JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2026, VOL 9, NO 1

Modernizing Foreign Exchange Cross-Currency Trading: Challenges of Outdated Banking Infrastructure and the Path Toward Artificial Intelligence-Driven Low-Latency Technology

Ganesh Marimuthu

Comerica, USA

Abstract

The foreign exchange cross-currency trading sector is being faced with the enduring problems that are related to the realization of old technological infrastructure that compromises the competitiveness of institutions and efficiency in operations. The legacy systems that are typified by virtualized desktop systems, remote application delivery systems, batch-processing middleware systems, and monolithic server platforms impose significant latency, restrict real-time risk visibility, and constrain advanced execution capabilities that are necessary in the modern market. These technological limitations are in the form of delayed price discovery, poor quality of execution, higher operating expenses, and increased regulatory exposure. The shift to the modern infrastructure based on low-latency networking, distributed architecture using microservices, artificial intelligence-based decision management systems, and real-time risk management platforms resolves the essential constraints of the legacy environment. The machine learning approaches allow high-level functions such as smart routing of orders, predictive liquidity analytics, dynamic pricing optimization, and detection of anomalies that are beyond modern rule-based systems. A combination of kernel-bypass networking, event processing models, GPU-accelerated computing, and cloud native deployment architecture develops scalable, sustainable trading systems that can handle market data and place orders with response time on the order of a microsecond. Continuous monitoring of the exposure and detection of behavior patterns that are critical to regulatory compliance and prevention of risk before it occurs is achieved through real-time risk engines and AI-enhanced surveillance systems. The modernization framework provides quantifiable benefits in various areas such as the quality of execution, the precision of pricing, efficiency of operations, satisfaction by clients, and being competitive against electronically advanced market participants in the ever-technological-oriented financial markets.

Keywords: Foreign Exchange Trading, Low-Latency Infrastructure, Artificial Intelligence, Microservices Architecture, Real-Time Risk Management.

1. Introduction

The foreign exchange market is the biggest and most highly traded financial market in the world, and the cross-currency trading activities comprise a crucial element of international financial infrastructure that underpins multinational enterprises, institutional investors, as well as banking treasury activities. Although such operations are critical, large segments of the banking industry still use technological architectures that were created in previous decades when electronic trading was new, and market structures were completely distinct than those in the modern world. The inertia of legacy infrastructure such as virtualized desktop systems, remote-client-access systems, and price-oriented middleware that is batch-oriented, generates significant operational drag that is reflected in the form of slow price discovery, poor quality execution, and limited competitive positioning as compared to technology-oriented trading firms. Studies on the connection between latency in trading systems and profitability in market making reveal that increases in speed in processing orders and consumption of market data produce significant benefits in competitiveness of quoted prices and in adverse selection prevention [1]. The technological barriers linked to the use of archaic systems are not just about speed but also about other underlying architectural constraints such as limited scalability at volatile economic times, lack of real time risk transparency, and the ability to execute advanced algorithmic trading patterns. The deployment of artificial intelligence methodologies in the trading infrastructure is a paradigm shift that has allowed a capability that could not be achieved using a conventional rule-based framework and especially in areas like intelligent order routing, predictive liquidity analysis, and dynamic pricing optimization [2]. This paper discusses the complex nature of the difficulties associated with legacy technology infrastructure in the context of foreign exchange cross-currency trading and provides a holistic framework of modernization incorporating low-latency connectivity, distributed microservices architecture, machine learning empowered decision systems, and real-time risk management capabilities that are crucial in establishing competitive positioning in the modern markets.

2. Systemic Challenges of Legacy Infrastructure in Foreign Exchange Operations

The technological debt that has been accrued as a result of such incremental system modification of the system over a long period of time has brought about structural barriers to the smooth cross-currency trading activity within the banking institutions. In legacy trading settings, the virtualized desktop infrastructure is often used as the main means of interaction between the traders, and it can introduce a significant round trip delay between the moment when the user enters the data and when the response is sent by the system, as it involves a series of network hops and graphics rendering needs, which are inherent to the presentation protocols that may be used remotely. Such architectural limitations are especially troublesome at the time of high market volatility when swift price changes require immediate trader responsiveness, which cannot be sufficiently provided by virtualized environments. The algorithmic trading literature underlines that the execution latency directly affects order fill rates and pricing competitiveness, and slower systems record higher rejection rates when in competitive quote situations, when several market makers respond to client requests concurrently to the requesting client [3]. These performance constraints are exacerbated by the application service provider architectures used in the legacy banking systems, which are centralized monolithic server architectures that centralize pricing calculations, risk calculations, and order management functionalities on a single computational node that is not capable of horizontal scalability. The time lapse in risk visibility occasioned by the batch-based paradigm of processing used in these systems is that position exposure computation is periodically updated at fixed intervals instead of being continuously updated with every transaction and market action.

Table 1: Table 1: Legacy System Limitations in FX Trading [3, 4]

Component	Legacy Approach	Impact
Desktop Access	Virtualized/Citrix	High latency, rendering delays
Server Architecture	Monolithic centralized	No horizontal scaling
Risk Calculation	Batch processing	Delayed exposure visibility
Market Data	Batched updates	Information asymmetry

Curve Building	Simple interpolation	Pricing inaccuracies
Order Routing	Static hierarchies	Poor execution quality

The inadequacy of the market data infrastructure of legacy systems significantly impedes the quality and real-time availability of pricing data to the trading systems, which generates information asymmetry in comparison to other trading firms with modern electronic infrastructures. Conventional market data feeds are updated in batched messaging protocols with large refresh rates, whereas modern electronic communication networks are updated in a streaming format with low latency. Both money market rates and foreign exchange spot rates, forward points, and basis swap spreads should be synchronized across different maturity levels and produce an accurate cross-currency forward curve, the accuracy of which in the construction of a curve directly determines the accuracy of prices and the effectiveness of hedges. As noted in the market microstructure literature on foreign exchange operations, the ability to price and the competitiveness of the spreads is highly dependent on the temporal resolution of consumption of market data and the processing speed of the system, with slower systems suffering greater adverse selection costs [4]. Older curve-building algorithms typically employ simplistic interpolation schemes that are unable to reflect complicated non-linear interactions amid currency pairs and interest rate term cases, especially in emerging market currencies, where the liquidity fragmentation defects continuous pricing surfaces. The execution infrastructures of legacy banking systems are usually not engineered to have complex order routing, but rather a fixed set of liquidity provider hierarchies, which are unable to respond dynamically to changing market dynamics of the temporary contraction of liquidity, venue-specific latency, or counterparty credit limits. Lack of smart venue selection tools leads to the deterioration of quality of execution on a variety of levels such as increased effective spreads, increased cost of market impact, and lower fill rates during times of market stress, where the quality of execution can be most important in safeguarding client relationships and institutional profitability.

3. Modern Architectural Foundations and Low-Latency Connectivity Infrastructure

Foreign exchange trading technology transformation demands radical conceptualization of system architecture on the principles of distributed computing, event-based messaging model, and ultra-low-latency networking realizations. Contemporary trading systems provide direct market access connectivity based on kernel-bypass networking technologies, which bypass operating system overhead, providing trading applications with simple access to network interfaces, with latency improvements in the orders of magnitude better than traditional networking stacks. These enhanced networking strategies use user-space drivers, which do not involve the classic kernel network processing, and the result is that they avoid the context switches and the waiting times of interrupt processing, which are significant contributors to total system latency. Technical analysis Studies into low-latency messaging systems in financial applications show that specialized inter-process communication models allowing minimal-latency delivery can provide message delivery between distributed systems components in the microsecond range, where trading infrastructure fundamentally changes its responsiveness [5]. The nature of the architectural change between monolithic application servers and microservices-based distributed applications brings the necessary attributes needed to support modern trading operations, such as independent scaling of components, fault isolation, and continuous deployment features that allow rapid feature development without affecting the entire system.

Container orchestration systems enable dynamic distribution of resources and auto-failover resilience systems, which serve as a significant increase in the availability of a system over the traditional models of static deployment, which means that trading operations continue even when there is a failure or maintenance on the infrastructure. Deployment models based on cloud-native that use hybrid infrastructure designs can help institutions achieve cost-efficiency through leveraging cloud computing resources to perform computationally-intensive analytics and reporting services, and keeping ultra-low-latency trading cores on dedicated on-premise hardware. Combining programmable network infrastructure with software-defined networking functionality allows dynamic traffic scheduling that provides time-sensitive market data streams and order execution streams with prioritized treatment even in the face of network congestion. Studies on microservices architecture underscore the fact that distributed systems architecture has been found to be more scalable and resilient than monolithic ones, with a well-thought-through service decomposition allowing independent scaling of computing resources in line with the workload characteristics [6]. Direct connectivity facilities with various liquidity providers, such as electronic communication networks, swap execution facilities, and bilateral trading relationships, offer platforms with a variety of execution venues through which sophisticated routing algorithms are applied to maximize the quality of executions in view of real-time liquidity evaluation, spread assessment, and historical performance measures. Co-location services, where trading infrastructure is physically co-located in data centers near large liquidity venues, reduce network propagation delay, offering competitive advantages to price discovery and execution quality, especially to latency-sensitive trading programs such as cross-currency arbitrage and statistical trading programs.

4. AI Integration for Pricing Optimization and Intelligent Execution

The use of the artificial intelligence and machine learning methodology in cross-currency trading in foreign exchange is a revolutionary development that has allowed the potential of functionalities that are not practically achievable with traditional quantitative techniques. AI-based pricing models use advanced machine learning models such as ensemble models and deep neural networks to form multi-dimensional pricing surfaces that dynamically integrate a wide range of information sources such as signals on market microstructure, features of order flows, inventory levels, and future volatility. These learning models are in a continuous adaptation of streaming market information, which involves enormous volumes of price observations, which determine the existence of non-linear relationships between currency patterns, interest rate patterns, cross-currency basis patterns, and macro-economic events that cannot be sufficiently explained by the traditional models. Literature studies analyzing deep learning in predicting the financial market reveal a finding that the long short-term memory neural networks have the potential to model temporal effects in financial time series, and they have the ability to learn patterns that can be used to predict stock market outcomes better than traditional econometric models [7]. Natural language processing algorithms examine heterogeneous information feeds, such as news feeds, central bank communications, economic releases, and social media sentiment, in order to create predictive signals of volatility regime changes and directional price changes, which are incorporated into automated pricing adjusting mechanisms that reactively adjust spreads on anticipated market conditions.

AI methodologies in reinforcement learning algorithms: Smart order routing systems are a high-stakes domain of application. AI-based systems have been shown to provide significant performance improvements compared to the traditional rule-based routing logic. These systems are trained on the best execution strategies with lots of simulation and experience, and assess multi-dimensional decision spaces such as venue choice among a large number of liquidity providers, order sizing strategies between immediate execution and algorithmic time-slicing, timing optimization based on predicted price movements and liquidity patterns, and dynamic rebate capture opportunities. Market-making studies based on reinforcement learning show that market agents based on deep Q-learning and policy gradient algorithms can learn and identify execution strategies that surpass traditional reference points by refining, encouraging feedback in the market [8]. Hedge optimization AI-based systems compute the real-time portfolio exposures of currency pairs, maturity structure, and product type, such as spot, forward, swap, and option positions, to approve the most suitable hedging instruments and timing to execute them at the lowest cost without breaching institutional risk exposures. These systems use sophisticated portfolio optimization methods that use transaction costs, market liquidity limits, correlation structure predictions, and market impact predictions to determine combinations of hedges that meet the necessary risk reduction at the lowest possible cost of implementation. Unsupervised learning based anomaly detection models continuously track pricing, execution, and risk indicators to detect unusual trends that may indicate disruption in the market, system malfunction, or behavior that needs urgent intervention to offer early warning abilities, which can be used to manage risk proactively, as opposed to responding to risk.

5. Real-Time Risk Management and Advanced Surveillance Capabilities

Modernization of risk management infrastructure is one of the key components of a wholesome foreign exchange trading platform change, which will also deal with regulatory needs, mitigation of operational risks, and competitive positioning needs at the same time. The inherently out-of-date approach of traditional batch-based risk systems computing position exposures and risk measures at fixed time intervals is fundamentally poor in dealing with dynamic risk in volatile cross-currency markets when major currency pairs undergo large changes in price in a time-limited timeframe as a result of major economic announcements or geopolitical events. Current high-frequency risk engines use streaming computation architectures that process each transaction, update of market data, and external events to ensure that exposure metrics are continually updated with very low calculation latency between trade execution and risk metric update. These systems utilize graphics processing unit accelerated calculations to do parallelized risk calculations such as scenario analysis, stress testing and Monte Carlo simulations which would have taken prohibitively long on traditional central processing unit approaches. Literature review: GPU computing applications Literature review A literature review on the use of parallel processing architectures in financial risk management has shown that computationally intensive financial functions found in quantitative finance can be run in real-time with a significant speedup factor [9]. The timely notification of limit breaches via continuous risk monitoring is also much quicker to identify and remediate than the traditional batch systems and can help limit risk management much more quickly mitigate any potential losses in the unfortunate event of a negative market movement, and can also provide risk managers with timely information to make proactive changes to their portfolio.

Table 2: Risk Management System Comparison [9, 10]

Feature	**Batch System**	**Real-Time System**
Update Frequency	Hourly/Daily	Continuous streaming
Calculation Method	Sequential CPU	GPU parallelized
Breach Detection	Delayed	Immediate
Scenario Analysis	Overnight	On-demand
Surveillance	Rule-based	AI-enhanced patterns
Alert Accuracy	High false positives	ML-optimized

The artificial intelligence-enhanced surveillance systems denote a marked improvement over traditional rule-of-thumb trade surveillance, as machine learning algorithms are used to identify sophisticated patterns of behavior that are indicative of manipulation of the market or errors in the functioning of the operations or some other form of misconduct by the trader in a way that does not necessarily involve specification of rules. Such unsupervised learning schemes examine trading behavior in many ways, such as time series patterns of order placement, patterns of quote updates, patterns of execution prices that are not consistent with market levels, abnormal correlation structure of trading behavior, and communication structure among market participants. ML-based surveillance tools build the baseline of behavioral profiles of traders, trading desks, and client relationships and automatically detect anomalies that violate set behavioral patterns and raise a red flag to be investigated by the compliance team, with a small number of false positives, which inundate the compliance unit. The studies in the field of foreign exchange markets focus on the significance of the efficient surveillance and control systems due to decentralized over-the-counter trading markets with historical cases of coherent manipulation, which have led to serious regulatory fines and reputational losses of associated institutions [10]. The analysis of trader communications through natural language processing of electronic messaging and recorded conversations can be used to automatically identify potentially problematic communications about inappropriate trading behavior and give compliance teams a set of prior review priorities instead of spending their time listening to large volumes of communications manually. Explainable AI methods give transparency to surveillance alert generation, allowing compliance analysts to know which behavioral patterns triggered an alert and aid regulatory examination needs of provisional control effectiveness and governance management.

6. Technology Stack Implementation Framework and Migration Strategy

The actual operation of modernized foreign exchange cross-currency trading infrastructure must take great care in the selection of technology, a gradual migration schedule, as well as thorough validation to sustain the continuity of operations whilst achieving the targeted performance goals. The fundamental trading engine can be based on high-performance time-series databases that are optimized for financial tick data storage and retrieval, high-speed query execution of intricate historical analysis using extensive price history over multiple years of market behavior. The event-based architecture provided by distributed messaging systems represents the infrastructure foundation that ties microservices such as market data normalization, pricing calculation engines, order management systems, risk aggregation services, and reporting parts. These messaging systems are capable of supporting high throughput rates with little end-to-end latency to guarantee trading processes, which are time sensitive, do not slow down when the market is in peak mode, with a good number of messages being delivered in large numbers. The choice of programming language can greatly influence characteristics of system performance, with modern systems typically having latency-critical parts of their systems implemented in systems programming languages to provide the lowest latency overhead and deterministic memory management, and higher-level programming languages used to provide business logic services where microsecond-level latency is less important.

Table 3: Migration Strategy Phases [1, 2, 9]

Phase	Scope	Validation	Rollback Plan
Phase 1	Low-risk pairs	Synthetic replay	Immediate revert
Phase 2	Major pairs	Shadow testing	Gradual reduction
Phase 3	All products	Live monitoring	Staged fallback
Phase 4	Full migration	Performance metrics	Legacy standby

Computing platforms accelerated with graphics processing units can be used to perform computationally intensive tasks in parallel, such as Monte Carlo risk simulation and constructing curves over a wide range of currency pairs and tenors, and inferring machine learning models to provide continuous prediction requests. The contemporary GPU designs with thousands of parallel processing cores are significantly faster than the traditional CPU designs with respect to highly parallelizable code often found in the quantitative finance applications. Selection of cloud infrastructure used in architectures of hybrid deployment takes into account various factors such as network latency to major financial centres, the presence of high-performance computing instances, regulatory compliance needs such as data residency and access controls, and integration capabilities with on-premise infrastructure by means of dedicated private connectivity. Testing and validation systems with synthetic market replay, chaos engineering failure injection, and production traffic shadowing can be used to fully verify system reliability and performance prior to the migration of live trading flows between legacy systems. Staged migration strategies often start with less risky currency pairs or even product categories, gradually expanding coverage as operational confidence is gained and operational performance in the form of execution quality, system availability, and the rate of operational incidents is proven.

Table 4: Performance Metrics Overview [1, 3, 5, 9]

Metric	Legacy	Modern	Improvement
Order Latency	Hundreds of ms	Microseconds	Orders of magnitude
System Availability	Standard	High availability	Significant increase
Risk Update	Hourly intervals	Continuous	Real-time visibility
Scalability	Limited vertical	Elastic horizontal	Dynamic scaling
Execution Quality	Static routing	AI-optimized	Enhanced performance
Surveillance	Rule-based	ML-powered	Pattern detection

Conclusion

The banking institutions that continue their reliance on the old technological infrastructure in conducting their foreign exchange cross-currency trading activities have material competitive disadvantages and needless operational exposures in the modern electronic markets. The essence of virtualized desktop environments, remote application delivery systems, monolithic architectures, and batch-oriented processing paradigms is incapable of providing the performance characteristics, scalability requirements, real-time analytics capability, and intelligent automation aspects required to participate effectively in the market. The transformation model includes several technological aspects such as ultra-low-latency layered networking infrastructure to enable direct connectivity to liquidity venues, with short propagation delay, distributed microservices architecture to enable elastic scalability and operational resilience through independent component management, artificial intelligence-driven systems to optimize pricing accuracy and high execution quality through continuous learning of market patterns and real-time risk management platform to enable proactive exposure control and regulatory compliance through streaming computation. Combining these technological capabilities provides quantifiable operational efficiencies that are expressed in the form of higher quality execution that leads to stronger client relationships, lower transaction costs by effectively using machine learning models to capture the complex market dynamics, more accurate pricing through machine learning models, proactive portfolio management through enhanced risk breach detection, and enhanced surveillance capabilities to meet regulatory compliance needs. Bank of America financial institutions that effectively implement extensive technology transformation programs place themselves in competitive positions with electronically advanced market participants as well as provide scalable and technology-ready platforms to respond to the changing business needs, regulatory demands, and market structure changes across the global foreign exchange markets, where operational excellence and technological capability are the key competitive differentiators.

References

[1] Joel Hasbrouck and Gideon Saar, "Low-latency trading," ScienceDirect, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1386418113000165

[2] Álvaro Cartea et al., “Algorithmic and High-Frequency Trading,” Cambridge University Press, 2015. [Online]. Available: https://books.google.co.in/books?id=5dMmCgAAQBAJ&dq=Algorithmic+and+High-Frequency+Trading&lr=&source=gbs_navlinks_s

[3] Terrence Hendershott et al., "Does algorithmic trading improve liquidity?" Journal of Finance, 2011. [Online]. Available: https://faculty.haas.berkeley.edu/hender/Algo.pdf

[4] Michael R. King et al., "The market microstructure approach to foreign exchange: Looking back and looking forward," ScienceDirect, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0261560613000594

[5] Don MacVittie, "The BIG-IP System and Message Assurance for Low Latency Financial Information eXchange (FIX)". [Online]. Available: https://cdn.studio.f5.com/files/k6fem79d/production/cf8696b8e22fdf34c2b7e29290406c130ffd9f4c.pdf

[6] Nicola Dragoni et al., "Microservices: yesterday, today, and tomorrow," arXiv:1606.04036v4, 2017. [Online]. Available: https://arxiv.org/pdf/1606.04036

[7] Thomas Fischer and Christopher Krauss, "Deep learning with long short-term memory networks for financial market predictions," ScienceDirect, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0377221717310652
[8] Thomas Spooner et al., "Market Making via Reinforcement Learning," arXiv:1804.04216, 2018. [Online]. Available: https://arxiv.org/abs/1804.04216
[9] Abhishek Bhattacharya et al., “Data Mining and Information Security: Proceedings of ICDMIS 2024, Volume 1,” Springer Nature, 2025. [Online]. Available: https://books.google.co.in/books?id=go6LEQAAQBAJ&lpg=PP1&pg=PP1#v=onepage&q&f=false
[10] Alain Chaboud et al., "The foreign exchange market," BIS, 2023. [Online]. Available: https://www.bis.org/publ/work1094.pdf