t_9d) and it's triggering transaction (t_3a, t_3d).

The latency circle shows the median latencies for the request-response/EOBI market data path for Eurex futures sent via high frequency sessions measured on 26 April 2023.
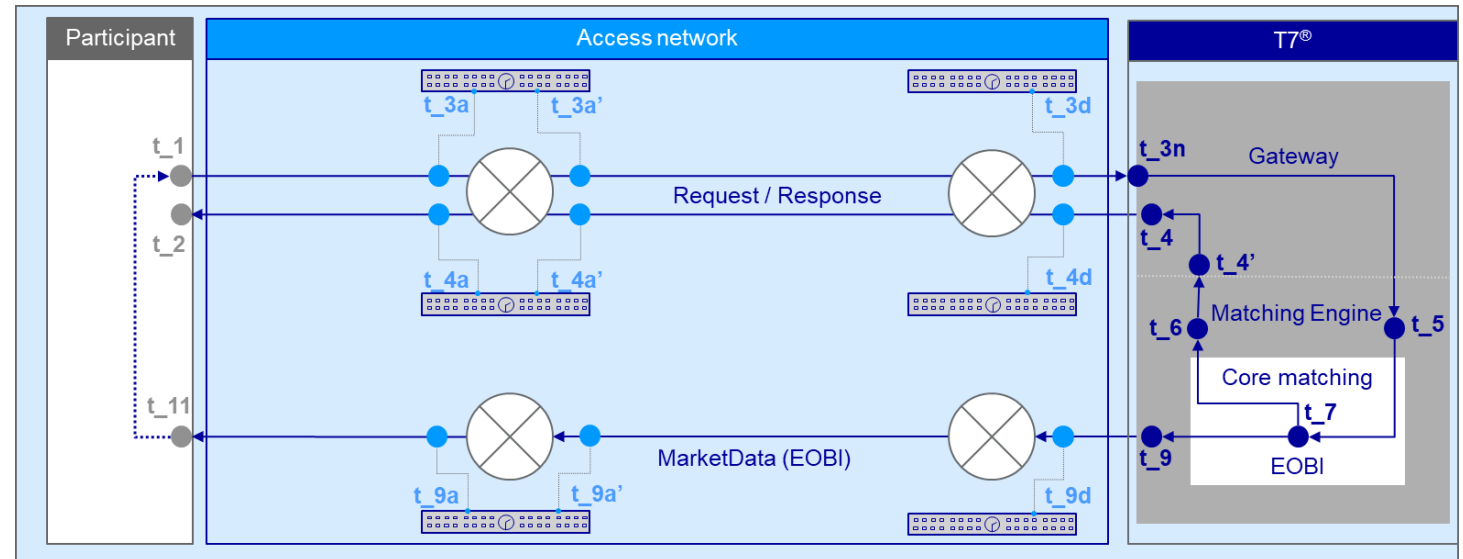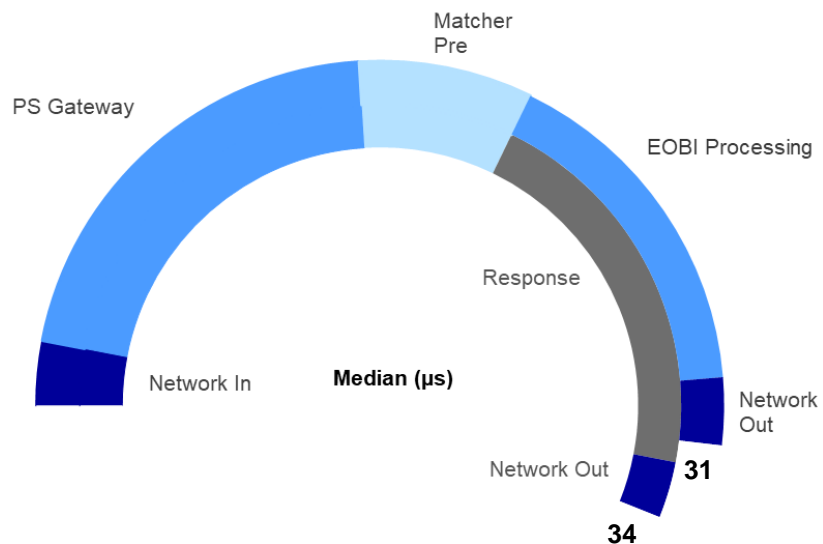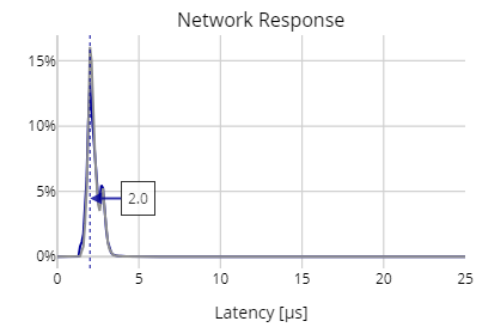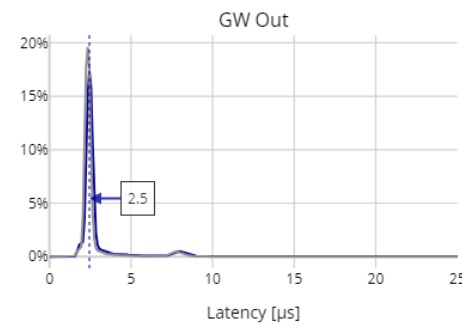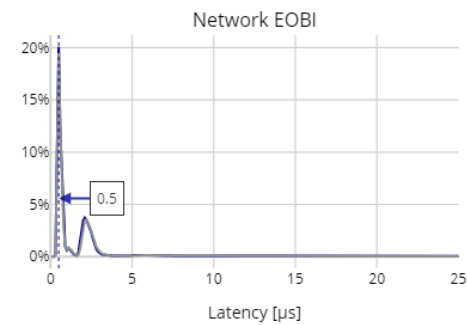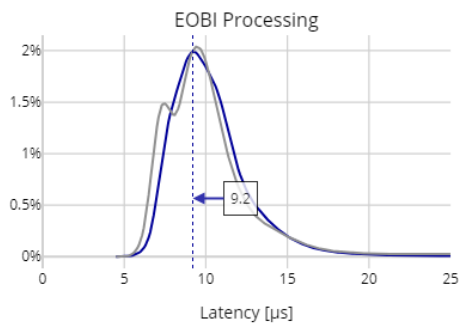
# T7® Latency
## Composition (continued)

The below charts show a comparison of current latencies (26 April 2023) with the spectrum from 7 September 2022. The data is for Eurex futures sent from HF sessions. 'Network response/EOBI' include the TCP/UDP stack on the respective server.
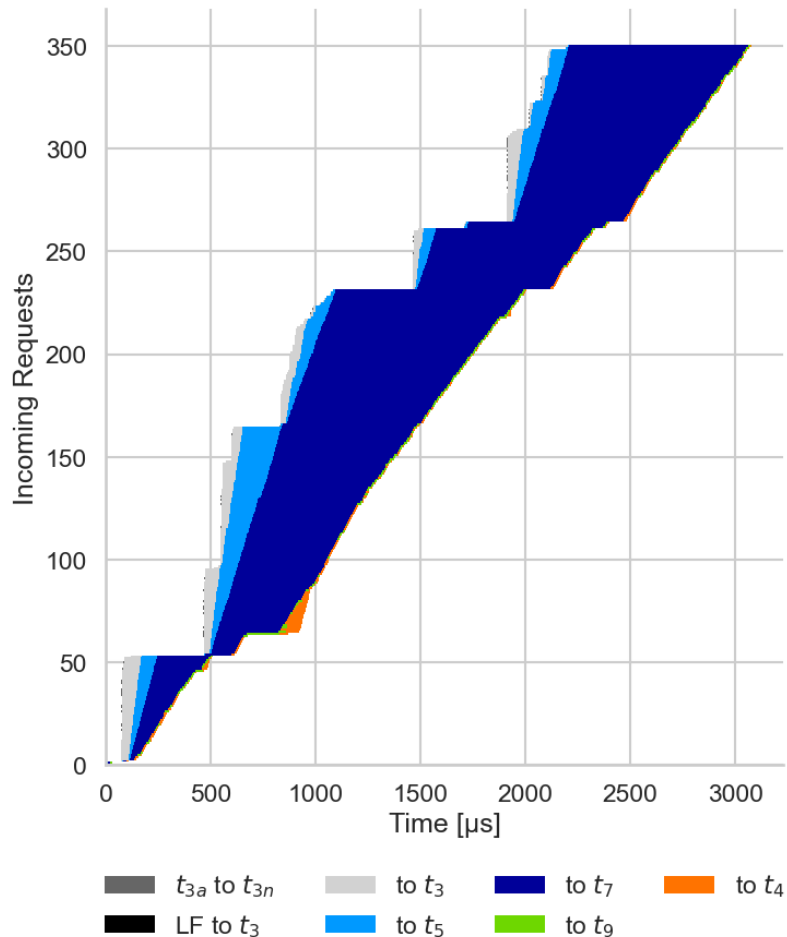


Futures latency compared with 2022-09-07
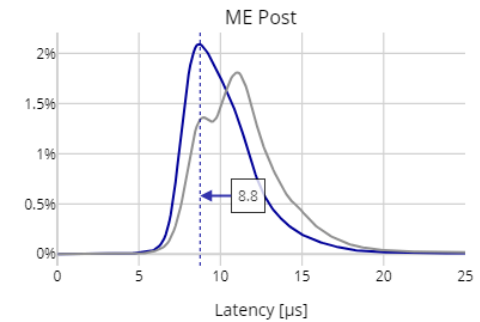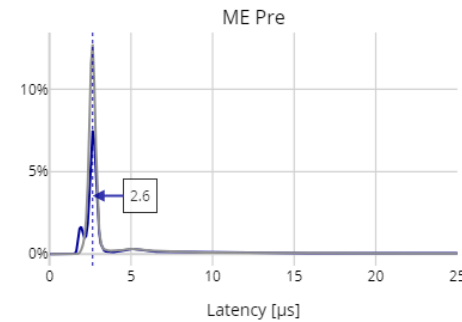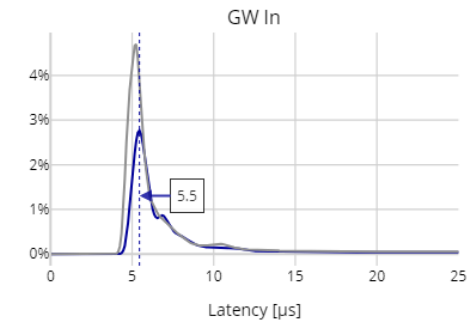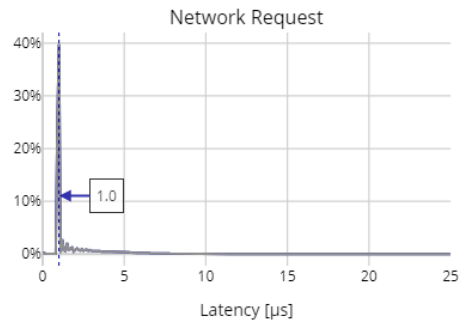
# T7® Latency

## Eurex Micro-burst dynamics



During micro-bursts, the input into the trading system may be greater than the throughput capabilities. This in turn causes queuing which results in higher latencies.

Higher latencies cause risk (i.e. it takes longer to place/pull an order).

T7 provides real-time performance insights by providing up to seven timestamps with each response and key timestamps with every market data update.

The chart on the left shows the paths:

- Network access layer ($t_{3a}$) or LF gateway In ($t_3$) to PS gateway In ($t_{3n}$) to

- Matching Engine in ($t_5$) to

- Start matching ($t_7$) to

- EOBI SendingTime ($t_9$) [where available] to
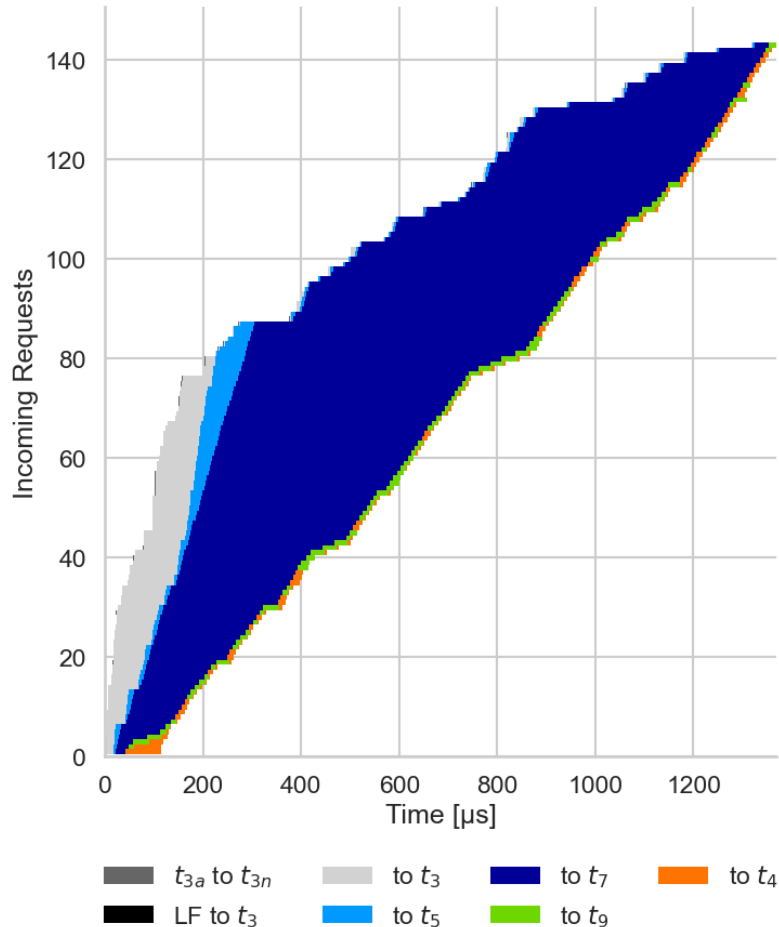
- ETI SendingTime ($t_4$).

Typical throughput rates in kHz (1/ms) are 8000 at $t_{3n}$, ~400 at $t_5$, ~200 at $t_7$.

EOBI send times are usually before the gateway send time of responses.

Note that base latency for requests entered via LF gateways is ~ 32 µs higher. As all requests are routed via PS gateways no overtaking between LF gateway and PS gateway requests is observed.

# T7® Latency
## Xetra Micro-burst dynamics



During micro-bursts, the input into the trading system may be greater than the throughput capabilities. This in turn causes queuing which results in higher latencies.

Higher latencies cause risk (i.e. it takes longer to place/pull an order).

T7 provides real-time performance insights by providing up to seven timestamps with each response and key timestamps with every market data update.

The chart on the left shows the paths:

- Network access layer (t_3a) or LF gateway In (t_3) to PS gateway In (t_3n) to

- Matching Engine in (t_5) to

- Start matching (t_7) to

- EOBI SendingTime (t_9) [where available] to

- ETI SendingTime (t_4).

Typical throughput rates in kHz (1/ms) are 8000 at t_3n, ~400 at t_5, ~200 at t_7.

EOBI send times are usually before the gateway send time of responses.

Note that base latency for requests entered via LF gateways is ~ 32 µs higher. As all requests are routed via PS gateways no overtaking between LF gateway and PS gateway requests is observed.

# T7® Latency

## Order entry network

The order entry network in Colocation 2.0 uses a two staged hierarchical funnel in approach.

All cables are normalized to guarantee a maximum deviation of +/- 0.5 m (+/- 2.5 ns) between any two cross connects to the exchange. Deutsche Börse is in the process of reducing this deviation to +/- 35 cm (+/-1.75 ns) – to be finished in the course of 2023.
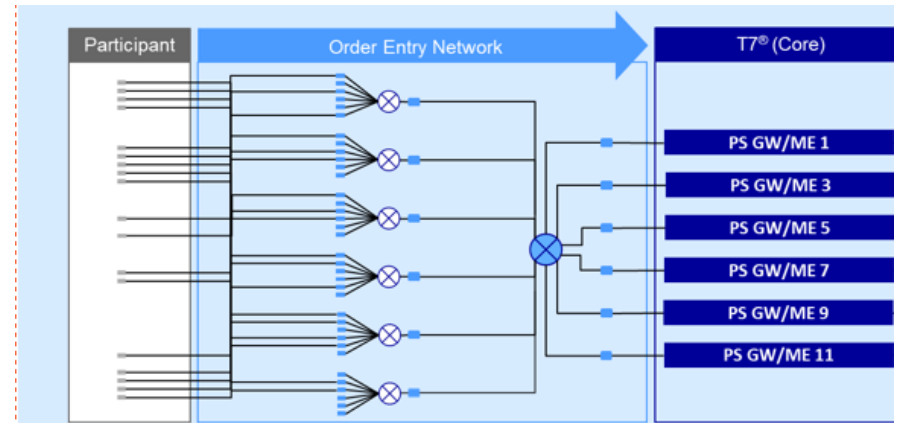


Every inbound and outbound packet on this path is captured via passive network TAPs at 3 different locations.

The packets are timestamped with nanosecond resolution and accuracy.

This capture infrastructure allows early detection and analysis of many kinds of technical network problems and an in-depth latency analysis on network level, like overtaking probabilities between measurement points.

The high precision timestamp file service enables participants access to timestamps t_3a, t_3d and t_9d for each request leading to a an EOBI market data update.

# Order Entry Network
## Latency aspects
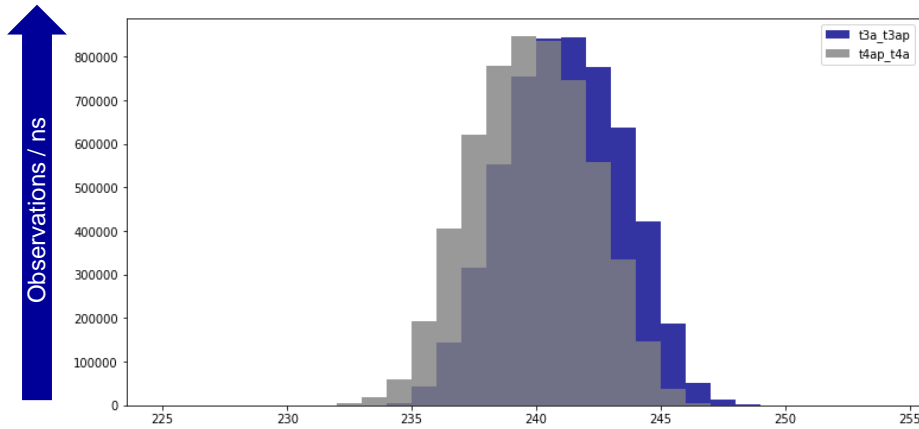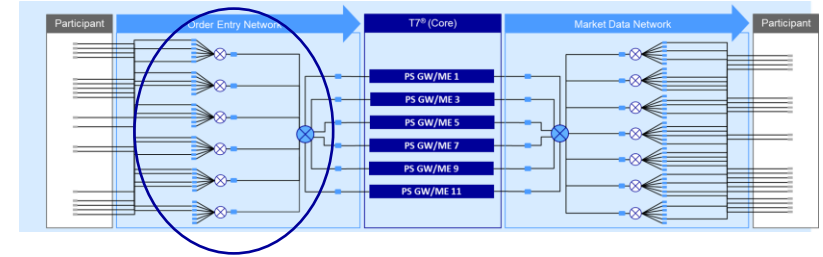




The base latency and latency jitter is identical for all access layer switches within the measurement precision. The spectrum is tight with a standard deviation of under 3 ns (top left and inset).

Note that bursts of traffic may lead to queues in access and distribution layer switches because multiple lines funnel into a single uplink (bottom left, note the logarithmic scale).
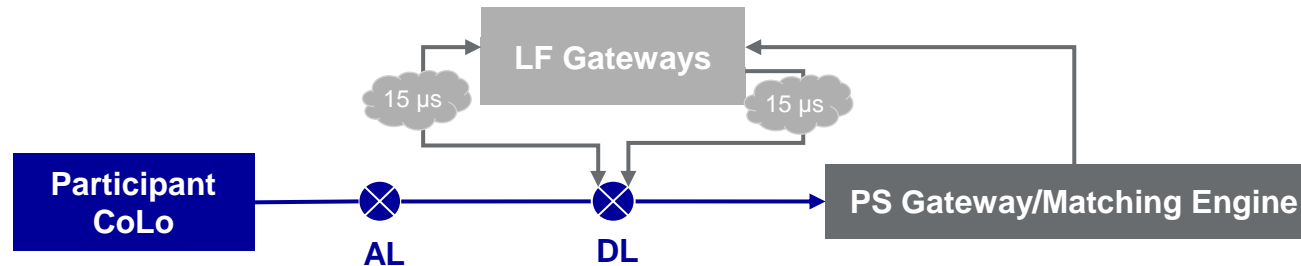
The chart on the bottom right shows the overtaking probability on a switch between a first message in a burst and consecutive messages. The overtaking drops sharply and there is no observed overtaking beyond 4 nanoseconds delta between messages.

Data is taken from the Eurex B side.
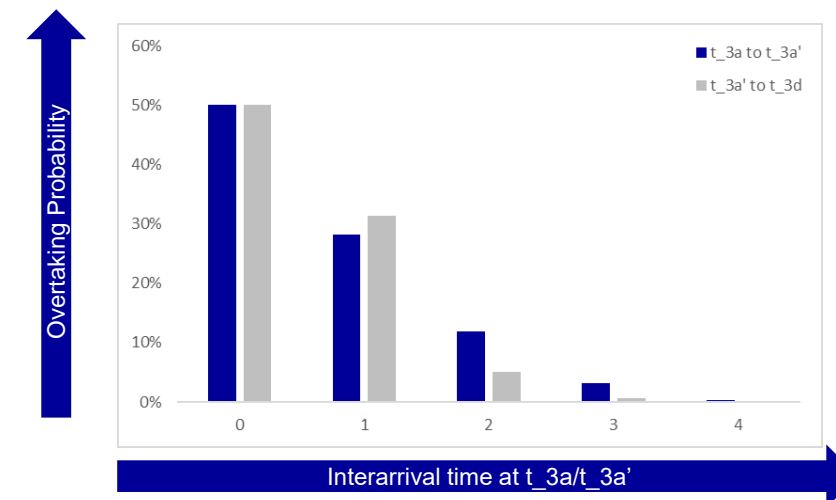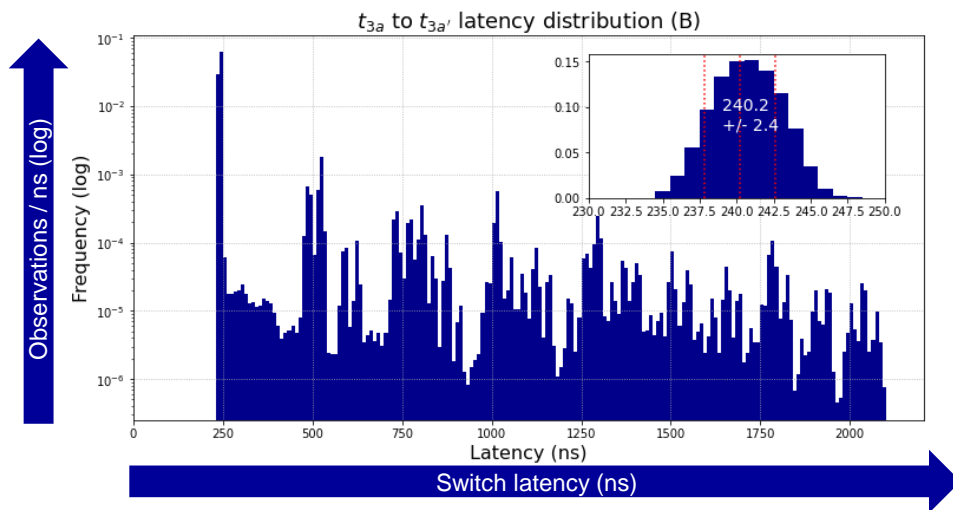
# PS Gateway versus LF Gateways

## Latency comparison



- Using 10 Gbit/s cross connects and access via PS gateways provides the fastest way for order and quote management in T7.

- LF gateways on the other hand allow access to all partitions of a market via a single session.

- Some markets (e.g. XMAL) are using LF gateways only.

- Most markets have combined PS/gateway Matching Engines in place for which all Matching Engine bound requests sent to LF gateways are routed via PS gateways.

- The base latency of the path to the PS gateway is around 32 µs higher for LF gateways when compared to directly accessing the PS gateways. Note also that requests that have to cross sides between LF and PS gateways take another ~45 µs longer to reach the Matching Engine.
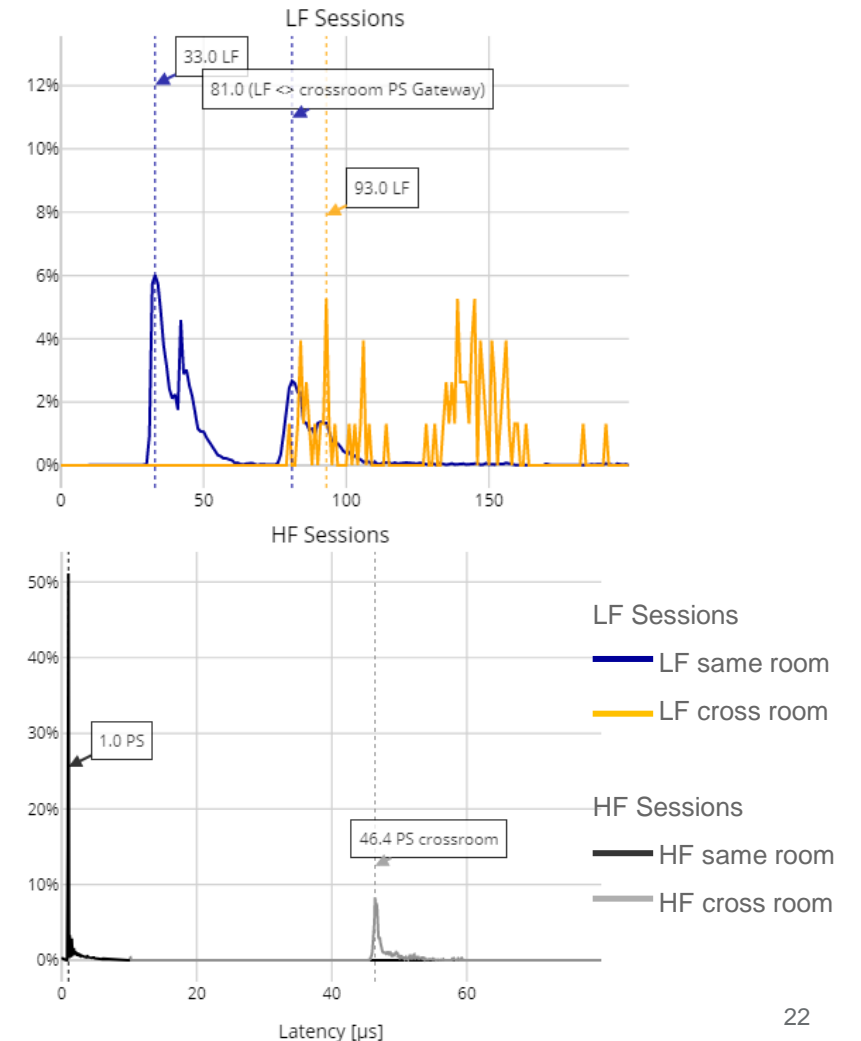
# Order Entry Latency
## Comparison of access types

The table below shows a comparison of different access options to the T7 system. All times given are in microseconds.

Network timestamps (t_3a) are synchronized using white rabbit. The time synch quality between these timestamps is thus ~1 ns. Other T7 timestamps are subject to jitter of up to +/- 50 ns.

| Gateway type | Same side line | Same side partition | t_3a to t_3n uncongested latency |
|---|---|---|---|
| **PS** | ☑ | ☑ | 1 |
|  | ☐ | ☑ | 46 |
| **LF** | ☑ | ☑ | 33 |
|  | ☑ | ☐ | 81 |
|  | ☐ | ☑ | 93 |
|  | ☐ | ☐ | 137 |



LF Sessions

33.0 LF
81.0 (LF <> crossroom PS Gateway)
93.0 LF



HF Sessions

1.0 PS
46.4 PS crossroom

LF Sessions
— LF same room
— LF cross room

HF Sessions
— HF same room
— HF cross room
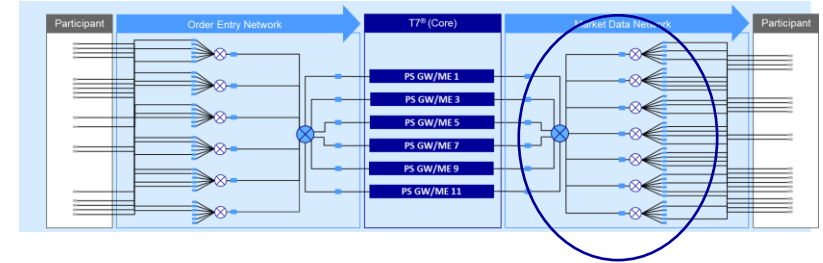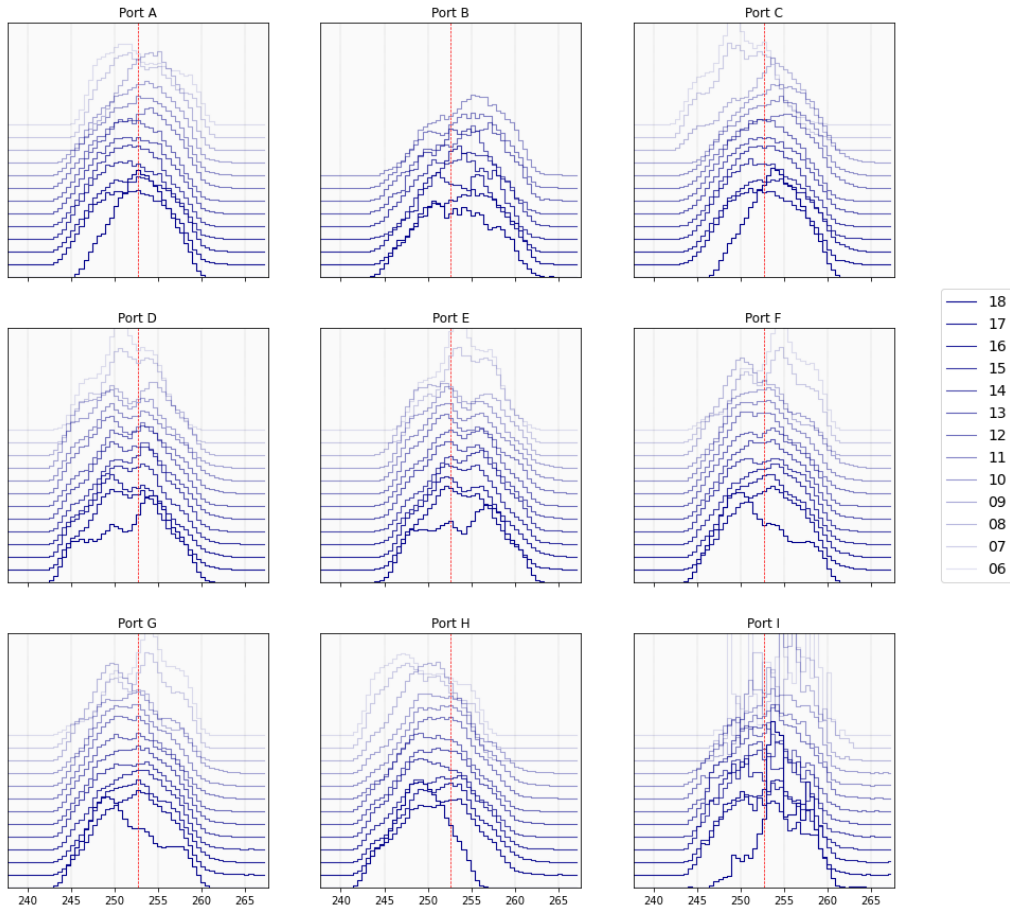
Latency [μs]

# 23
## Market Data

# Market Data Network
## Latency aspects



t_9a' to t_9a latency (ns)



The market data network has a funnel in – fan out topology. It funnels in data from different market data disseminators (on the distribution layer switch) and fans it out via multiple access layer switches.
We took extra care to establish a balanced and deterministic network. Additionally, static forwarding is configured to ensure equal multicast load towards access layer switches.

We observe a semi static latency difference of up to 10 ns (comparing t_9a of two different ports on the same switch) due to internal multicast processing of the switch.

The Cisco Nexus 3548-X switch divides the on-board Ethernet-Ports into three buffer areas. Serial replication for multicast takes place within these areas. In terms of latency this means any port within this area could be faster than other ports, which is an artifact of variable internal pointers to packet queues. The starting port for replication of a packet in a buffer area is identified by an internal pointer.
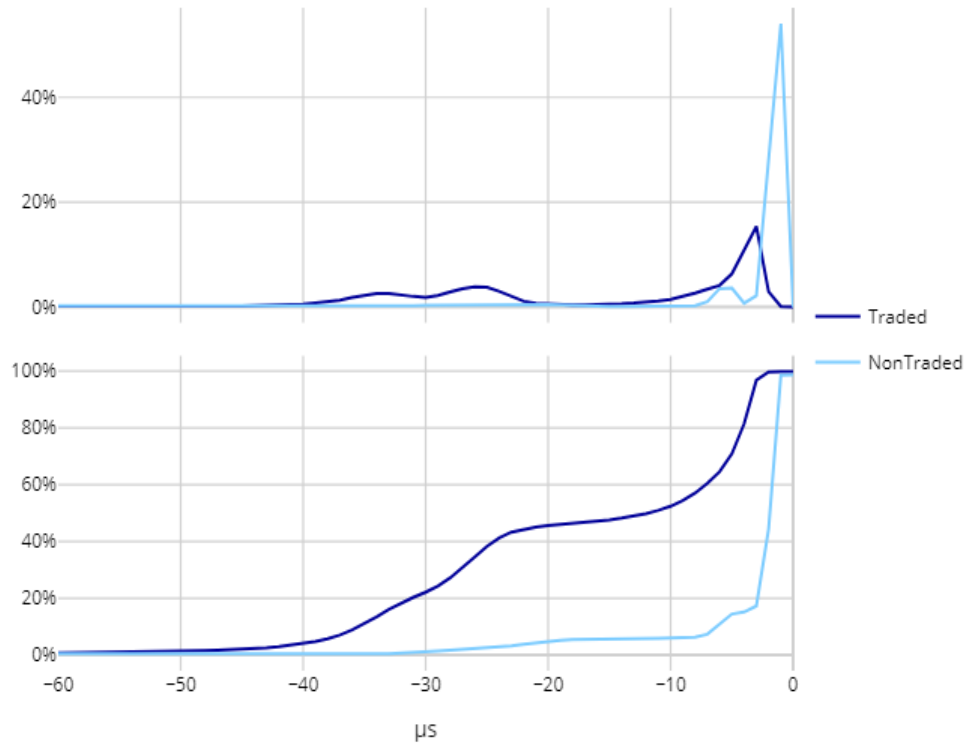
The figure on the left shows the latency distributions (t_9a' to t_9a) for nine ports belonging to one buffer area of one access layer switch.

Each distribution represents one minute of the specified hour of the day.

# Trading System Dynamics

## Latency characteristics of EOBI versus ETI for Futures

ETI vs EOBI for FESX 2023-04-26

T7 is designed to publish order book updates first on its public data feed.

The diagram shows the time difference distribution between public and private data in microseconds (EOBI first datagram vs ETI responses, t_9 - t_4).

The data is a production sample from 26 April 2023.

EOBI market data is in median 1.8 μs faster than the ETI response for order book updates and 12 μs faster for executions.

The first EOBI datagram is faster in more than 99.96 percent of the cases compared to the ETI response and also the first passive ETI book order notification for simple transactions.

Public data has become even faster with the latest EOBI optimization measures.

# Trading System Dynamics
## Latency characteristics of EOBI versus ETI for Options

ETI vs EOBI for OESX 2023-04-26
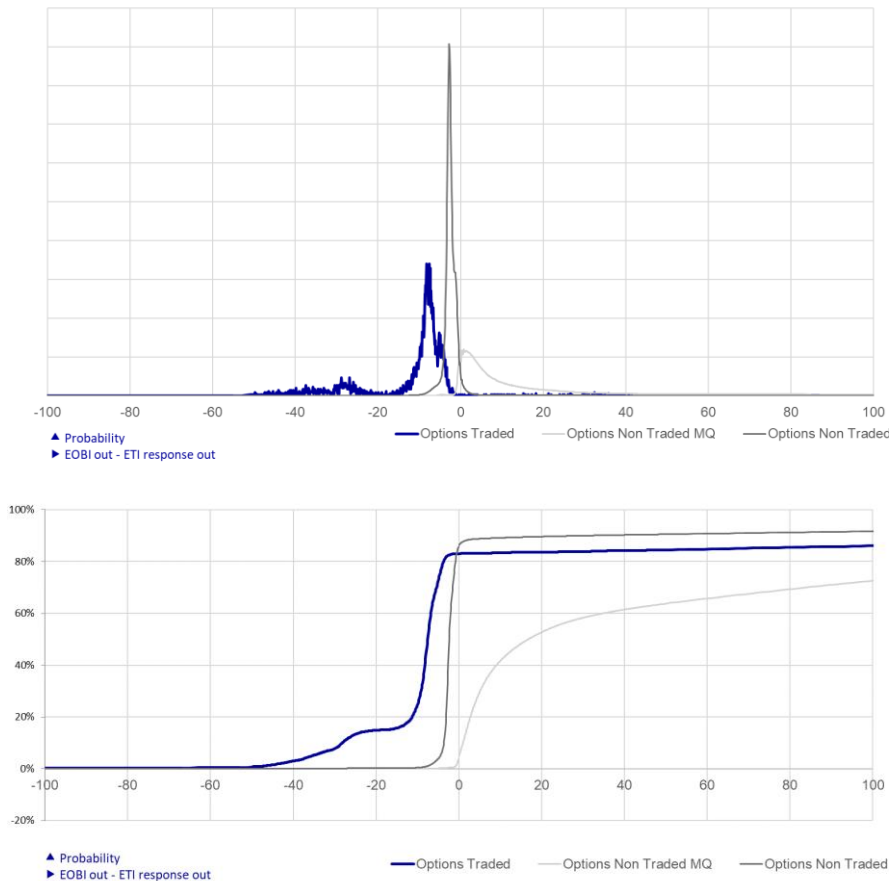


The data is a production sample from 26 April 2023 for OESX Options.

We distinguish between orders leading to a trade, mass quotes with more than one quote pair and orders and single quote updates.

The difference between public and private data has slightly been narrowed with the introduction of the consolidated PS Gateway/Matcher process.

Trades are received first on EOBI in around 82% of the cases with a median latency advantage of 7.5 µs, single orders in around 86% of the cases and a median latency advantage of 2.3 µs.

There are two main reason for EOBI delays:
The transaction is delayed by preceding messages (queues)
The transaction causes a market maker protection with many quote deletions

The latency profile for mass quotes is dominated by larger mass quotes, where the EOBI publisher has to broadcast each quote, leading to longer delays and queues in the EOBI path, while the ETI path only deals with a simple mass quote ack.

# Trading System Dynamics

## Latency characteristics of EOBI versus ETI for Xetra
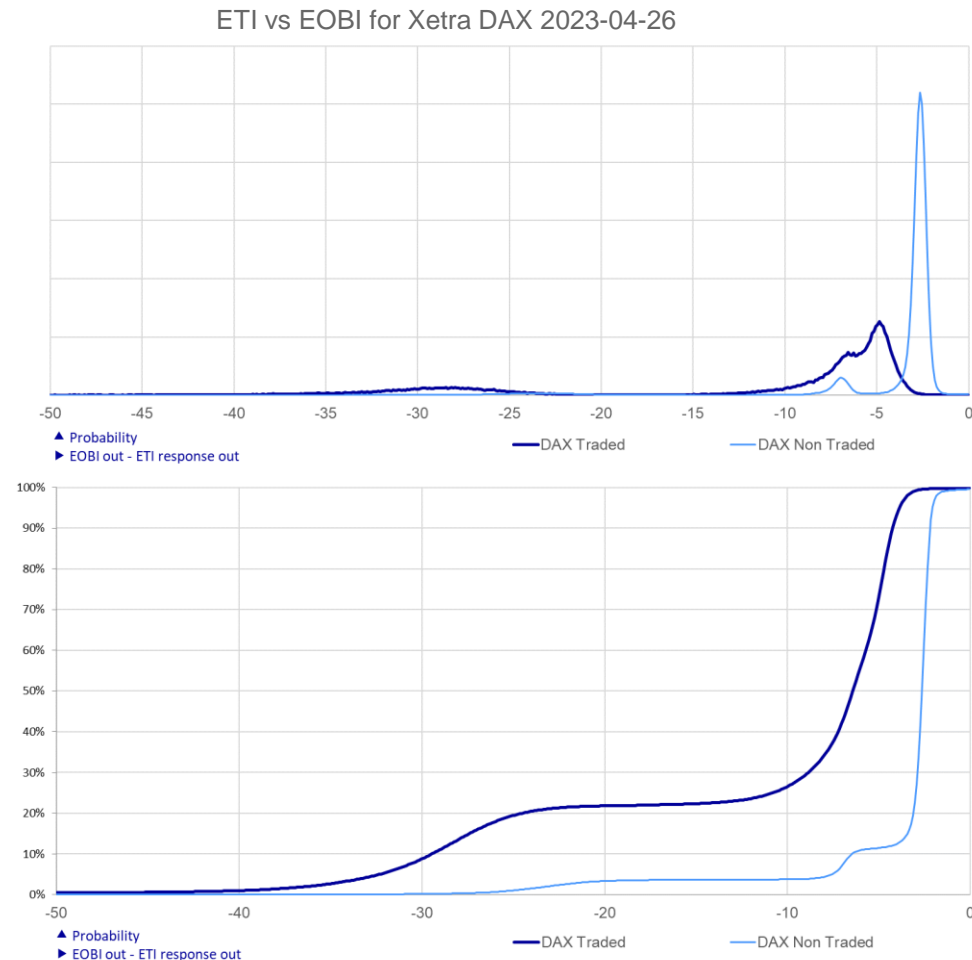
ETI vs EOBI for Xetra DAX 2023-04-26

The diagram shows the time difference distribution between public and private data in microseconds for XETRA DAX products (EOBI first datagram vs ETI responses, t_9 - t_4).

The data is a production sample from 26 April 2023.

The latency distribution is similar to the Eurex futures, trades are received on EOBI in median 2.5 µs faster, whereas single orderbook updates are usually 6.5 µs faster on EOBI.

# Trading System Dynamics
## Latency characteristics of EOBI versus EMDI

Market data updates provided via EOBI are almost always faster than EMDI.

The left graph shows the distribution of t_9 minus t_8, i.e. EOBI first datagram versus EMDI sending time for 26 April 2023 (negative values: EOBI is faster). It shows that EMDI is significantly slower especially for mass quotes.

The right table shows the difference of EOBI versus EMDI at various percentiles in µs.

The difference will be smaller (i.e. EMDI will become faster) after the planned consolidation of the EMDI process into the Matching Engine. However, the consolidation guarantees that EOBI will always be sent first and thus faster than EMDI.
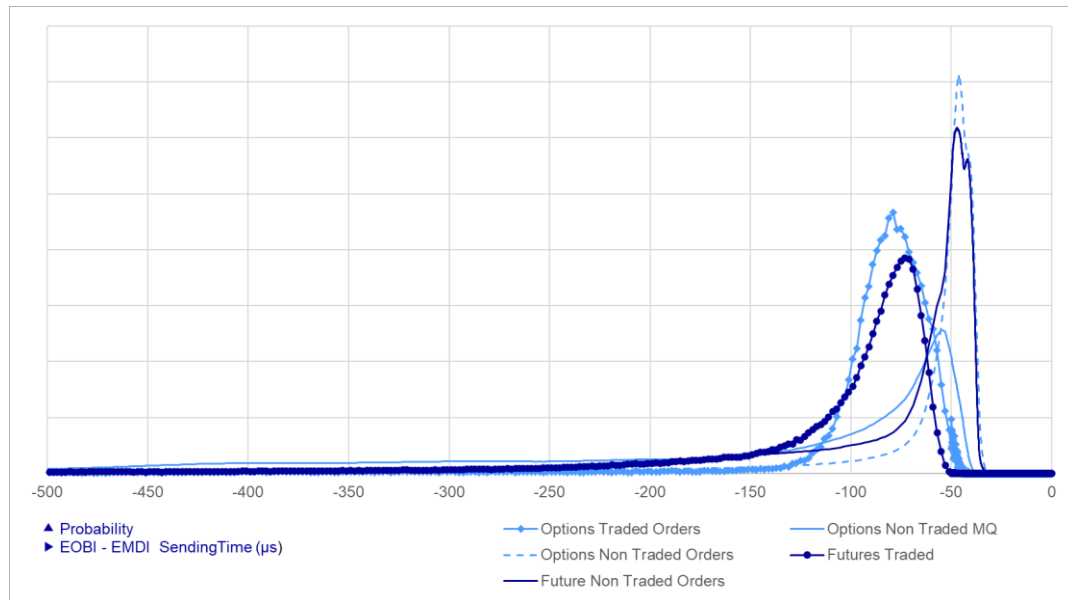


| Percentile | Options Traded Orders | Options Non Traded MQ | Options Non Traded Orders | Futures Traded | Future Non Traded Orders |
|---|---|---|---|---|---|
| 0.1 | -68125 | -68125 | -68125 | -61375 | -4125 |
| 1 | -2875 | -14125 | -37625 | -8125 | -243 |
| 10 | -117 | -499 | -8375 | -329 | -79 |
| 25 | -93 | -329 | -443 | -135 | -67 |
| 50 | -79 | -127 | -55 | -89 | -59 |
| 75 | -67 | -65 | -46 | -73 | -53 |
| 90 | -59 | -51 | -41 | -65 | -50 |
| 99 | -50 | -44 | -36 | -55 | -46 |
| 99.9 | -46 | -40 | -34 | -51 | -42 |
| EOBI faster | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

# Trading System Dynamics
## Latency characteristics of ETI versus EOBI versus EMDI

This diagram displays the dependency of the median latency on the complexity of a trade for ETI
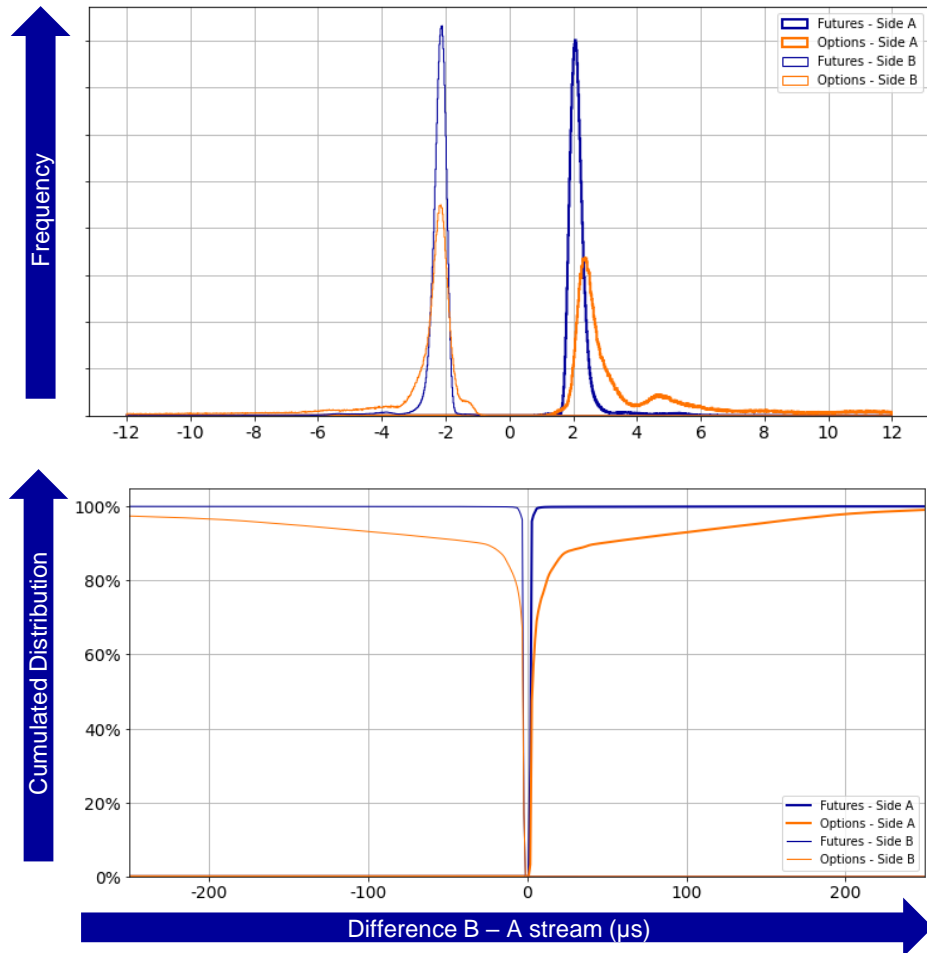
($t\_4 - t\_7$), EMDI ($t\_8 - t\_7$) and EOBI ($t\_9 - t\_7$). Note that for ETI we display the gateway sending time of the first passive notification and for EOBI the sending time of the UDP datagram containing the Execution Summary message.

The difference between public and private data has slightly been narrowed with the introduction of the consolidated PS Gateway/Matcher process. However, the 'public data first' principle is still being ensured. In over 99% of all trades, we disseminate order book data on EOBI first (also for larger trades).
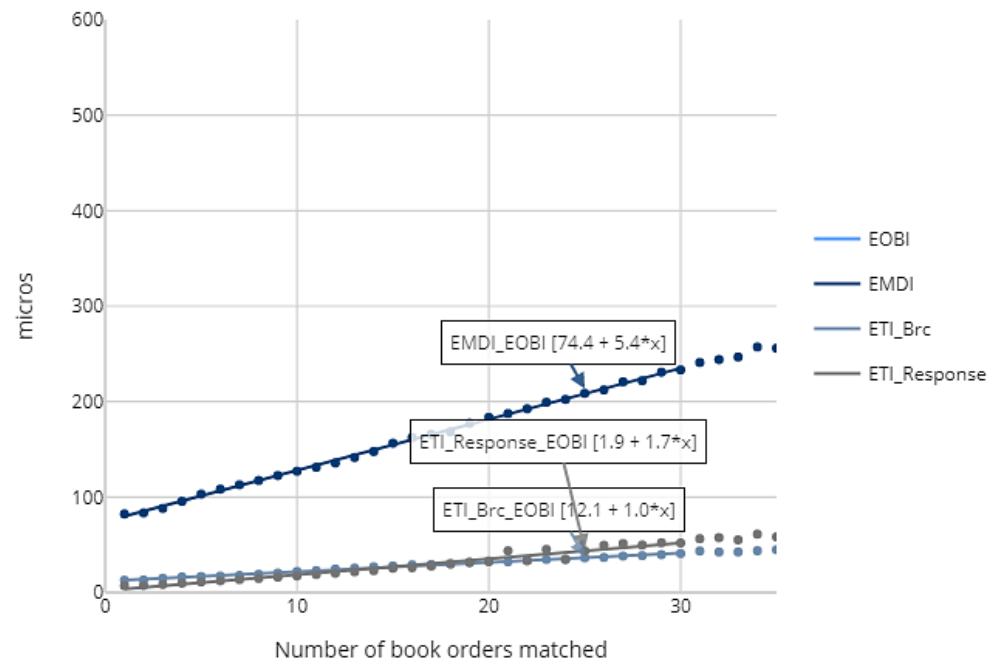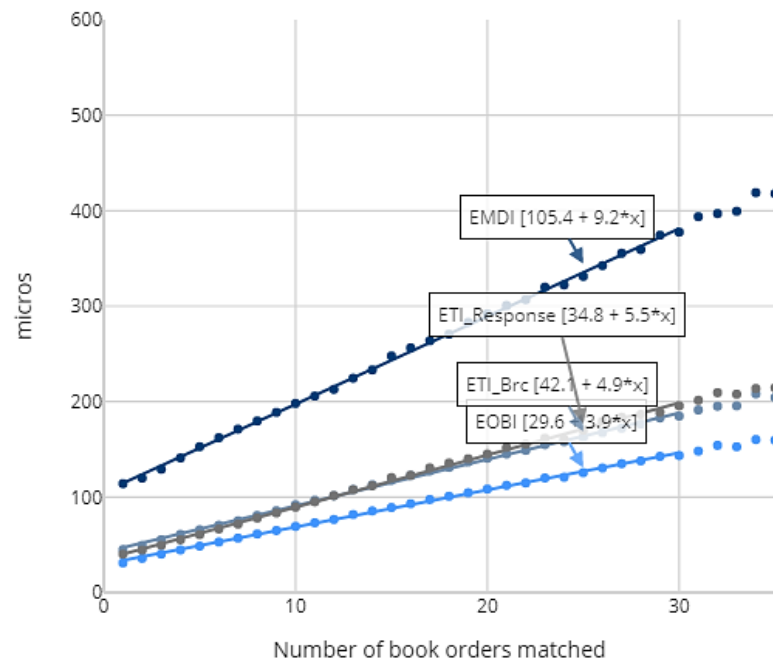
# Trading System Dynamics

## EOBI latency difference of primary and secondary feed



For products assigned to even partitions, market data is published first on the A and then on the B stream. For products assigned to odd partitions market data is published first on the B and then on the A stream.

The partition ID / product ID is contained in the UDP datagram header of the order book incremental messages and can be used for filtering on UDP datagram level for EMDI / EOBI.
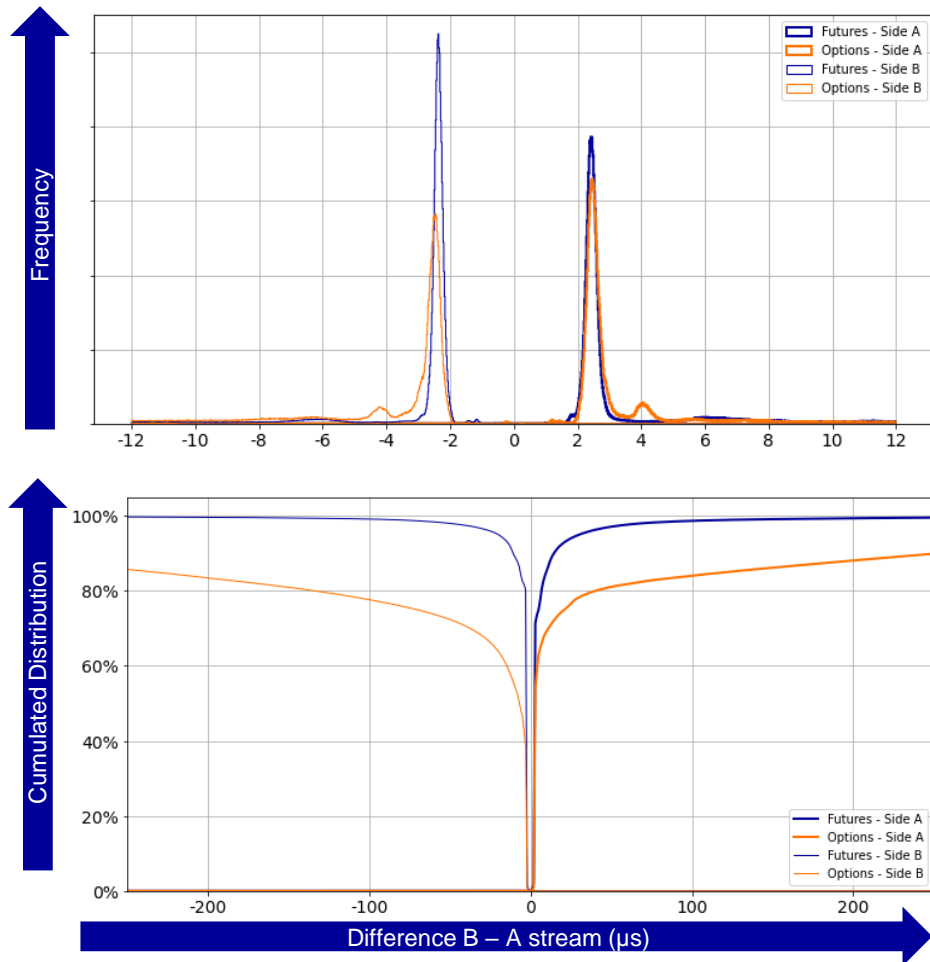
Furthermore, a UDP datagram on the T7 EMDI / EOBI order book delta or snapshot channel contains data of exactly one product (e.g. EURO STOXX 50® Index Futures).

The data for both primary and secondary streams is disseminated by the same server using two ports connected to the two sides of the network. The process sends the data first on the primary interface. After all datagrams of a transaction are sent it starts sending on the secondary interface.

The median latency difference between the A and the B EOBI incremental feed is about 2 µs for futures and Cash products. For options the median is slightly higher and there are far more outliers (i.e. much slower secondary feed). The reason is that since data is published on the secondary feed only after all datagrams of a transaction are sent on the primary feed, the latency difference depends on the complexity of the transaction, i.e. a mass quote with 200 quote updates will lead to a higher delay than a single order entry.

# Trading System Dynamics
## EMDI latency difference of primary and secondary feed



For products assigned to even partitions, market data is published first on the A and then on the B stream. For products assigned to odd partitions market data is published first on the B and then on the A stream.

The partition ID / product ID is contained in the UDP datagram header of the order book incremental messages and can be used for filtering on UDP datagram level for EMDI / EOBI.

Furthermore, a UDP datagram on the T7 EMDI / EOBI order book delta or snapshot channel contains data of exactly one product (e.g. EURO STOXX 50® Index Futures).

The data for both primary and secondary streams is disseminated by the same server using two ports connected to the two sides of the network. The process sends the data first on the primary interface. After all datagrams of a transaction are sent it starts sending on the secondary interface.

The median latency difference between the A and the B EMDI incremental feed is about 2 to 3 µs for futures and Cash products. For options the median is slightly higher and there are far more outliers (i.e. much slower secondary feed). The reason is that since data is published on the secondary feed only after all datagrams of a transaction are sent on the primary feed, the latency difference depends on the complexity of the transaction, i.e. a mass quote with 200 quote updates will lead to a higher delay than a single order entry.

# Eurex: Market data volume
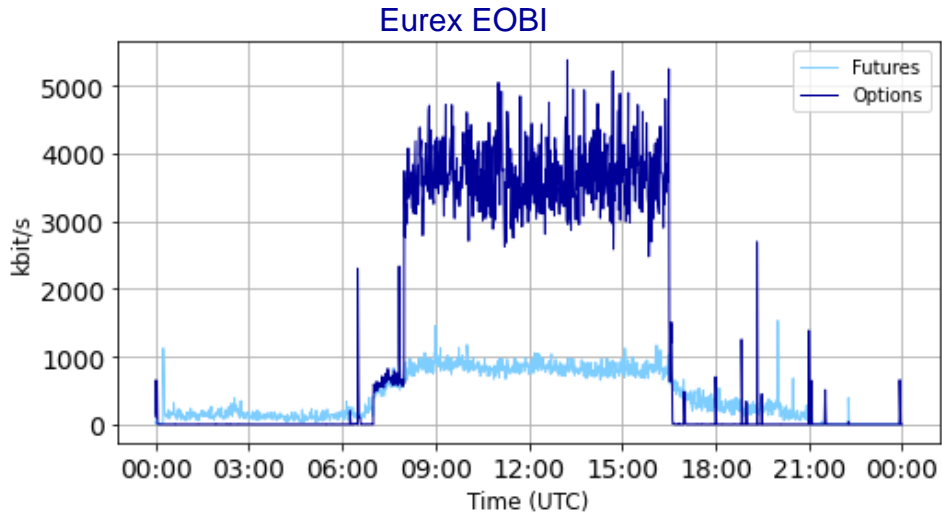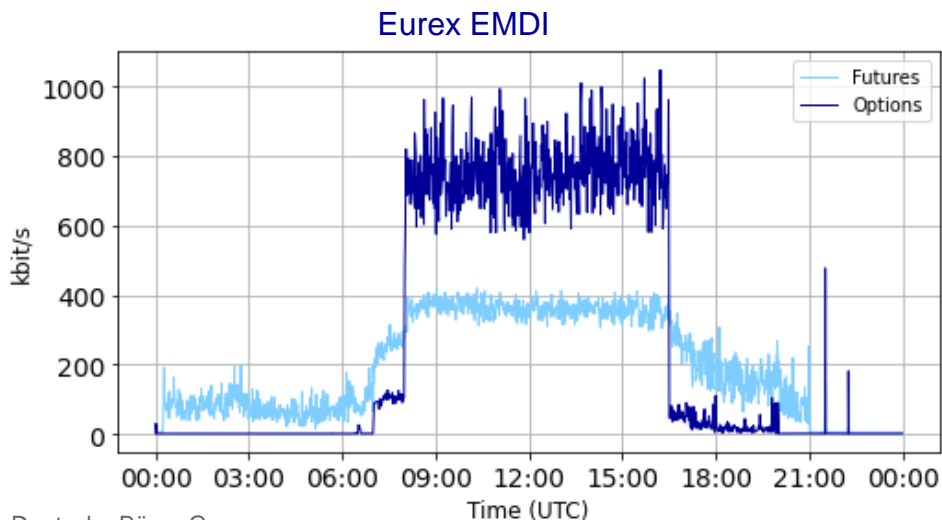

Eurex EOBI


Eurex EMDI

Each data point equals the maximum bandwidth produced on a 1 millisecond scale by the incremental B stream in Mbit/ms.

The provided data shows one data point per minute for 13 March 2023 – a very busy trading day.
Enhanced Order Book Interface (EOBI) peak volume is significantly higher than price level aggregated data volume EMDI. EOBI market data is therefore currently only available to trading participants using 10 Gbit/s connections.

The EOBI for options incremental data stream peaks around 5.5 Gbit/s on millisecond level, while the futures stream peaks at 1.5 Gbit/s.

Participants that want to receive data for Eurex Exchange's products on EMDI with less than 1 ms queuing delays need to use a connection with a bandwidth of more than 1 Gbit/s (options) or 400 Mbit/s (futures) respectively. Trading participants are advised to use two 10 Gbit/s connections (one for each market data stream) in Co-Location to receive market data.

# Xetra: Market data volume

Each data point equals the maximum bandwidth produced on a 1 millisecond scale by the incremental B stream in Mbit per ms.

The provided data shows one data point per minute for 13 March 2023 (a busy trading day).

Enhanced Order Book Interface market data is currently only available to trading participants using 10 Gbit/s connections.
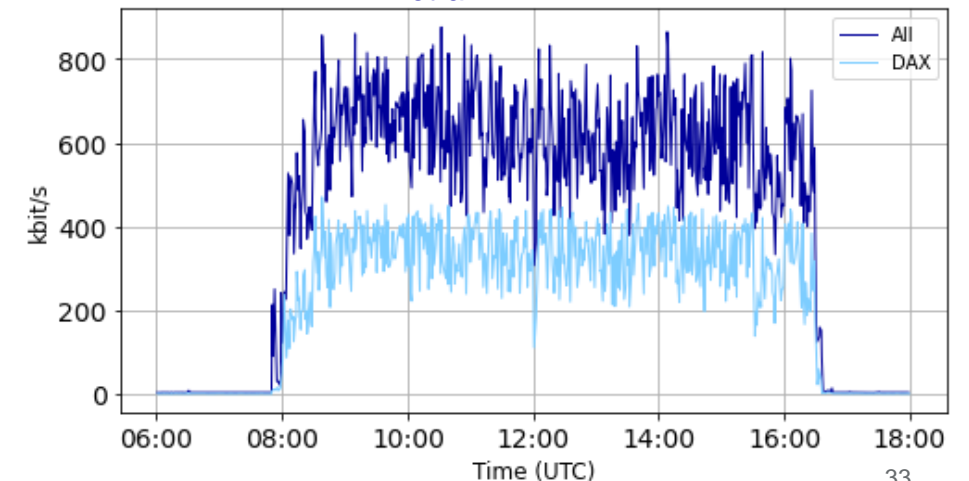Trading participant are advised to take two cross connects (one for each market data stream) in Co-Location to receive market data.

Participants that want to receive EMDI data with less than 1 ms queuing delays need to use a connection with a bandwidth of at least 850 Mbit/s (All products) or 450 Mbit/s (DAX® equities only).



Xetra EOBI
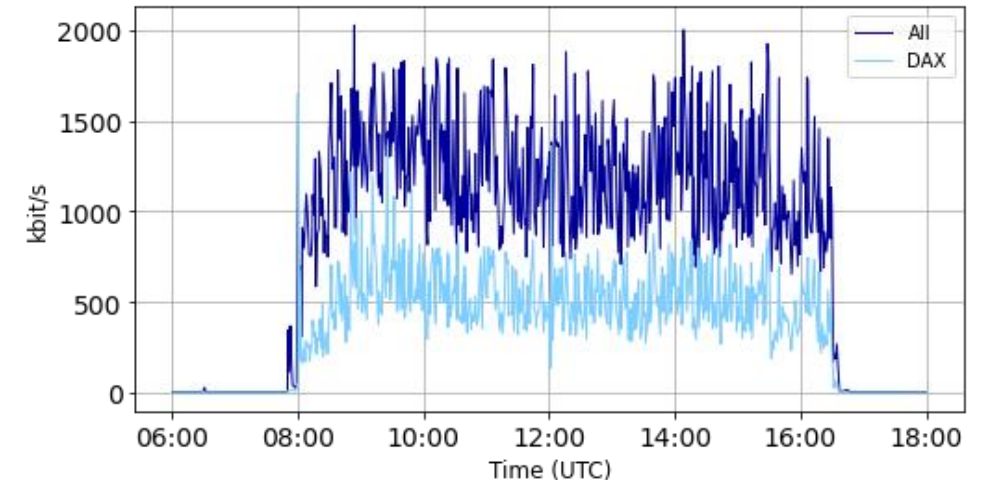


Xetra EMDI

**34**
**What you need to be fast**

# What you need to be fast…

## A few recommendations to achieve the low latency

Use the Equinix Co-Location facility to be close to Deutsche Börse T7.

Use state-of-the-art switches (if any) and only have at most one hop between the exchange network and your server. Alternatively, use hardware solutions to connect directly without hops (e.g. FPGA).

Use good network interface cards and TCP/IP acceleration, e.g. a kernel-by-pass library.

Use at least two 10 Gbit/s (cross-) connections in Co-Location for EOBI market data and two 10 Gbit/s connections for T7 ETI.

Use HF sessions to connect to PS gateways and make sure you use the cross connect on the same side as the gateway you are connecting to (compare time-to-live values in the IP header in the responses from both sides).

Measure and analyze your own timestamps (e.g. the reaction time as recommended on the next slide).

Use state of the art time synchronization, i.e. GPS clocks and a high-quality time distribution. The PTP signal you can get from us has a quality of +/- 50 ns. For our network timestamps we use the White Rabbit protocol and PPS breakouts. You can connect to our white rabbit time service providing you a time synchronization quality of 1-2 ns max, see https://www.deutsche-boerse.com/dbg-en/products-services/ps-technology/ps-connectivity-services/ps-connectivity-services-time-services

We provide highly accurate network timestamps of all orders leading to a market data update via the high precision timestamp file service, see https://www.mds.deutsche-boerse.com/mds-en/analytics/high-precision-timestamps.

Use the EOBI Execution Summary for fast trading decisions and position keeping (passive executions). For a consistent order book, all incremental updates following the Execution Summary should always be processed. For fastest decisions evaluate the market data classification based on the DSCP flags in the IPv4 header of EOBI market data packets.

Trade notifications need to be processed to create safety. We recommend to use either a low-frequency ETI session or a FIX trade capture drop copy to confirm the fast execution information provided by the execution reports via high-frequency sessions

# What you need to be fast…

## Participant reaction time measurement
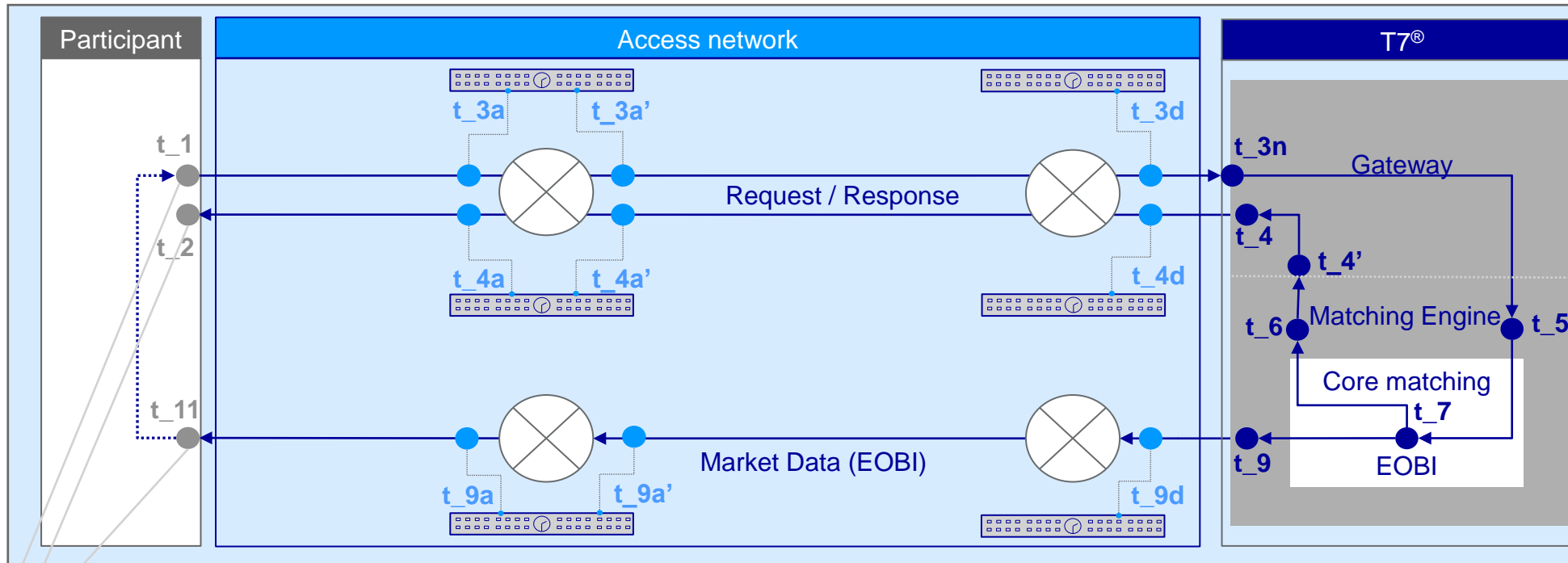


Measure the time between market data reception (t_10/t_11) and your reaction (t_1) and align with timestamps from the high precision timestamp file

t_3a, t_3d and t_9d are available via the high precision timestamp file service, see https://www.mds.deutsche-boerse.com/mds-en/analytics/high-precision-timestamps
Take our white rabbit signal to compare your timestamps with ours with ns accuracy
https://www.deutsche-boerse.com/dbg-en/products-services/ps-technology/ps-connectivity-services/ps-connectivity-services-time-services#tab-1556860-1556864

**37**
**T7® Overview**

# T7® Architecture
## Overview

T7® architecture developed by Deutsche Börse:

- o Uses state-of-the-art infrastructure and hardware for high performance
- o Offers reliable connectivity and enriched trading functionality
- o Is multi-market capable, operates Derivatives (Eurex, EEX) and Cash markets (Xetra, Börse Frankfurt, Vienna, etc.)

- T7® consists of partitions. A partition is a failure domain in charge of matching, persisting and producing market data for a subset of products. Each T7 partition is distributed over two rooms in the Equinix data centre.

- There are 12 Eurex T7 and 11 Xetra T7 partitions.

- Separate partitions are used for markets of other exchanges hosted on T7 (e.g. Vienna (XVIE), EEX (XEEE), Bulgaria (XBUL), …).

- The reference data contains the mapping of products to partition IDs.

- 4 LF gateways and one FIX LF gateway allow access to all Eurex partitions and the separate EEX partition.

- 4 LF gateways and one FIX LF gateway allow access to all Xetra partitions.

- 2 LF gateways and one FIX LF gateway are shared between Vienna and their partner exchanges.

- 2 LF gateways and one FIX LF gateway are shared between XBUL and XMAL

- Note that the active half of a partition is either on side A (for even partitions) or on side B (for odd partitions).

- Only in case of the failure of a PS Gateway/Matching Engine or a market data publisher, the active half of the service will shift to the other room.

- With consolidated PS Gateway/Matcher processes (since Q2 2021) the active PS gateway and active Matching Engine act as a single failure domain within each partition, i.e. they will always fail as a single logical group.
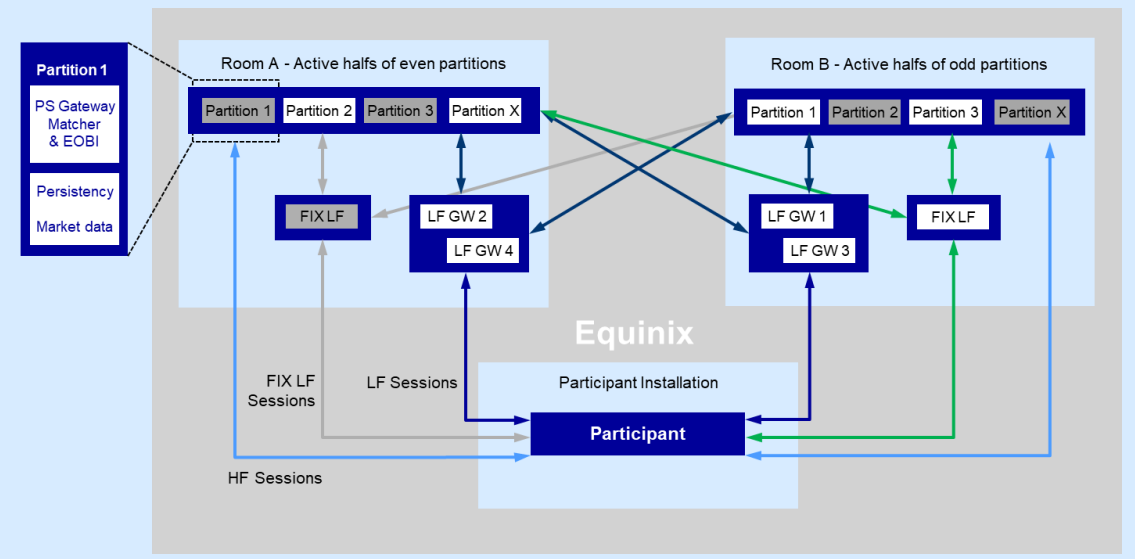
# T7® Topology
## Overview
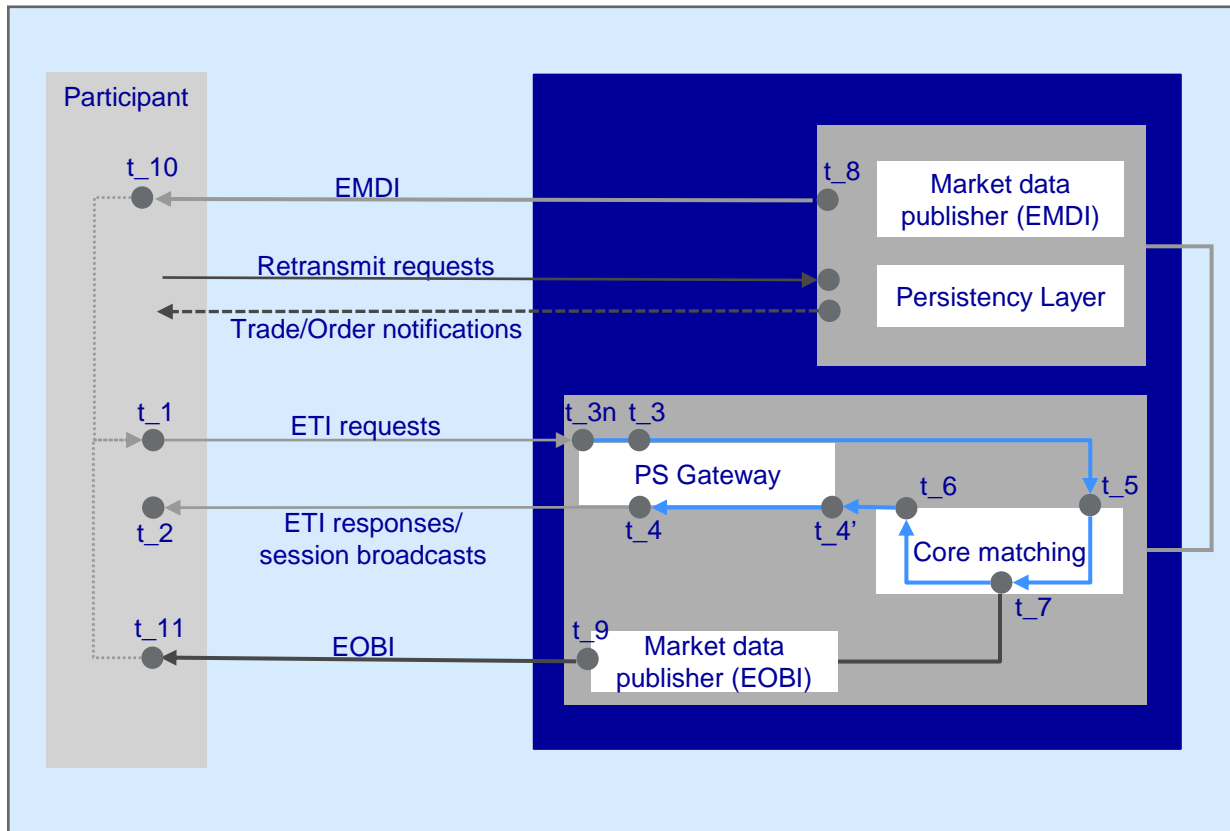


**Matching Engine:**

- order book maintenance & execution
- creation of direct responses as well as execution messages for passive orders/quotes
- creation of EOBI order book messages
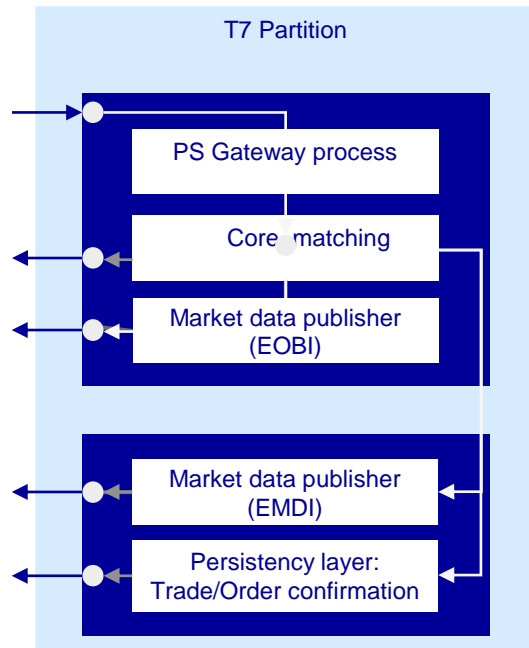- creation of EOBI order book snapshot messages

**Persistency:**

- persistent order storage
- trade/execution history
- transaction history for standard orders
- creation of listener broadcast for standard orders

**Market Data (EMDI):**

- creation of order book delta messages
- creation of order book snapshot messages

# T7® Topology
## Partitions

**T7 Partition**

PS Gateway process

Core matching

Market data publisher (EOBI)

Market data publisher (EMDI)

Persistency layer: Trade/Order confirmation

Orders/quotes entered for a specific product are sent by the PS gateway process to the core Matching Engine process (both residing on the same server in the same partition).

The matching priority is assigned when the orders/quotes are read into the Matching Engine.

The core matching component works as follows:
- when an order/quote arrives, it is functionally processed (e.g. put in the book or matched),
- handover of data to the EOBI market data publisher and
- handover of all data resulting from the (atomic) processing of the incoming order/quote to the market data and persistency components in the partition.

Resulting responses and private broadcasts are sent out in the following order:
- direct response to the order/quote entered (for persistent as well as for non-persistent orders and quotes),
- fast execution information for booked orders/quotes (in case of a match).

The generation of listener broadcasts, trade confirmations (by the persistency server) and of non-EOBI market data (by the market data publisher) is done on separate servers. Hence the order of the resulting messages from these servers is not strictly deterministic.

Note that the Matching Engine holds states of orders in memory. All responses, broadcasts and EOBI market data thus are preliminary by nature.
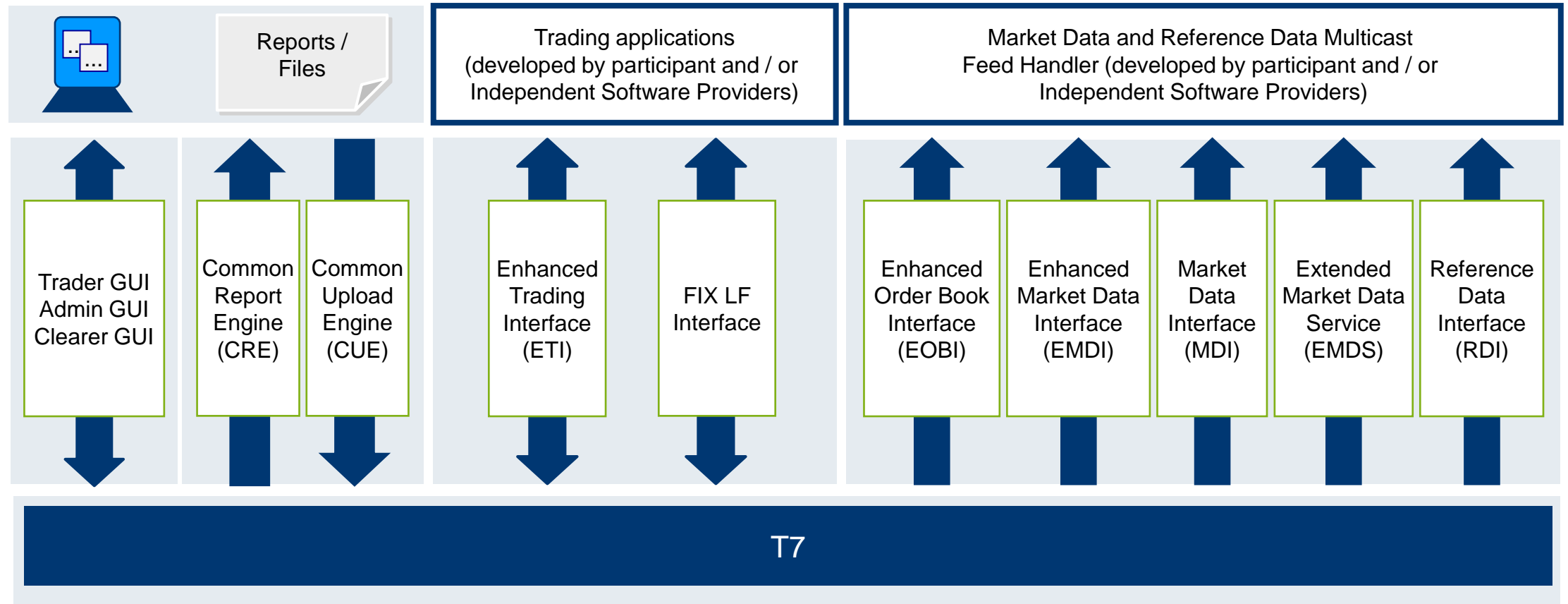
# Network Topology in Co-Location

## Eurex



Customer non 10 Gbit/s
A connection

Customer non 10 Gbit/s
B connection

plus ~ 10µs

LF 2   LF 4

LF 1   LF 3

Side  A

Side  B

PS/Matcher  A1 · · · · PS/Matcher An

PS/Matcher  B1 · · · · PS/Matcher Bn

plus ~ 45µs

Customer room 1

Customer room 2

Customer room …

Customer room 6

Customer room 7
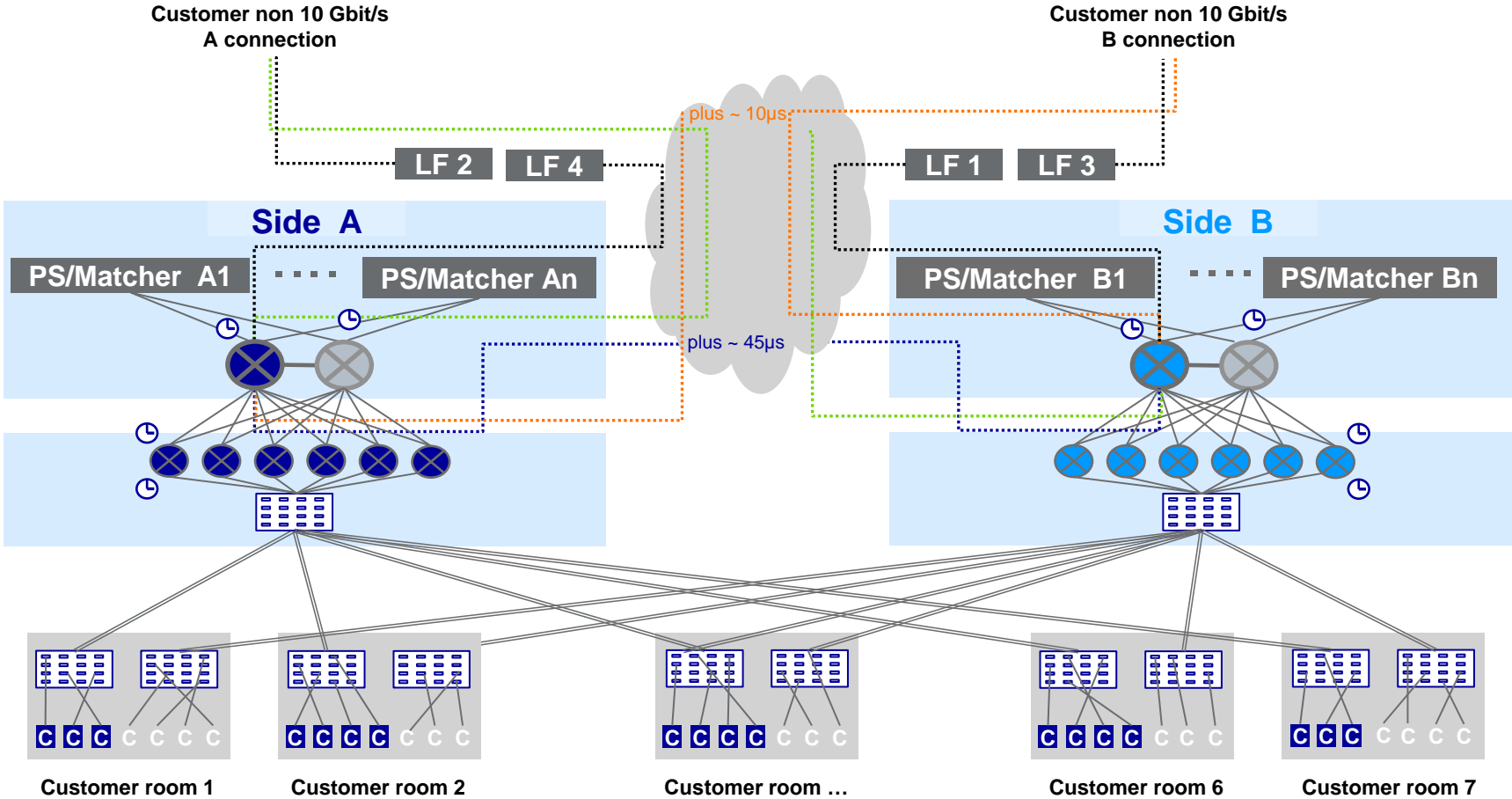
# T7® Trading System
## Interfaces



| | | | | |
|---|---|---|---|---|
| Reports / Files | | Trading applications (developed by participant and / or Independent Software Providers) | | Market Data and Reference Data Multicast Feed Handler (developed by participant and / or Independent Software Providers) |

Trader GUI Admin GUI Clearer GUI

Common Report Engine (CRE)

Common Upload Engine (CUE)

Enhanced Trading Interface (ETI)

FIX LF Interface

Enhanced Order Book Interface (EOBI)

Enhanced Market Data Interface (EMDI)

Market Data Interface (MDI)

Extended Market Data Service (EMDS)

Reference Data Interface (RDI)

**T7**

# Order Entry
## Introduction

Requests sent to T7 will be routed via an access network and a gateway.

There are the following basic connection alternatives:

**Choice of Network**

There are two network classes connecting a participant's installation with the T7 gateways:

- Co-Location with 10 Gbit bandwidth and a one way base latency of around 2 µs.
- Other networks with less than 1Gbit bandwidth and a one way base latency of minimum about 50 µs.

**Choice of Session Type**

T7 supports three session types:

- High frequency sessions connect to PS gateways for low latency access to a single partition (flat binary protocol: ETI).
- Low frequency sessions connect to LF gateways for convenient access to all partitions, with a considerably higher base latency (ETI).
- FIX LF sessions connect to FIX LF gateway for convenient access to all partitions using the FIX protocol, with a considerably higher base latency than LF gateways.

**Remarks**

- LF gateway and FIX LF requests are routed via PS gateways.

# Order Entry
## Co-Location Network

Participants may use Co-Location to place their infrastructure in the datacentre that hosts the T7 system.

The Co-Location 10 Gbit network has the following properties:

**Fair and equal access**

Regardless of the Co-Location room we ensure all lines are created equal.

More precisely the latency between the handover point in the participant's rack and the first (access layer) switch is calibrated to below +/- 2.5 ns.

Deutsche Börse is in the process of reducing this deviation to +/- 35 cm (+/-1.75 ns) – to be finished in the course of 2023.

**Two redundant halves ('A' and 'B')**

There are two independent order entry network halves.

As active gateways are placed in either half there is an optimal side for each gateway (even numbered gateways are on the A side, odd number on the B side). The only exception is FIX LF: There is only one active FIX LF gateway which is by default located on the B-side.

Crossing sides, e.g. connecting to a B side gateway via an A network, is possible but results in at least 45 μs higher base latency.

**Two hierarchical switch layers**

Participants connect to access layer switches (currently 6 Eurex*, 2 Xetra per side).

The uplink of each access layer switch is connected to a distribution layer switch.

The distribution layer switches have a direct connection to the active gateways on the respective side.

* For capacity reasons a 6th access layer switch was added on Eurex A-Side on 13 March 2023.

# Order Entry
## Gateways

There are three gateway types to access the T7 system:

**Partition-specific (PS) Gateway combined with Matching Engine**

Protocol:                    flat binary (ETI)
Allowed session types:       High Frequency Sessions only
Sequencing:                  FIFO operation (Sequence guaranteed from network card to Matching Engine in)
Latency:                     lowest, median latency ~ 12 μs network card to Matching Engine in
Versatility:                 Allows routing to one partition only, only subset of broadcasts available

**Low Frequency (LF) Gateway**

Protocol:                    flat binary (ETI)
Allowed session types:       Low Frequency Sessions only
Sequencing:                  FIFO not guaranteed
Latency:                     medium, (additional ~32 μs latency compared to PS gateway direct access)
Versatility:                 Routes to all partitions (via PS gateway), all ETI broadcast types available

**FIX LF Gateway**

Protocol:                    FIX
Allowed session types:       Fix Sessions only
Sequencing:                  FIFO guaranteed
Latency:                     high, requests to the Matching Engine are routed via PS gateways
Versatility:                 Routes to all partitions (via PS gateway), all FIX broadcast types available

# Market Data
## Overview

Market Data can be consumed over two distinct types of networks and in various types

**Choice of Network**

There are two network classes available for market data:

- Co-Location with 10 Gbit bandwidth and a one-way base latency of around 2 µs.
  10 Gbit connections are equalized in length (cable latency difference of less than +/- 2.5 ns) and provide the lowest jitter.
  Deutsche Börse is in the process of reducing this deviation to +/- 35 cm (+/-1.75 ns) – to be finished in the course of 2023.

- Other networks with less than 1Gbit bandwidth with higher base latency.

**Choice of Market Data Type**

There are three market data types:

- Order by Order market data (EOBI) with highest granularity and lowest latency in flat binary format.
  EOBI is sent out directly from the Matching Engine and is only available via 10 Gbit network.

- Price level aggregated market data (EMDI) with slightly higher latency in FAST encoded format.

- Netted price level aggregated market data (MDI) in FAST encoded format.

# Appendix

# Middleware, Network, Hardware and OS Overview

**T7 uses state-of-the-art infrastructure components**

Intel Xeon Gold 6256 CPU (Cascade Lake Refresh) for Matching Engine and consolidated PS Gateway/Matching Engine.

Intel Xeon Gold 6148 CPU (Skylake)  for all other servers.

We are currently in the migration process to Red Hat Enterprise Linux 8.7.
T7 internal communication between its core components is based on Confinity Low Latency Messaging using an Infiniband network.
In 2024, we will migrate our T7 core network from Infiniband to Ethernet and deploy new server hardware.

**T7 network access**
Deutsche Börse offers trading participants to connect via 10 Gbit/s cross connects to its T7 platform in the Equinix data centre.

The Co-Location offering uses Cisco Nexus 3548-X switches operating in cut-through mode.
All cables are normalized to give an overall maximum deviation between any two cross connects of less than +/- 0.5 m(+/- 2.5 ns).
Deutsche Börse is in the process of reducing this deviation to +/- 35 cm (+/-1.75 ns) – to be finished in the course of 2023.

Insight into network dynamics is offered by the High Precision Timestamp File service (see https://www.mds.deutsche-boerse.com/mds-en/analytics/high-precision-timestamps).

Participant facing interface cards on the gateways and market data publishers use Solarflare EnterpriseOnload wire order delivery API to bypass the kernel TCP stack and deliver messages in the same order received by the network card.

Cables connecting Line-of-Sight antenna cables have been equalized to +/- 1m by Equinix.

# Throttle and Session Limits

**In order to protect the trading system, T7 has several measures in place to ensure that its most vital components are not harmed by a malfunctioning client application. Therefore, transaction limits are imposed on T7 sessions.**

ETI LF sessions are available with throttle values of 150 or 50 transactions/sec  ETI HF sessions are available with throttle values of 250*, 150 or 50 transactions/sec. Furthermore, LF sessions that cannot enter orders/quotes but can only receive trade and listener broadcasts are available (at a reduced price).

The disconnect limit is set at:

- 750 for HF Ultra sessions* with a throttle value of 250 transactions/sec. i.e. a session will be disconnected in case of more than 750 consecutive rejects due to exceeding the transaction limit (throttle).

- 450 for sessions with a throttle value of 150 transactions/sec, i.e. a session will be disconnected in case of more than 450 consecutive rejects due to exceeding the transaction limit (throttle).

- 150 for sessions with a throttle value of 50 transactions/sec, i.e. a session will be disconnected in case of more than 150 consecutive rejects due to exceeding the transaction limit (throttle).

Please note that in case the disaster recover facility is used, all ETI sessions will have a throttle limit of 30 transactions per second.
For both limits, all technical transactions are counted using a sliding window.

The number of ETI sessions which can be ordered is limited. Currently, up to 400 Eurex sessions and 200 Xetra sessions can be ordered.

There is also a limit on the maximum number of sessions that can connect to a PS gateway concurrently per participant. This limit is currently configured to 80 sessions, see Eurex Circular 122/17.

On 1 July 2019, we introduced a limit on the maximum number of outstanding session and trader login requests possible per business unit and per session at any given point in time. This limit is set to 50 on business unit level, 10 on session level. We recommend a synchronous login procedure, where a login request is sent on a session only after the previous login has been responded to. Please refer to the Incident Handling Guide for details.

For Eurex the number of order entry cross connects in colocation that may be used concurrently on a single day is limited to 6 per access layer switch. In addition, the number of allowed ethernet frames per cross connect is currently limited to 25.000 per second and 400.000 per minute.

* Introduced for Eurex and Xetra on 1 June 2023, see Eurex Circular 029/23 and Xetra Circular 021/23.

# T7® Topology
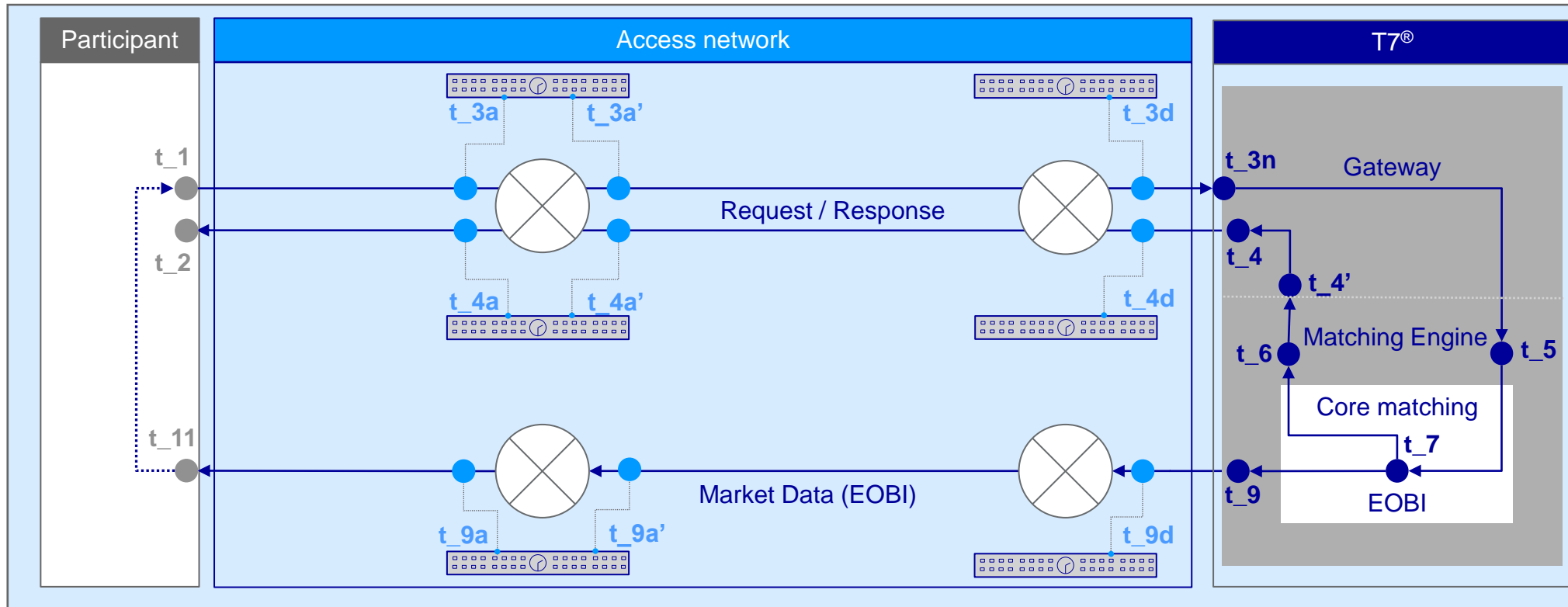
## Timestamps



Timestamps provided in T7 API (in real time) in dark blue (t_3n: taken by network card, other: application level)

Network timestamps taken using TAPs and timestamping switches (Metamako)

Timestamps possibly taken by participants shown in grey

# T7® Timestamps

## Description

t_[x]a, t_[x]a' time taken by network capture devices in the access layers.

t_[x]d        time taken by network capture devices in the distribution layers.

t_1, t_2      can be taken by a participant (e.g. via a network capture) when a request/ response is read from/written to the network.

t_3n          time taken by the PS gateway when the first bit of a request arrives on the PS gateway NIC;
              contained in (private) ETI response for PS gateway enabled partitions.
              Consecutive messages via the same session may be assigned to the same t_3n in rare cases.

t_3           time taken by the ETI gateway application when a request is read from the socket on the participant´s side of the gateway;
              contained in (private) ETI response for transactions for non PS gateway enabled partitions (e.g. XVIE).

t_4'          time taken by the ETI gateway when a response/ notification is received by the ETI gateway from the Matching Engine;
              contained in (private) ETI response/ notification.

t_4           time taken by the ETI gateway when a response/ notification is written to the socket on the participant´s side of the gateway;
              contained in (private) ETI response/ notification.

t_5, t_6      time taken by the Matching Engine when a request/response is read/written; contained in (private) ETI response.

t_7           time at which the Matching Engine starts maintaining the order book

t_8           time taken by EMDI publisher just before the first respective UDP datagram is written to the UDP socket.

t_9           time taken by EOBI publisher just before the first respective UDP datagram is written to the UDP socket.

t_10, t_11    can be taken by a participant (e.g. via a network capture) when a UDP datagram is read from the UDP socket.

# T7® Timestamp Reference

**The timestamps t_3,…,t_9 are available via the following fields:**

| Timestamp | Tag no. | Field name | Present in |
|---|---|---|---|
| t_3, t_3n | 5979 | RequestTime | ETI Response<br>EMDI Depth Incremental message, in case a trade is reported<br>EOBI Execution Summary, Order Add, Order Modify, Order Modify Same Priority and Order Delete messages |
| t_4' | 7765<br>25043 | ResponseIn<br>NotificationIn | ETI Response (from the Matching Engine)<br>ETI Notification (from the Matching Engine) |
| t_4 | 52 | SendingTime | ETI Response and Notification |
| t_5 | 21002<br>2445 | TrdRegTSTimeIn<br>AggressorTime | ETI Response (from the Matching Engine)<br>EMDI Depth Incremental message, in case a trade is reported<br>EOBI Execution Summary message |
| t_6 | 21003 | TrdRegTSTimeOut | ETI Response and Notification (from the Matching Engine) |
| t_7 | 17<br><br>273<br>60<br><br>21008 | ExecID<br><br>MDEntryTime<br>TransactTime<br><br>TrdRegTSTimePriority | ETI Response (from the Matching Engine)<br>EOBI Execution Summary message<br>EMDI Depth Incremental and Top of Book Implied message<br>EMDI messages for other events<br>EOBI Order Modify Same Priority and Order Delete messages<br>EOBI Order Add and Order Modify messages |
| t_8 | No tag | SendingTime | T7 EMDI UDP packet header |
| t_9 | 60 | TransactTime | EOBI packet header |
| t_8 - t_5 | No tag | PerformanceIndicator | EMDI UDP packet header of the T7 EMDI Depth Incremental stream |

**Notes on timestamps:**

All timestamps provided are 8 byte integers (in nanoseconds after Unix epoch).

The PerformanceIndicator is a 4 byte integer (in nanoseconds).

The Network timestamps (t_[x]a, t_[x]a', and t_[x]d, t_[x]d') are not available in any protocol field but some via the High Precision Timestamp File service.

# Thank you for your attention

**Sergej Teverovski, Manfred Sand, Phuong Hieke, Simon Schwanz**
**Trading System Analytics**
Deutsche Börse AG
Mergenthalerallee 61
65760 Eschborn
Germany

E-mail: monitoring@deutsche-boerse.com

For updates refer to
https://www.eurex.com/ex-en/support/technology/t7 and
http://www.xetra.com/xetra-en/technology/t7/publications

DEUTSCHE BÖRSE
GROUP

# Disclaimer