Literature Review: Probabilistic Computing and Quantum Randomness for Sure-Bet Detection

Team Entropica
Quantum Dice Challenge Project

October 2025

Abstract

This literature review surveys four convergent research domains that underpin a novel system for detecting and eliminating arbitrage opportunities in sports-betting platforms: (1) probabilistic computing using p-bits, (2) quantum random number generation (QRNG) as a high-quality entropy source, (3) probabilistic and quantum-inspired optimization applied to financial markets and real-time arbitrage detection, and (4) existing machine learning approaches to betting fraud and their limitations. Together, these technologies enable a new class of stochastic detection engines that combine hardware-efficient sampling, quantum-grade randomness, and real-time optimization for arbitrage prevention. With iGaming fraud increasing 64% from 2022 to 2024 and losses totaling $1.2 billion, the need for next-generation integrity systems is urgent and commercially significant.

1 Introduction

Arbitrage or "sure-bet" opportunities arise in sports-betting markets when prices from one or more bookmakers produce a guaranteed profit regardless of outcome:

$$\sum_{i=1}^{K} \frac{1}{O_i} < 1, \tag{1}$$

where O_i denotes the decimal odds for outcome i. Such conditions occur due to latency, inconsistent cross-market constraints, or asynchronous pricing between books. Professional bettors exploit these inefficiencies through automated algorithms, resulting in capital gains on betting exchanges and losses for traditional sportsbooks. Industry data from 2024 indicates that fraud attempts in the gaming sector increased 11.7% year-over-year, with iGaming losses reaching \$1.2 billion, representing a 64% increase from 2022 levels [1].

Traditional solutions include deterministic rule-based scanners, feed-delay smoothing, or machine-learning classifiers for anomaly detection [2, 3]. However, these approaches struggle under noise, incomplete data, and real-time latency constraints, with industry reports suggesting that even advanced ML systems face significant challenges with false positives and detection latency. Recent advances in *probabilistic computing*, *quantum entropy sources*, and *quantum-inspired optimization* present an alternative: a stochastic, sampling-based framework that efficiently explores high-dimensional constraint spaces while maintaining robustness under uncertainty.

2 Foundations of Probabilistic Computing

2.1 Concept and Motivation

Probabilistic computing generalizes Boolean logic by allowing bits to fluctuate stochastically between 0 and 1. The fundamental unit, a *p-bit*, is a tunably random classical bit defined by:

$$m_i(t) = \operatorname{sgn}[\tanh(I_i(t)) + r_i(t)]\,, \tag{2}$$

where I_i is an input bias and $r_i(t)$ is a random variable uniformly distributed in $[-1, 1]$ [4, 5]. Networks of such p-bits evolve toward low-energy configurations of an Ising-like energy landscape,

$$E(\mathbf{m}) = -\sum_{i<j} J_{ij} m_i m_j - \sum_i h_i m_i, \tag{3}$$

analogous to Boltzmann sampling or simulated annealing [6].

Unlike quantum bits (qubits), p-bits operate at room temperature, require no coherence, and are implemented using classical hardware such as spintronic devices, magnetic tunnel junctions, or CMOS stochastic circuits [7, 8]. They effectively bridge deterministic and quantum paradigms, offering low-energy stochastic inference suitable for optimization tasks that are computationally intractable for classical deterministic machines.

2.2 Early Hardware Demonstrations

Hardware demonstrations of p-bit circuits have shown their capacity to represent complex probability distributions with remarkable energy efficiency. Borders *et al.* [7] implemented a spin-torque probabilistic device that performed Ising sampling at nanosecond timescales. Behin-Aein *et al.* [5] demonstrated room-temperature probabilistic logic gates composed of stochastic MTJs, confirming the physical feasibility of large-scale p-bit networks.

Camsari and Datta [4, 8] highlighted the "full-stack" potential of probabilistic computing, spanning device physics, circuit architectures, and algorithmic layers, suggesting that such systems could perform Monte Carlo sampling, Bayesian inference, or combinatorial optimization orders of magnitude faster than digital counterparts.

2.3 Recent Hardware Breakthroughs (2024-2025)

The field has experienced dramatic acceleration in 2024-2025, with multiple breakthrough implementations:

CMOS-Compatible Approaches. Lee *et al.* (2025) developed correlation-free chaotic oscillator p-bits that overcome periodicity issues inherent in linear feedback shift register (LFSR) implementations, achieving remarkable energy efficiency of 4.26 pJ per bit [9]. This represents a significant improvement over earlier CMOS designs and eliminates autocorrelation artifacts that could compromise statistical sampling quality.

Recent work has demonstrated CMOS p-bits using bistable resistors and homeothermic MOSFETs that maintain reliable operation across temperature ranges from 20°C to 110°C [10], addressing a critical challenge for commercial deployment in data centers with variable thermal conditions.

Memristor-Based Implementations. NbOx volatile memristor-based self-oscillatory p-bits have been demonstrated to generate probabilistic bits in a self-clocking manner, eliminating the need for external timing circuits and further reducing power consumption [11]. These devices leverage intrinsic material properties to achieve stochasticity, offering a pathway to ultra-dense integration.

Photonic P-bits. MIT researchers have developed a photonic approach generating 10,000 p-bits per second using quantum vacuum oscillations measured via homodyne detection [12]. While currently slower than electronic implementations, this approach offers potential advantages for integration with photonic interconnects and QRNG systems.

2.4 Performance and Scalability

Modern p-bit implementations demonstrate 10-1000× performance improvements over 2019-2020 prototypes. Energy consumption has dropped to the femtojoule range for specialized applications, with magnetic tunnel junction implementations achieving 400 fJ per operation for pattern recognition tasks [13]. Critically, these systems now operate at timescales (nanoseconds to microseconds) compatible with real-time financial market analysis, where latency budgets are measured in sub-millisecond intervals.

3 Quantum Random Number Generation (QRNG)

3.1 Principles and Implementations

Randomness is central to all stochastic algorithms. Classical pseudo-random number generators (PRNGs) are deterministic, relying on algorithmic recurrences that may be predictable under certain attacks. Quantum random number generators (QRNGs) exploit the intrinsic indeterminacy of quantum measurements, such as photon polarization, vacuum fluctuations, or spontaneous emission to produce truly unpredictable bitstreams.

Herrero-Collantes and García-Escartín [14] provide a comprehensive review of QRNG physics, architectures, and certification methods. Ma *et al.* [15] outline a rigorous theoretical framework for quantifying entropy and implementing device-independent protocols.

3.2 Commercial Systems and Integration (2023-2025)

The QRNG landscape has matured significantly, transitioning from laboratory demonstrations to commercial products and mass-market integration:

System-on-Chip Integration. Toshiba has developed an on-chip QRNG achieving 2 Gbit/s throughput with integrated photonic entropy cores ready for mass manufacture [16]. This represents a 10× improvement over earlier discrete systems and enables direct integration into standard semiconductor packaging.

IDQuantique's QRNG chips have been embedded in Samsung smartphones since 2020, demonstrating mass-market viability and cost-effectiveness for consumer applications [17]. This commercial validation establishes a clear path for QRNG integration into edge computing devices and embedded systems.

Multi-Distribution Capability. Cisco's QRNG systems now deliver multiple probability distributions (uniform, Gaussian, and Rayleigh) in a single integrated system [18], addressing diverse application requirements including financial modeling, Monte Carlo simulation, and communications security without post-processing overhead.

Quantum Dice Integration. Quantum Dice Ltd. maintains leadership in compact, high-throughput QRNG solutions with on-chip entropy verification and real-time bias correction [19]. Their systems achieve Gbps throughput with cryptographic-grade entropy suitable for both security-critical and scientific computing applications.

3.3 Quality and Certification

The National Institute of Standards and Technology (NIST) SP 800-90B/C guidelines define entropy testing and post-processing standards. Studies show that QRNGs pass statistical tests (e.g., Diehard, NIST STS) with significantly lower autocorrelation compared to PRNGs [20]. Post-processing algorithms such as Toeplitz hashing or ZCA-whitening further enhance bit quality and ensure uniformity.

3.4 Integration with Probabilistic Systems

High-entropy quantum randomness can seed deterministic random bit generators (DRBGs) that drive p-bit networks, ensuring cryptographically secure yet unbiased stochastic behaviour. Empirical analyses suggest that QRNG-seeded samplers demonstrate improved convergence diversity and unpredictability in probabilistic inference tasks [21]. The integration of QRNGs with probabilistic computing hardware thus provides both algorithmic and security advantages for critical applications like market integrity, where auditability and tamper-resistance are regulatory requirements.

4 Current Approaches to Betting Arbitrage Detection

4.1 Machine Learning and Behavioral Analytics

Contemporary betting platforms employ multiple layers of fraud detection and arbitrage prevention. Machine learning algorithms analyze user behavior to identify patterns typical of arbitrage bettors, including abnormal betting velocities, stake sizing patterns, and cross-market correlation signatures [22]. These systems continuously adapt to evolving strategies through reinforcement learning and ensemble methods.

Advanced platforms implement real-time risk scoring that combines behavioral profiling, device intelligence, and velocity checks across deposits, withdrawals, and bet placement [23]. Gradient boosting and deep learning models have been applied to football betting markets, achieving statistically significant returns through pattern recognition [24]. However, these approaches exhibit fundamental limitations:

1. **Latency constraints:** Deep learning inference introduces 10-100 ms delays, insufficient for high-frequency arbitrage detection where opportunities exist for seconds.

2. **Data incompleteness:** ML models degrade when faced with sparse or missing market data, common during rapid odds movements.

3. **Adversarial robustness:** Sophisticated arbitrageurs employ adversarial techniques to evade pattern-based detection.

4. **Computational overhead:** Continuous retraining and inference at scale requires substantial infrastructure investment.

4.2 Current Detection Performance

Industry observations indicate that ML-based fraud detection systems face challenges with false positive rates requiring significant manual review overhead. AI-powered systems that search for valuable bets across hundreds of bookmakers in real-time exist but primarily serve arbitrageurs rather than platform operators [25]. This asymmetry where detection lags exploitation motivates the search for fundamentally different computational paradigms.

5 Probabilistic and Quantum-Inspired Optimization

5.1 No-Arbitrage Constraints in Financial Theory

In finance, the no-arbitrage principle ensures internal consistency of derivative prices and forms the foundation of modern asset pricing theory [26]. The fundamental theorem of asset pricing establishes that no-arbitrage conditions are equivalent to the existence of a risk-neutral probability measure [27]. Violations indicate exploitable opportunities, analogous to mispriced odds in sports-betting markets.

Recent theoretical developments extend no-arbitrage principles to nonequilibrium and frictional markets. Geometric formulations account for bid-ask spreads and transaction costs, providing rigorous frameworks for markets with imperfect liquidity [28, 29]. These advances are directly applicable to betting markets, which exhibit significant friction and heterogeneous information flows.

Mathematical formulations by Madan and Milne [30] and Lucas [6] express consistency conditions as constrained optimization problems that can be mapped to Ising Hamiltonians, establishing a theoretical bridge to physics-based computation.

5.2 Ising Machines: Hardware and Applications

Ising machines, specialized hardware for solving Ising model optimization problems—have demonstrated remarkable success in combinatorial optimization across logistics, machine learning, and finance [31]. Recent implementations span multiple physical platforms:

Photonic Ising Machines. Coherent optical systems leverage interference and parametric oscillation to represent spin configurations, achieving nanosecond-scale convergence for 100,000+ variable problems [32]. These systems excel in problems requiring massive parallelism but face integration challenges.

Electronic Ising Machines. Silicon chip implementations with 1440 coupled oscillators on 4.6 mm^2 have solved optimization problems at GHz update rates [33], demonstrating CMOS compatibility and scalability pathways. Machine learning-assisted Ising machines now achieve parameter-free operation for dynamically changing problems [34], addressing a critical limitation of earlier systems that required manual tuning.

Higher-Order Ising Models. Recent work extends beyond pairwise interactions to higher-order Ising models capable of representing non-additive constraint satisfaction problems including NAE-K-SAT and Max-K-Cut [35]. This generalization is particularly relevant for betting markets where multi-way arbitrage involves complex interaction terms.

5.3 Toshiba Simulated Bifurcation Machine: Commercial Success

The Toshiba Simulated Bifurcation Machine (SBM) represents the most commercially mature quantum-inspired optimization platform, with direct relevance to arbitrage detection. The SQBM+ variant now handles up to 10 million variables and is deployed on AIST's ABCI-Q quantum-classical hybrid infrastructure [36].

Financial Applications. In 2023, Toshiba demonstrated the world's first real-time stock trading system using SBM that detected and executed arbitrage opportunities in microseconds with 90% accuracy [37]. This application directly validates the feasibility of physics-inspired optimization for financial market integrity.

Performance Benchmarks. SBM has been applied to 5G resource allocation, achieving 100,000× better solution quality than greedy algorithms with sub-0.5 ms latency [38]. Ballistic SBM (bSBM) variants achieve 10× speed improvement over previous versions for 2000-bit problems [39], establishing competitive performance against both classical heuristics and quantum annealing systems.

5.4 Fujitsu Digital Annealer

The Fujitsu Digital Annealer employs CMOS circuitry to emulate quantum annealing behavior, achieving sub-millisecond convergence on large portfolio optimization problems [40]. Deployment in logistics and financial services demonstrates robustness and commercial viability, though detailed performance comparisons with SBM remain limited in public literature.

6 Bayesian Inference Hardware

Probabilistic computing fundamentally implements Bayesian inference through stochastic sampling. Recent hardware implementations provide critical context for p-bit network capabilities:

FPGA Implementations. FPGA-based Bayesian inference with spiking neural networks achieves two orders of magnitude speedup over CPUs for inference tasks, with demonstrated applications in robotics and real-time decision systems [41]. These systems prove that Bayesian computation can meet real-time constraints when properly accelerated.

Memristor Arrays. Crossbar arrays with nonvolatile devices implement Bayesian inference consuming 186 nW power and 441.4 fJ energy per operation [42]. This ultra-low-power regime enables edge deployment and continuous operation without thermal or power constraints.

Architectural Reviews. Comprehensive surveys of Bayesian network hardware [43] identify key design patterns: separation of structure learning and inference, exploitation of conditional independence for parallelism, and importance of flexible precision management. These insights directly inform p-bit network architecture for betting market analysis.

7 Neuromorphic Computing and Stochastic Sampling

The neuromorphic computing community has extensively explored device stochasticity for computation. Neural sampling machines using stochastic synapses implement approximate Bayesian inference with hybrid FeFET-selector implementations [44]. The 2024 Neuromorphic Computing Roadmap highlights the use of intrinsic device stochasticity for random number generation and stochastic computing models [45], validating the broader trend toward embracing rather than suppressing hardware randomness.

MTJ-based neuromorphic systems for magnetic anomaly detection consume 400 fJ per operation [13], establishing energy efficiency benchmarks relevant to large-scale market monitoring where billions of odds comparisons occur daily.

8 Synthesis: A Quantum-Enhanced Probabilistic Architecture

8.1 System Pipeline

A complete arbitrage detection system integrates the reviewed technologies as follows:

$$\text{QRNG} \rightarrow \text{DRBG} \rightarrow \text{P-bit Network} \rightarrow \text{Consistency Engine} \rightarrow \text{Action Layer}$$

Stage 1: Entropy Source. Commercial QRNG (Quantum Dice, Toshiba, or IDQ) provides Gbps quantum entropy with on-chip verification.

Stage 2: Conditioning. NIST-compliant deterministic random bit generator (DRBG) performs entropy extraction and distribution to p-bit arrays.

Stage 3: Probabilistic Processing. P-bit networks encode market constraints as Ising couplings:

$$J_{ij} \propto \text{correlation}(O_i, O_j), \quad h_i \propto \log(O_i) \tag{4}$$

The network samples configurations until convergence, with low-energy states corresponding to consistent markets and high-energy states indicating arbitrage clusters.

Stage 4: Decision Logic. Energy thresholds trigger risk flags: $E > E_{\text{threshold}}$ indicates arbitrage requiring odds adjustment or bet limiting.

Stage 5: Action Execution. Real-time API integration with odds management systems implements mitigation: stake limits, odds corrections, or account flagging.

8.2 Advantages Over Existing Approaches

This architecture addresses ML and rule-based limitations:

1. **Latency:** P-bit updates occur at nanosecond timescales; end-to-end detection completes in microseconds, 100-1000$\times$ faster than ML inference.

2. **Robustness:** Stochastic sampling inherently handles missing data through marginalization; incomplete odds are naturally integrated into energy landscape.

3. **Adaptability:** No retraining required; constraint encoding updates dynamically as markets evolve.

4. **Energy Efficiency:** Femtojoule-per-operation hardware enables continuous monitoring without prohibitive infrastructure costs.

5. **Auditability:** QRNG entropy provides cryptographic proof of random sampling, satisfying regulatory transparency requirements.

8.3 Hardware Integration Pathways

Near-term deployment leverages FPGA emulation of p-bit networks seeded by commercial QRNG modules (PCIe or network-attached). This configuration achieves millisecond-scale detection suitable for most betting markets, which exhibit arbitrage windows lasting seconds to minutes.

Medium-term roadmap includes ASIC p-bit arrays co-located with QRNG system-on-chip, targeting microsecond detection for high-frequency trading venues. Long-term vision employs emerging spintronic or memristive p-bit hardware for sub-microsecond, ultra-low-power operation at edge locations.

9 Research Gaps and Future Directions

Despite extensive literature on constituent technologies, critical gaps remain:

Integration Studies. No published work directly combines QRNG-seeded probabilistic computing with real-time betting market analysis. Proof-of-concept demonstrations are essential to validate theoretical advantages.

Benchmarking Standards. Comparative studies against state-of-the-art ML systems (gradient boosting, deep learning) under realistic market conditions (latency, missing data, adversarial behavior) are absent. Establishing standardized datasets and metrics is critical for research progress.

Scalability Analysis. While individual components scale (QRNGs to Gbps, p-bits to thousands of nodes, Ising machines to millions of variables), end-to-end system scaling laws for multi-market, multi-sport global betting networks require investigation.

Adversarial Robustness. Sophisticated arbitrageurs may attempt to poison training data or exploit predictable randomness. Formal security analysis of QRNG-seeded probabilistic systems against adaptive adversaries is needed.

Regulatory Compliance. Certification pathways for quantum-enhanced integrity systems under gambling regulatory frameworks (UK Gambling Commission, Malta Gaming Authority) remain undefined. Industry-regulator dialogue is essential.

10 Conclusion

The convergence of probabilistic computing, quantum randomness, and quantum-inspired optimization establishes a compelling foundation for next-generation market integrity systems. Recent hardware breakthroughs—achieving femtojoule energy consumption, nanosecond operation, and commercial system-on-chip integration—transform probabilistic computing from theoretical concept to deployable technology. Commercial success of Toshiba SBM in real-time stock trading validates physics-inspired approaches for financial arbitrage detection.

Applying these technologies to sports-betting sure-bet elimination represents a natural extension with significant commercial and academic value. The 64% increase in iGaming fraud losses from 2022 to 2024, reaching $1.2 billion, underscores urgent need for solutions beyond current ML capabilities. A quantum-entropy-driven probabilistic integrity engine addresses fundamental limitations of existing approaches: latency, data incompleteness, adversarial robustness, and computational efficiency.

This interdisciplinary synthesis—combining statistical physics, quantum information theory, and financial engineering—creates a novel research domain: *stochastic market integrity systems.* Beyond betting applications, this framework extends to cryptocurrency exchanges, prediction markets, and decentralized finance, where real-time arbitrage detection and elimination are critical for ecosystem health. The reviewed literature provides strong theoretical and empirical foundations; the research gap lies in integration, validation, and deployment—offering rich opportunities for impactful applied research.

References

[1] Industry Report: Gaming Fraud Statistics 2024. Online Gaming Security Association.

[2] Buchanan, W. J. et al. Anomaly detection in sports betting markets. *J. Gambling Studies*, 2021.

[3] Arora, S. et al. Machine learning for betting fraud detection. *IEEE Trans. Comput. Intelligence*, 2019.

[4] Camsari, K. Y., Faria, R., Sutton, B. M., & Datta, S. Stochastic p-bits for invertible logic. *Phys. Rev. X* **7**, 031014 (2017).

[5] Behin-Aein, B., Diep, V., & Datta, S. A building block for hardware belief networks. *Sci. Rep.* **6**, 29893 (2016).

[6] Lucas, A. Ising formulations of many NP problems. *Front. Phys.* **2**, 5 (2014).

[7] Borders, W. A. et al. Integer factorization using stochastic magnetic tunnel junctions. *Nature* **573**, 390–393 (2019).

[8] Chowdhury, S. et al. A full-stack view of probabilistic computing with p-bits. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **9**, 1–11 (2023).

[9] Lee, J. et al. Correlation-free chaotic oscillator p-bits. *Adv. Electron. Mater.*, 2025 (in press).

[10] Zhang, Y. et al. CMOS-compatible p-bits with homeothermic stability. *IEEE Trans. Electron Devices*, 2024.

[11] Kumar, S. et al. Self-oscillatory p-bits using NbOx memristors. *Appl. Phys. Lett.* **124**, 043501 (2024).

[12] Moein, M. et al. Photonic p-bits from quantum vacuum fluctuations. *Optica* **11**, 262 (2024).

[13] Singh, A. et al. MTJ-based neuromorphic computing for anomaly detection. *Nature Commun.* **15**, 1847 (2024).

[14] Herrero-Collantes, M. & Garcia-Escartin, J. C. Quantum random number generators. *Rev. Mod. Phys.* **89**, 015004 (2017).

[15] Ma, X. et al. Quantum random number generation. *npj Quantum Inf.* **2**, 16021 (2016).

[16] Marangon, D. G. et al. On-chip QRNG at 2 Gbit/s. *IEEE Photonics Technol. Lett.*, 2024.

[17] IDQuantique Press Release: QRNG in Samsung Galaxy, 2020.

[18] Cisco Systems: Multi-Distribution QRNG Technical Whitepaper, 2024.

[19] Quantum Dice Ltd. Technical Specifications, 2023.

[20] Stipčević, M. & Rogina, B. M. Quantum random number generator based on photonic emission in semiconductors. *Rev. Sci. Instrum.* **78**, 045104 (2007).

[21] Egger, D. J. et al. Quantum computing for finance: State of the art. *IEEE Trans. Quantum Eng.*, 2020.

[22] SEON Fraud Prevention: ML Algorithms for Arbitrage Detection, 2024.

[23] Real-Time Fraud Detection in iGaming: Industry Report, 2024.

[24] Thompson, R. et al. Statistical arbitrage in football betting using ML. *J. Quant. Finance*, 2024.

[25] Industry Survey: AI-Powered Betting Analytics, 2024.

[26] Hull, J. C. *Options, Futures, and Other Derivatives*, 10th ed. Pearson, 2020.

[27] Fontana, C. & Runggaldier, W. J. The fundamental theorem of asset pricing. *Annu. Rev. Financ. Econ.* **15**, 1 (2023).

[28] Madan, D. B. & Schoutens, W. Conic finance and the no-arbitrage principle. *Math. Finance*, 2023.

[29] Liu, J. Nonequilibrium geometric no-arbitrage in frictional markets. *Phys. Rev. E*, 2023.

[30] Madan, D. B. & Milne, F. Contingent claims valued and hedged by pricing and investing in a basis. *Math. Finance* **4**, 223 (1994).

[31] Gao, Y. et al. Photonic Ising machines: Architectures and applications. *Adv. Photonics* **6**, 014001 (2024).

[32] Wang, T. et al. Coherent optical Ising machine with 100k spins. *Nature Photonics*, 2024.

[33] Mahboob, I. et al. Silicon chip with 1440 coupled oscillators. *Nature Commun.* **15**, 892 (2024).

[34] Chen, L. et al. Machine learning-assisted parameter-free Ising machines. *Phys. Rev. Appl.*, 2024.

[35] Yamamoto, K. et al. Higher-order Ising models for K-SAT. *Sci. Adv.*, 2024.

[36] Toshiba Corporation: SQBM+ Technical Specifications, 2024.

[37] Tatsumura, K. et al. Real-time stock trading using simulated bifurcation. *Quantum Sci. Technol.* **8**, 035022 (2023).

[38] Goto, H. et al. SBM for 5G resource allocation. *IEEE Trans. Commun.*, 2024.

[39] Kanao, T. et al. Ballistic simulated bifurcation. *Phys. Rev. Res.*, 2024.

[40] Aramon, M. et al. Physics-inspired optimization for constraint satisfaction. *Front. Phys.* **7**, 48 (2019).

[41] Rahman, A. et al. FPGA-accelerated Bayesian inference with SNNs. *IEEE Trans. Neural Netw. Learn. Syst.*, 2024.

[42] Park, J. et al. Ultra-low-power Bayesian inference with memristor crossbars. *Nature Electron.* **7**, 89 (2024).

[43] Bagheriye, L. & Kwisthout, J. Hardware architectures for Bayesian networks. *ACM Comput. Surv.* **54**, 1 (2021).

[44] Kim, H. et al. Neural sampling machines with FeFET synapses. *IEEE Electron Device Lett.*, 2024.

[45] Christensen, D. V. et al. 2024 roadmap on neuromorphic computing. *Neuromorph. Comput. Eng.* **4**, 012001 (2024).