



Markov Chain Abstractions of Electrochemical Reaction-Diffusion in Synaptic Transmission for Neuromorphic Computing

Margot Wagner^{1,2,3*}, Thomas M. Bartol^{3,4}, Terrence J. Sejnowski^{1,2,3,4,5} and Gert Cauwenberghs^{1,2,5}

¹ Department of Bioengineering, University of California, San Diego, La Jolla, CA, United States, ² Institute for Neural Computation, University of California, San Diego, La Jolla, CA, United States, ³ Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, United States, ⁴ Center for Theoretical Biological Physics, University of California, San Diego, La Jolla, CA, United States, ⁵ Division of Biological Sciences, University of California, San Diego, La Jolla, CA, United States

OPEN ACCESS

Edited by:

Erika Covi,
NaMLab gGmbH, Germany

Reviewed by:

Stanisław Woźniak,
IBM Research-Zurich, Switzerland
Fabien Allibert,
UMR8520 Institut d'électronique, de
microélectronique et de
nanotechnologie (IEMN), France

*Correspondence:

Margot Wagner
mwagner@ucsd.edu

Specialty section:

This article was submitted to
Neuromorphic Engineering,
a section of the journal
Frontiers in Neuroscience

Received: 21 April 2021

Accepted: 04 November 2021

Published: 29 November 2021

Citation:

Wagner M, Bartol TM, Sejnowski TJ
and Cauwenberghs G (2021) Markov
Chain Abstractions of Electrochemical
Reaction-Diffusion in Synaptic
Transmission for Neuromorphic
Computing.
Front. Neurosci. 15:698635.
doi: 10.3389/fnins.2021.698635

Progress in computational neuroscience toward understanding brain function is challenged both by the complexity of molecular-scale electrochemical interactions at the level of individual neurons and synapses and the dimensionality of network dynamics across the brain covering a vast range of spatial and temporal scales. Our work abstracts an existing highly detailed, biophysically realistic 3D reaction-diffusion model of a chemical synapse to a compact internal state space representation that maps onto parallel neuromorphic hardware for efficient emulation at a very large scale and offers near-equivalence in input-output dynamics while preserving biologically interpretable tunable parameters.

Keywords: neuromorphic, synapse, Markov chain, Monte Carlo, synaptic transmission, computational efficiency

1. INTRODUCTION

It has been known since the pioneering of computer architecture by John von Neumann that brains are far more effective and efficient in processing sensory information than digital computers, owing to the massively parallel distributed organization of neural circuits in the brain that tightly couple synaptic memory and computing at a fine grain scale (von Neumann, 1958). Modern day computers still follow the “von Neumann” architecture where computing and memory are kept separate, incurring severe penalties in computing bandwidth due to the bottleneck in data flow between centralized processing and vast memory. Moore’s law’s relentless scaling of semiconductor technology, with a doubling of integration density every 2 years, has allowed the von Neumann architecture to remain fundamentally unchanged since its advent. As the shrinking dimensions of transistors supporting the progression of Moore’s law are approaching fundamental limits, it has become essential to consider alternative novel computing architectures to meet increasing computational needs in this age of the deep learning revolution, which itself is driven by advances rooted in a deeper understanding of brain function (Sejnowski, 2020). At the forefront of this movement are neuromorphic systems, introduced by Mead (1990) as a solution to these limitations. Neuromorphic engineering looks toward human brains as inspiration for hardware systems due to their highly efficient computational nature. The human brain is regarded as the pinnacle of efficient computing, operating at an estimated rate of 10^{16} complex

operations per second while consuming less than 20 W of power (Churchland and Sejnowski, 1992; Cauwenberghs, 2013). Therefore, neuromorphic engineering looks to mimic the function and organization of neural structures using hybrid analog and digital systems. This is possible because there is significant overlap in the physics of computation between the brain and neuromorphic engineering (Figure 1). In both systems, information is carried in the form of charge, and, in hardware, neuronal membrane dynamics are represented using metal-oxide-semiconductor field-effect transistors (MOSFETs) (Mead, 1989). In the MOSFET sub-threshold region of operation, electrons and holes are the carriers of current between *n*- or *p*-type channels and behave akin to ions flowing through ion channels that mediate current across the neuronal cell membrane. Fundamentally, these hardware systems share analogous properties to their biological counterparts, including charge stochasticity, diffusion as the primary mechanism of carrier transport, and energy barriers modulated by gating voltage. Paired with Boltzmann distributions of charge, these systems are able to emulate current as an exponential function of the applied voltage, capturing the same biophysics underlying the neuronal dynamics (Mead, 1989; Broccard et al., 2017).

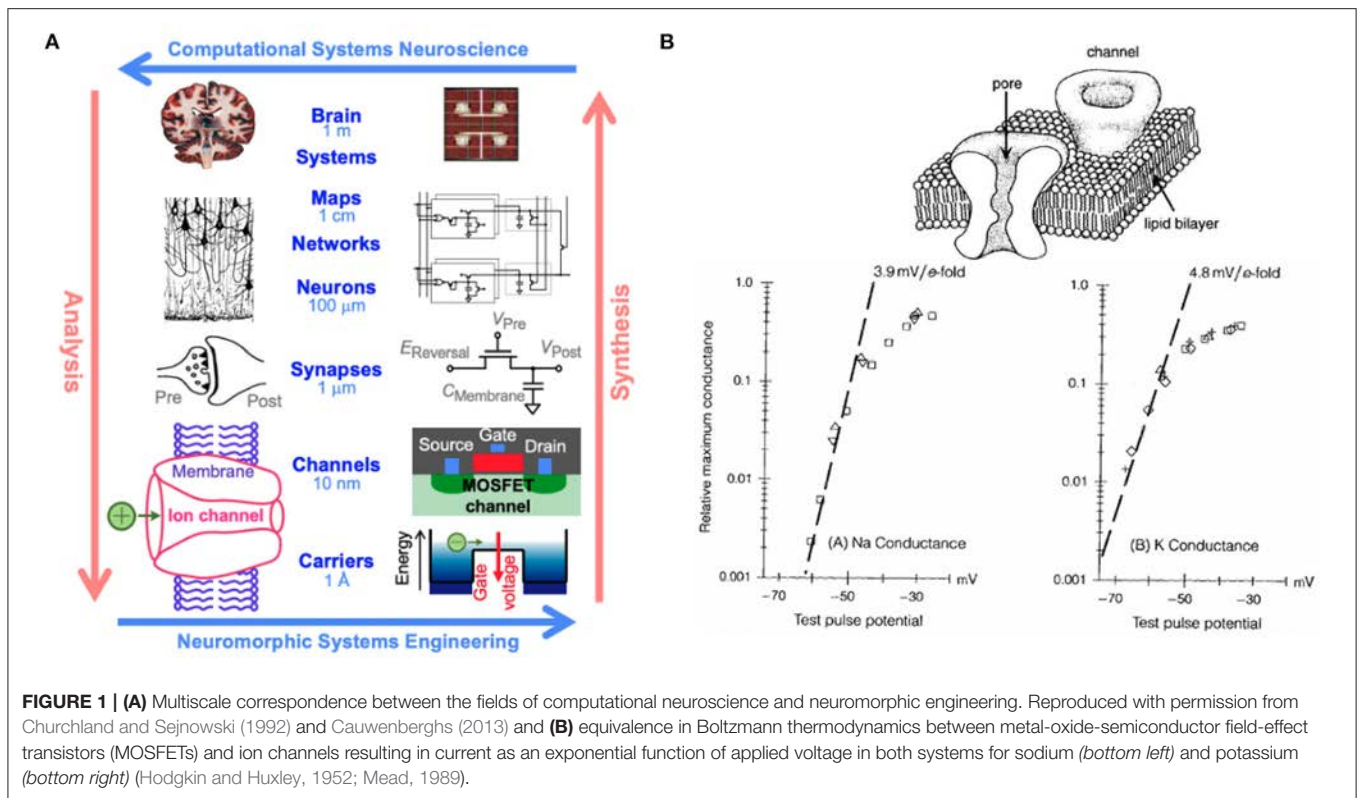
Since the introduction of neuromorphic engineering, computational models of different complexity have been introduced to describe neuronal dynamics, typically ranging from more detailed and realistic conductance-based Hodgkin-Huxley models to simpler integrate-and-fire models allowing for better scalability. Synaptic connectivity between neurons is of primary concern in the field currently because synaptic strength and plasticity are fundamental to learning and memory in both biological and artificial representations of neural networks (Indiveri et al., 2011; Broccard et al., 2017). In neuromorphic architectures, synapses instantiate both computation and memory, and a new focus on compact electronic implementations of this computational memory has been emerging recently including the use of memristors (Boybat et al., 2018). Efficient representation of synapses is a crucial topic of concern as there are roughly 10^4 synapses for each neuron, totalling approximately 10^{16} in the human brain. They are diverse in nature and have highly complex temporal and spatial dynamics, which further complicates their representations (Broccard et al., 2017). Currently, there is a push for efficient synaptic models while maintaining the intricate dynamical behavior exhibited biophysically. Current models include time-multiplexing synapses, analog bistable synapses, and binary synapses to name a few, but the need for scalable and dynamically complex models of synaptic function and transmission is still existent and critical (Bartolozzi and Indiveri, 2007; Broccard et al., 2017).

Modeling synapses is a challenging task due to their intricacy and sheer quantity. As noted above, there are an estimated 10^{16} synapses in the human brain. They vary in function and type, including both chemical and electrical synapses and exhibit behavior spanning multiple different temporal and spatial scales, as well as being highly stochastic in nature (van Rossum et al., 2003; Wang et al., 2020). Additionally, synaptic plasticity causes changes in synaptic strength over time associated with learning

and memory. Synaptic transmission involves a multitude of mechanisms and molecular components, making simulations including all components not readily scalable. In order to capture the sophisticated dynamics of synapses in a scalable manner, abstractions have to be made according to the research problem in question. The stochastic nature of synapses also makes large scale simulations more complicated as modeling stochastic processes is typically more computationally demanding. It has been shown in multiple instances that the noise present due to the stochastic variability in synapses is highly integral to synaptic transmission, so this becomes an important feature to maintain (Malagon et al., 2016). For example, Moezzi et al. (2014) proved that models including ion channel noise in calcium channels paired with the existence of a presynaptic mechanism causing random delays in synaptic vesicle availability best capture the interspike interval behavior of auditory nerve fiber models. Additionally, multiple experimental works have found the existence of presynaptic vesicles that are released into the synaptic cleft with some probability (Castillo and Katz, 1954; Korn and Faber, 1991). There are multiple similar conclusions found in modeling and experimental results as recently discussed by McDonnell et al. (2016).

Synapses form the connections between neurons and the strength of these connections changes over time, forming the basis of learning and memory in both biological and artificial neural networks. The computations involved in accurately modeling the biophysics of synapses are complex due to the highly nonlinear nature of their dynamics, yet most of the neural network models in use today abstract synaptic strength to a single or small number of scalar values, tuned to a specific task. The learning rule for updating synaptic strength is then typically applied using abstractions of synaptic plasticity such as spike-time dependent plasticity and its causal extensions for scalable real-time hardware implementation (Pedroni et al., 2019). Physical constraints and limitations in VLSI implementations restrict the functional form of synaptic representation. In turn, these abstractions restrict the potential computing power of neuromorphic systems and restrain achievable benchmarks in approaching the functional flexibility, resilience, and efficiency of neural computation in the biological brain. Our work addresses the need for a more biophysically realistic model of the synapse with biologically tunable parameters to represent synaptic dynamics while offering a path toward efficient real-time implementation in neuromorphic hardware.

Synaptic transmission is dictated by a series of events initiated by presynaptic stimulation in the form of action potentials. An action potential causes membrane depolarization which leads to stochastic opening and closing of voltage-dependent calcium channels (VDCCs) lying on the presynaptic membrane and a resulting influx of calcium to the presynaptic terminal. Neurotransmitter release is modulated by calcium binding to calcium sensors near the neurotransmitter filled vesicles at the active zone, but calcium has other fates as it diffuses from the VDCCs. In addition to binding to the calcium sensors, it can bind to calbindin, which acts as a buffer, or it can be removed by plasma membrane calcium ATPase (PMCA) pumps. If enough calcium is able to bind to the calcium sensors, though,



then neurotransmitters are released across the synaptic cleft and initiate downstream effects at the postsynaptic membrane (Bartol et al., 2015). This process of synaptic transmission is the basis of communication in the brain.

Abstracting this for computational efficiency, we created a series of Markov state transitions to realize the system with multiple internal states allowing for a biophysically tunable model of synaptic connectivity implementable in neuromorphic architectures. Markov models have a history of use as a stochastic discrete state alternative to Hodgkin-Huxley type formulations since their introduction (Hodgkin and Huxley, 1952; Armstrong, 1971; Colquhoun, 1973). Additional stochastic models have been introduced, including the Gillespie method (1977), which has been used to model neural channel noise (Gillespie, 1977; Skaugen and Walloe, 1979; Chow and White, 1996). Markov models have also found use in whole-cell models (Winslow et al., 1999). Further extensions utilize a particle model (Koch, 1999). The importance of the inclusion of stochasticity in ion channel behavior and synaptic transmission generally cannot be understated. Its inclusion has been demonstrated time and time again in experimental work and is thought to be integral in the form and function of synaptic transmission (McDonnell et al., 2016). This provides an additional complication in modeling synapses and has been handled at various different stages of transmission, including the stochastic models of vesicle release using probabilistically generated quantal components, stochastic models of transmitter diffusion, and stochastic models of receptors (Castillo and Katz, 1954; van Rossum et al., 2003; Bartol et al., 2015). These simulations are computationally expensive

due to the high transition rates paired with the small number of transitions necessitating a small timepoint. Specifically, Markov models have shown to be an effective method of modeling ion channels but require high computational cost to effectively do so.

This paper looks to abstract the computationally complex and nonlinear nature of synaptic transmission dynamics in a manner that is efficient and readily scalable for implementation in neuromorphic silicon very large-scale integrated (VLSI) circuits. This is done by introducing an efficient stochastic sampling scheme within a Markov chain representation of the components integral to stochastic presynaptic quantal transmission.

2. MATERIALS AND METHODS

2.1. Markov Chain Models

The cascade of events from the action potential stimulus input to the presynaptic neurotransmitter release output can be equivalently modeled as a Markov chain to realize the system with multiple internal states instead of directly tracking all molecules and their kinetics in a computationally complex spatiotemporal 3D reaction-diffusion model. Each internal Markov state is assumed to be dependent solely on the state at the previous timepoint and is conditionally independent of all previous timepoints, simplifying simulations. Therefore, the fully biophysically complex system of synaptic transmission can be abstracted and sampled to create a Markov Chain Monte Carlo (MCMC) simulation which answers the same

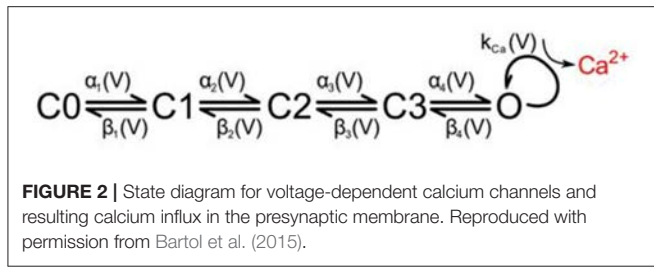


FIGURE 2 | State diagram for voltage-dependent calcium channels and resulting calcium influx in the presynaptic membrane. Reproduced with permission from Bartol et al. (2015).

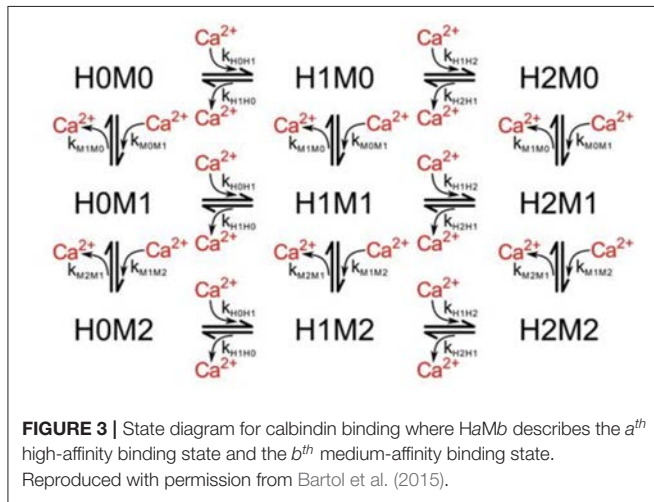


FIGURE 3 | State diagram for calbindin binding where HaMb describes the a^{th} high-affinity binding state and the b^{th} medium-affinity binding state. Reproduced with permission from Bartol et al. (2015).

question of neurotransmitter release utilizing tunable biophysical parameters while providing scalability for implementation in neuromorphic architectures.

For any given stimulus input, the VDCCs are assigned transition probabilities between states based on a five-state kinetic model (Figure 2) found experimentally and validated computationally with four closed states and a single open state (Church and Stanley, 1996; Bischofberger et al., 2002; Bartol et al., 2015). Prior to the stimulus, all VDCCs begin in the initial closed state, C0, and the concentration of calcium in the presynaptic terminal is at steady-state. The transition probabilities are voltage dependent akin to a Hodgkin-Huxley model where $\alpha_i(V) = \alpha_{io} \exp\left(\frac{V_m}{V_i}\right)$ and similarly $\beta_i(V) = \beta_{io} \exp\left(\frac{V_m}{V_i}\right)$ with parameter values from Bischofberger et al. (2002). The number of open VDCCs at any given moment is used to determine the number of calcium entering the presynaptic terminal based on experimental I-V curves and the resulting I-V equation found in Bischofberger et al. (2002) and used in Bartol et al. (2015), which gives the value for k_{Ca} . Calcium influx is captured by including transitions from the final closed VDCC state, C3, to the open VDCC state and an internal calcium generation. Using this, influx of calcium is modeled over the entire stimulus input due to the VDCCs opening.

Once calcium has entered the presynaptic terminal, much of it binds to calbindin, which acts as a buffer and primarily modulates the amount of calcium that is able to reach the calcium sensors at the active zone. The state transitions are reversible first-order

reactions, thus transition probabilities are dependent on the free calcium in the system and updated as that amount changes over time. Calbindin has four binding sites, two of high affinity and two of medium affinity, leading to a nine-state calcium concentration-dependent kinetic model (Figure 3; Nagerl et al., 2000). By modeling the binding and unbinding of calcium to calbindin as a loss or gain of free calcium, respectively, calcium transients can also be elucidated.

Our Markov chain is a discrete-state chain in discrete time. Markov chains are modeled by a probability that the chain will move to another state given its current state and is conditionally independent of all previous timesteps. The probabilities are by nature only dependent on the current state of the Markov chain. The probability of the state of a molecule X can typically be predicted for a certain timepoint $t + \Delta t$ as some particular state x_j using the states at all previous timepoints from the start of the simulation, $t = 0$, to the timepoint just before that in question, t . For a Markov chain simulation solely dependent on the previous timepoint, it is possible to predict the probability that a molecule is in a given state, x_j at the timepoint $t + \Delta t$ using solely the state of the single timepoint just before, X_t , which is known to be a particular state x_i . Thus, the probability of the molecule being in state x_j given that at the previous timepoint it was in state x_i is given as P_{ij} . Succinctly, this is written as

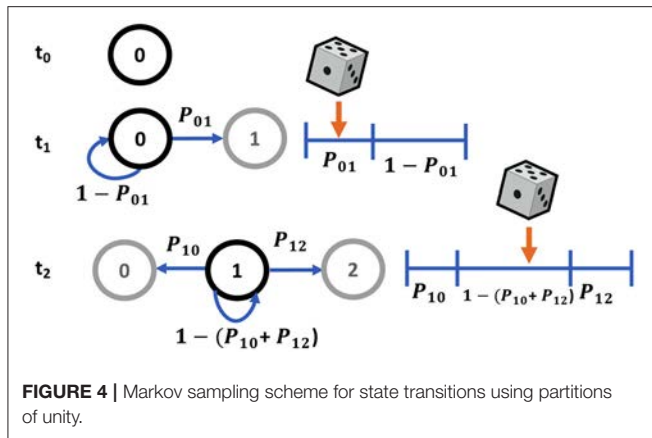
$$P(X_{t+\Delta t} = x_j | X_t = x_i, X_{t-\Delta t}, \dots, X_0) = P(X_{t+\Delta t} = x_j | X_t = x_i) = P_{ij} \tag{1}$$

For state transitions, the probability of transitioning to an adjacent state is the transition rate inherent in the system (k_{ij} for the transition from state i to state j , and k_{ij} is not necessarily equal to k_{ji}) times the timepoint, Δt . In the case of calbindin transitions, this is further multiplied by the amount of free unbound calcium for forward reactions as it is a first-order reaction. For the VDCCs, the transition rates are the α , β , and k_{Ca} . The probability that a molecule stays in its current state is the sum of the probabilities it transitions to an adjacent state subtracted from unity. For a multi-state system, this gives a transition probability matrix for the likelihood of transition from a given state at the current timepoint to any other state at the next timepoint. This matrix is sparse, with nonzero probabilities only for adjacent states to which a transition is possible. In the case of the five-state VDCC system, this gives the probability of a transition from state i to state j as

$$P_{ij} = \begin{cases} k_{ij} \Delta t & j = i \pm 1 \\ 1 - \sum_{k=0}^{N_{adj}} P_{ik} & j = i, k \neq i \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

where transitions to adjacent states are given by the transition rate k_{ij} times the timepoint, Δt ; the probability of staying in the current state is the sum of probabilities of adjacent state transitions subtracted from unity, where N_{adj} is the number of possible adjacent states. The probability of transitioning to a non-adjacent state is set to zero.

Typically Markov state transitions are modeled via a discrete inverse transform method, where given a random variable X , the



transition probabilities P_{ij} describe a partition of unity (Figure 4). Therefore, we can generate a random number uniformly, $R \sim U(0, 1)$ and map it onto discrete values of X . For example, in a two state system, $X_j = 0$ if $R \leq P_{i0}$ or $X_j = 1$ if $P_{i0} < R \leq P_{i0} + P_{i1} = 1$. This involves searching the state space for the next state given the current state for each molecule in the system at each timepoint, which can be a slow process for systems with a large number of states and molecules.

Here we have implemented a more efficient MCMC sampling strategy involving sampling from a multinomial distribution. Therefore, instead of sampling from a uniform distribution for each of n molecules, we sample from a multinomial distribution once for each state, using n molecules as the number of experiments, where $X \sim \text{Multi}(n, p_1, \dots, p_k)$. For simulations where the number of possible states is less than the number of molecules, this is a more efficient sampling strategy. Since we are particularly interested in the number of molecules in each state at each timepoint, this is an effective approach. Multinomial sampling thus describes the distribution of the n experiments across k possible outcomes each with a probability of p_k , where n_k is the number of experiments falling into the k^{th} outcome following a probability mass function of

$$f(n_1, \dots, n_k; n, p_1, \dots, p_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} \quad (3)$$

In our model, for each state i , we have an initial number of molecules in that state at a given timepoint t , or $n_{i,t}$. As previously described, there exists a probability that the molecules will transition to any state at the next timepoint, including staying in the original state given by P_{ij} . Thus, to determine the distribution of molecules $n_{i,t}$ across all states at the next timepoint, we sample from a multinomial distribution according to

$$X_{i,t+\Delta t} \sim \text{Multi}(n_{i,t}, P_{i1}, \dots, P_{ik}) \quad (4)$$

for k possible states. We do this sampling for each state at each timepoint and sum accordingly. This expedites computation by only requiring a single computation at each timepoint, sampling the distribution of all n molecules at once. **Algorithm 1** highlights the pseudocode for this process.

Algorithm 1: Markov Multinomial Reaction Sampling.

Result: The number of molecules in each state at each timepoint for a simulation.

initialize number of states;

initialize matrix of number of molecules per state per timepoint;

for each timepoint do

for each state do

 calculate transition probability according to Equation

 (1);

 sample multinomial distribution with current

 number of molecules in the state and transition

 probability;

end

 update number of molecules in each state using samples;

end

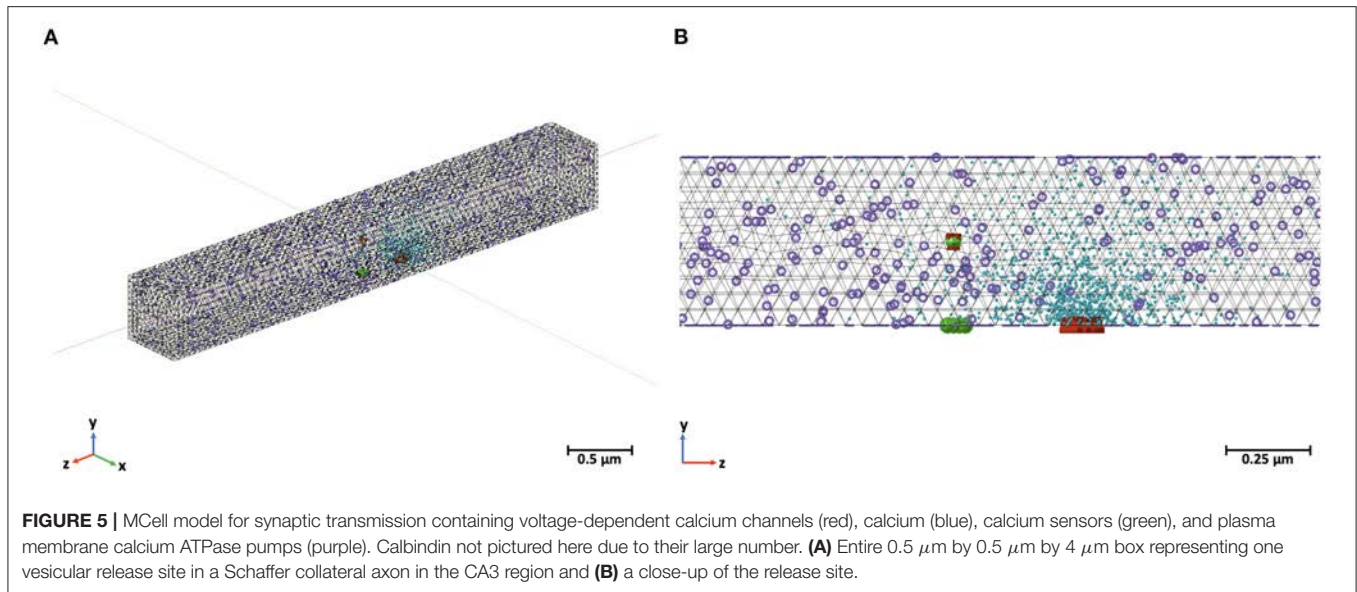
Markov simulations for the VDCCs were run for 65 VDCCs all starting in the closed state, C0. Calbindin molecules were initiated in the different binding states according to the steady-state concentration of calcium and at a baseline concentration of $4.5 \cdot 10^{-5}$ M. All simulations were run for 10 ms with a timestep of $1 \mu\text{s}$. The simulations were repeated 1,000 times to obtain an average and standard deviation. Markov simulations were implemented using Python.

2.2. MCell Models

MCell is a modeling software that uses spatially realistic 3D geometries and Monte Carlo reaction-diffusion modeling algorithms, which allows for biophysically realistic simulations of high complexity as it specifically tracks the state of every molecule in space and time (Bartol et al., 2015). Due to the accuracy and specificity, it provides a ground truth for biological simulations but does so at the cost of computational complexity.

To validate and compare our Markov models of synaptic transmission, we built a biophysically realistic stochastic 3D reaction-diffusion system with all major components for presynaptic vesicle release variability in response to a stimulus input (Figure 5) based on the models of Nadkarni et al. (2010) and Bartol et al. (2015). The model includes realistic geometry for a CA3-CA1 *en passant* synapse focusing primarily on the presynaptic Schaffer collateral axon of a CA3 pyramidal cell found in the hippocampus with parameters set from experimental data (Nadkarni et al., 2010; Bartol et al., 2015). The CA3-CA1 synapse was chosen for the simulations as it is highly studied experimentally and is important for learning and memory. Furthermore, CA3-CA1 synapses are relatively small, containing one to two neurotransmitter release zones. Release from this region is also known to be highly stochastic in nature, necessitating the inclusion of stochasticity in biologically realistic models (Nadkarni et al., 2010). All kinetics and parameters match those used for the equivalent Markov models.

The MCell model includes the canonical presynaptic geometry for an average CA3-CA1 synaptic terminal as a rectangular box



measuring $0.5\mu\text{m}$ by $0.5\mu\text{m}$ by $4\mu\text{m}$. This box captures the dynamics of a single synaptic active zone, referring to the region on the presynaptic membrane specialized for neurotransmitter release. Initially, the terminal contains the calbindin buffer, steady-state calcium concentration, PMCA pumps, VDCCs and calcium sensors modulating neurotransmitter release (Nadkarni et al., 2010). The detailed diffusion dynamics and kinetics of these systems are based on experimental data and have been discussed in further detail in Bartol et al. (2015). The active zone is based on that of an average presynaptic active zone containing seven docked neurotransmitter vesicle release sites. The VDCCs, of type P/Q, are stationed at a biophysically realistic distance from the active zone. They transition states in response to the membrane depolarization. The location, number, and calcium conductance of the VDCCs is replicated from experimental data (Nadkarni et al., 2010). PMCAs are homogeneously placed across the presynaptic membrane while calbindin molecules are in a uniform concentration within the volume. This is a flexible architecture that can respond to any stimulus input and allows for monitoring of the states of each molecule in the system. The MCell CA3-CA1 synaptic transmission models were originally created and validated in Nadkarni et al. (2010) and Bartol et al. (2015). To compare with the Markov models, we used the same single action potential stimulus.

MCell models were also run 1,000 times for 10 ms with a timestep of 1 μs .

3. RESULTS

3.1. Voltage-Dependent Calcium Channels

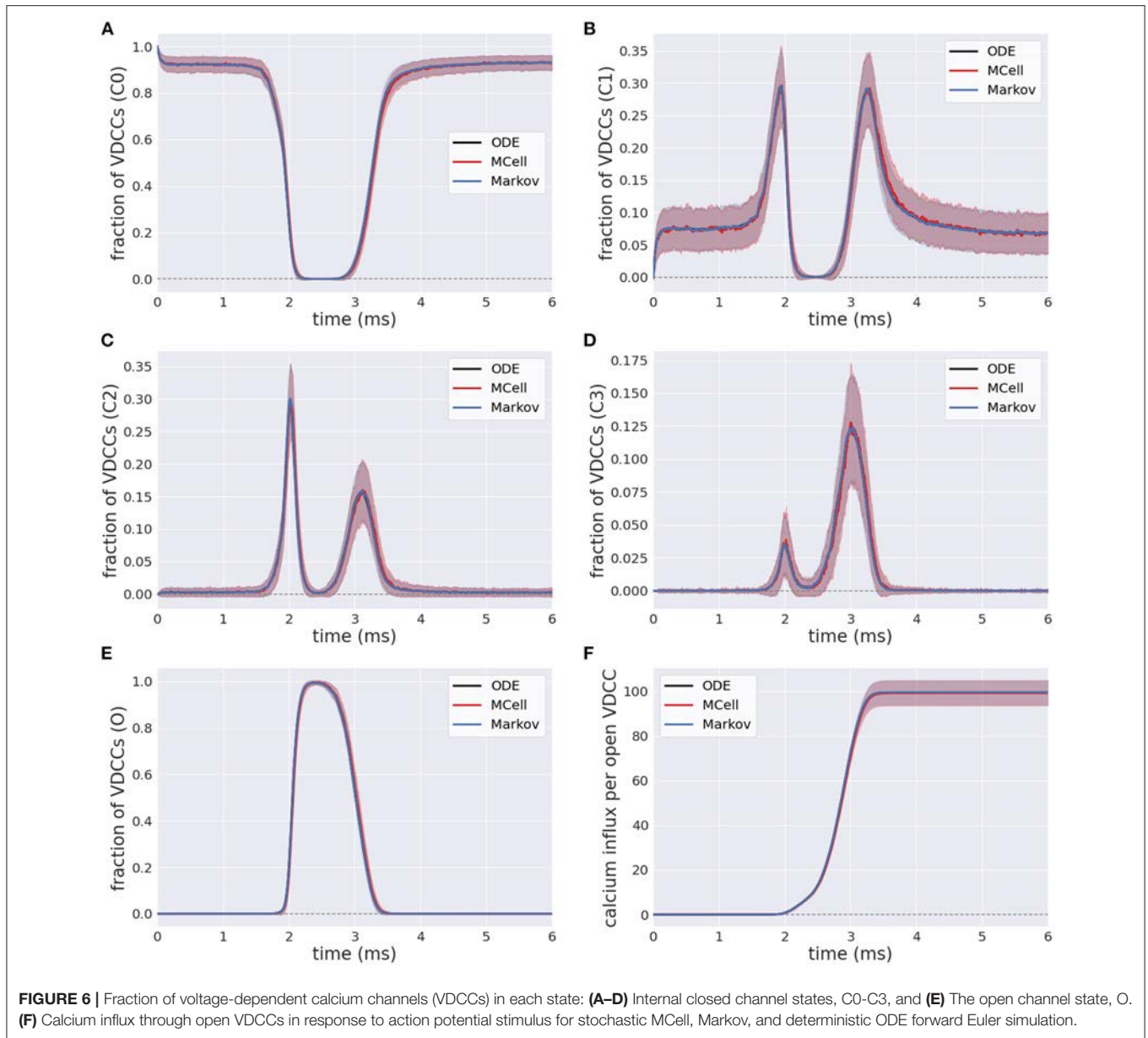
The efficient Markov chain implementation has strong agreement with the full MCell model in terms of the internal state transients in response to an external stimulus. The number of closed VDCCs (state C0) decreases over the duration of the stimulus (Figure 6A). The internal states (C1-C3) subsequently increase

and decrease as the membrane voltage increases and the forward rates for the VDCCs increase (Figures 6B–D), leading to an exponential increase in the open VDCCs while the membrane depolarizes. Figure 6E shows the fraction of open VDCCs over time in response to the action potential, which controls the amount of calcium influx to the system. At the maximum membrane potential, almost all VDCCs are in the open state. As the membrane repolarizes, the reverse reaction rate constants increase, and the VDCCs close. This leads to another increase and decrease in the internal VDCC states as the receptors go from their open to resting closed state (C0).

In its open state, VDCCs allow for the probabilistic influx of calcium through the channels into the presynaptic bouton. This is exemplified in Figure 6F, where there is an increase in the calcium influx through the open VDCCs over the course of the stimulus. Again, there is strong agreement between the more computationally complex MCell model and the computationally efficient Markov equivalent model.

3.2. Calbindin Buffer

Simulations of homogeneous calcium and calbindin were run using the MCell, Markov and deterministic simulation schemes. In the presence of calcium, the forward binding reaction is heavily favored, and this is highlighted in Figure 7A where free calcium exponentially decreases. A similar transient is apparent for the unbound state of calbindin, as it quickly transitions to different stages of high and medium binding Figure 7B. Over the course of the simulation, all the free calcium is removed from the system, and calbindin states reach a new steady-state where there is still unbound calbindin. Similarly, the fully bound state, H2M2, rapidly increases and reaches a new steady state that is still only 1% of all calbindin Figure 7C. This is due to the high concentration of calbindin in the presynaptic bouton. Even once all the calcium is in a bound state, there is still plenty of unbound or partially bound calbindin remaining in the

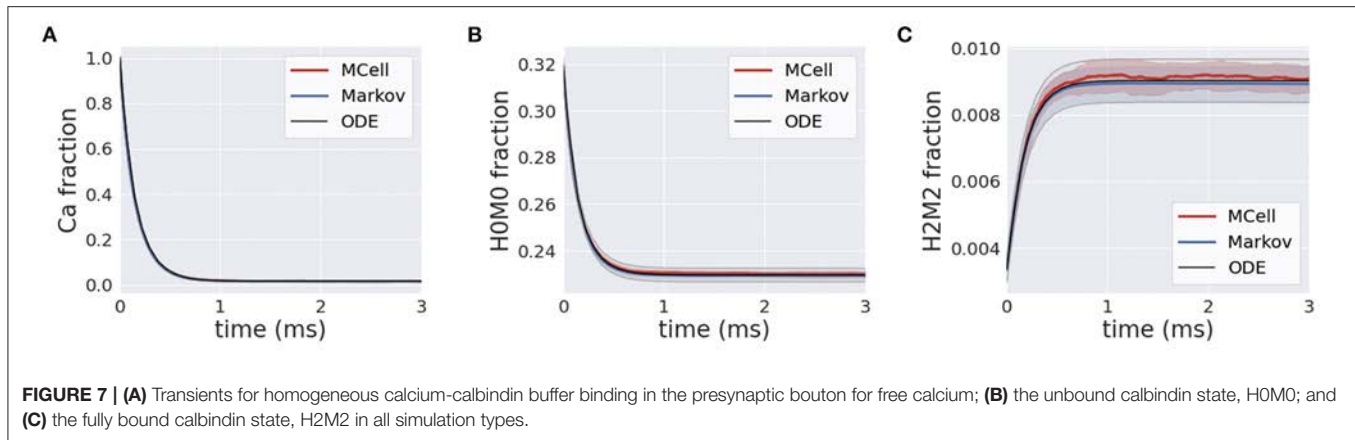


system. Calbindin acts as a strong buffer allowing for calcium storage and asynchronous neurotransmitter release, so this and slow unbinding rates become an important feature of calbindin. The rapid extent to which calcium binds to calbindin shows the impact of buffering on calcium's ability to diffuse and bind to the calcium sensors in the active zone. The inclusion of calbindin at such high concentrations becomes a key feature of maintaining the relatively low release rates of neurotransmitters even in the presence of a stimulus.

3.3. Complexity Analysis

MCell uses a scheduler which allows for only making changes to the scheduled particles, though in the worst-case, this still scales with the total number of particles in the simulation,

n , where n_{VDCC} is 65 and n_{calb} is 2.7×10^4 . It also scales with the length of the simulation, t , described by the number of time points for a discrete simulation. The simulations for the VDCC and calbindin both use 10k timepoints. At each timepoint, a particle can transition to any of its adjacent or branched states, b , which is similarly described by a fan-out factor in electronic implementation. From the VDCC kinetic model described in **Figure 2**, b_{VDCC} is 1–2 depending on the state while the calbindin kinetic model in **Figure 3** gives b_{calb} of 2–4. The overall time complexity for MCell is $\mathcal{O}(bnt)$. The classical Markov representation tracks every particle. It also searches through the space of each adjacent state for potential state transitions at each time point. Therefore, classical Markov implementation similarly results in an $\mathcal{O}(bnt)$ time complexity, or $\mathcal{O}(nt \log_2 b)$ at best



for implementation with an efficient search algorithm. Both the multinomial Markov model and the Euler ODE implementation describe the system in terms of the number of molecules in each state leading to a dependence on the total number of states, $s \geq b$, rather than the total number of particles. The total number of states for VDCC is 5 (Figure 2) while the number of states for calbindin is 9 (Figure 3). Due to efficient sampling methods, the multinomial Markov method is independent of the number of adjacent states, leading to a time complexity of $\mathcal{O}(bst)$ for both the multinomial Markov and Euler ODE methods. Thus, our stochastic multinomial Markov model is equally amenable to large scale simulations as the deterministic ODE method that is typically used in simulations involving more synapses.

The traditional Markov sampling model and the MCell representation store the molecular states in bits for each particle as well as the states adjacent to the current state, leading to a space complexity of $\mathcal{O}(bn \log_2 s)$. The efficient Markov model and ODE solution both simply store the number of molecules represented by bits in each state at each timepoint as well as the branched states resulting in a space complexity of $\mathcal{O}(bs \log_2 n)$. There exists a trade off here between the number of particles in each state compared to the number of states where one is stored directly and one is stored as an index. Thus, for simulations where the number of states is less than the number of particles, the multinomial Markov model is an efficient representation of the system, which is typically the case for biochemical simulation. MCell is more efficient with large state-space systems, but the number of states could be sparsified in a multinomial Markov representation by implementing dynamic instantiation and annihilation of states. Additionally, unseen or rarely seen states could be ignored by truncating based on probability of a particle being in that state. This would functionally decrease the number of states in the system allowing for use of the multinomial Markov simulation method.

3.4. Benchmarks

Runtime and total floating point operations were used as metrics for comparison between the simulation methods (Table 1). We also looked at the number of pseudorandom number generator calls (n_{PRNG}) between the simulations as this provides

a metric to elucidate the differences observed in execution time between the simulations. Here we compare MCell, the standard Markov model, and the multinomial Markov stochastic models. The deterministic Euler solution is included as well for a non-stochastic comparison. Again, it is valuable to note the importance of stochasticity in these models. Significant work has shown the necessity of stochasticity in models of synaptic transmission in order to match experimental work. It has been demonstrated that deterministic models at this scale generally underestimate quantal release as concentration fluctuations are not captured (van Rossum et al., 2003; Bartol et al., 2015; McDonnell et al., 2016). Thus, while deterministic ODE models provide efficient simulation techniques, they are not able to capture the full complexity of the dynamics of synaptic transmission, hence motivating the need for an efficient stochastic model.

In the VDCC simulations, the multinomial sampling MCMC model has a runtime on the order of the forward Euler deterministic solution. The MCell and the standard Markov stochastic models exemplify a runtime an order of magnitude higher. The number of operations is also higher for the MCell and the standard Markov models compared to the multinomial Markov model. The standard Markov case generates a pseudorandom number for each molecule and each timestep, so n_{PRNG} is equivalent to the number of molecules multiplied by the number of timesteps, $n_{VDCC}t$. In the multinomial Markov simulation, a pseudorandom number is generated for each occupied state and possible branching points at each timepoint, which gives $(bs)_{VDCC}t$ in the worst-case scenario. Therefore, n_{PRNG} is smaller for the multinomial case as long as $(bs)_{VDCC} < n_{VDCC}$, which is always the case here.

For the calbindin model, the multinomial Markov method is again an order of magnitude faster than the MCell model although it is also an order of magnitude slower than the deterministic model. The standard Markov model is an order of magnitude slower than the multinomial model. The operations are also fewer for the multinomial case than the standard case. Again, the standard Markov case gives n_{PRNG} equal to $n_{calb}t$ while the multinomial Markov simulation is $(bs)_{calb}t$. Again we see a smaller n_{PRNG} in the multinomial case because

TABLE 1 | Benchmarks for different simulation types for both voltage-dependent calcium channel and calbindin binding simulations.

VDCC	Runtime (sec)	No. Pseudorandom number generator calls (n_{PRNG})	Total floating-point operations (FLOPs)
MCell	139.20	1.5×10^9	1.7×10^{11}
Standard markov	109.16	6.5×10^5	7.0×10^7
Multinomial markov	9.40	1.0×10^5	1.3×10^7
ODE (Euler)	10.05	0	7.7×10^6
Calbindin	Runtime	No. Pseudorandom number generator calls (n_{PRNG})	Total floating-point operations (FLOPs)
MCell	29.34	6.9×10^7	8.7×10^9
Standard Markov	76.78	2.7×10^8	1.4×10^{10}
Multinomial Markov	2.37	3.6×10^5	2.7×10^7
ODE (Euler)	0.72	0	1.6×10^6

$(bs)_{\text{calb}} < n_{\text{calb}}$ even in the worst-case scenario where b is at its maximum value. Simulations are not currently optimized on hardware suggesting opportunities for further decreases in runtime. Overall, the multinomial Markov simulation provides a computationally efficient alternative to stochastic MCell simulations while maintaining the biological accuracy.

3.5. Neuromorphic Implementation

Thermodynamic foundations of neuromorphic engineering suggest direct biophysical implementation of populations of ion channels with individual stochastic opening and closing of gating variables driven by thermal noise fluctuations (Mead, 1989). So it seems only natural to consider implementations using stochastic ODEs describing the rates of reaction kinetics under additive white Gaussian noise (AWGN):

$$\frac{dX_{i,t}}{dt} = \sum_j k_{ij} X_{j,t} + \xi_{i,t} \quad (5)$$

where $\xi_{i,t}$ is normally distributed with zero mean and variance dependent on the magnitude of $X_{i,t}$. Fully parallel, continuous-time analog implementation of reaction kinetic rate equations of the type (Equation 5) have been demonstrated in micropower integrated circuits, e.g., cytomorphic chips in BiCMOS integrated silicon technology (Woo et al., 2018). Abundant intrinsic noise present in these micropower cytomorphic circuits can serve as AWGN, although its magnitude is determined by thermal processes that are hard to control and other non-white Gaussian sources of intrinsic noise contribute strongly colored low-frequency spectra. Thus, discrete-time implementation of the ODEs (Equation 5) through Euler integration on a digital computer offers greater control over the shape and amplitude of the AWGN distribution, limited by the quality of pseudo-random number generation by deterministic algorithms.

Although purely digital algorithmic implementations go against foundational principles of neuromorphic engineering rooted in the physics of computation (Mead, 1989), the convenience of their programmability and reproducibility have made ODE-based digital emulation platforms such as Loihi a popular choice among more software-focused neuromorphic computer scientists (Davies et al., 2018). The computation involved in such discrete-time ODEs (Equation 5) can be

performed at varying degrees of parallelism in custom or reconfigurable digital hardware, with the variables X_i being updated in sequence through time-multiplexing a single processing core in one extreme case, or all X_i updated in parallel with dedicated processing elements for each in the other extreme case. Ultimately in practice, the energy efficiency is relatively independent of the compute implementation, and depends more critically on the available memory bandwidth in accessing the rate parameters defining network connectivity (Pedroni et al., 2020). In essence, discrete-time Euler-integration ODE implementation of Equation (5) amounts to sampling from a normal distribution

$$X_{i,t+\Delta t} \sim \mathcal{N}(n_{i,t+\Delta t}, \sigma_{i,t}) \quad (6)$$

with mean and standard deviation

$$n_{i,t+\Delta t} = \sum_j P_{ij} n_{j,t} \quad (7)$$

$$\sigma_{i,t} = \sqrt{n_{i,t}(n - n_{i,t})} \quad (8)$$

incurring computational complexity $\mathcal{O}(bst)$ (section 3.3).

More fundamentally, the main disadvantage of implementing stochastic ODEs (Equation 5) or their discrete-time digital versions (Equation 6) is that they are primarily based on the Central Limit Theorem for very large number of variables, $n \rightarrow \infty$. As such, they have limited accuracy in approximating the reaction kinetics in systems with smaller numbers of molecular variables. While one may be tempted to assume that molecules are always excessively abundant, this is not typically the case since reactions are rate limited by the least abundant of reagents. Low numbers in molecular dynamics are prevalent in biologically relevant settings, giving rise to significant amounts of biological noise that are critical in neural dynamics, e.g., the highly stochastic quantal release of neurotransmitter in synaptic transmission. Thus, there is need for a mathematical description of stochastic synaptic transmission dynamics able to capture the accuracy in simulations with relatively small numbers of variables. Here we have shown that our multinomial Markov alternative, which directly samples the variables from the multinomial distribution (Equation 4) rather than the limiting normal distribution (Equation 6), produces accurate results for any value of n while offering

TABLE 2 | Space and time complexity for the various simulation strategies.

Complexity	Time complexity	Space complexity
MCell	$O(b n t)$	$O(b n \log_2 s)$
Standard markov	$O(b n t)$	$O(b n \log_2 s)$
Multinomial markov	$O(b s t)$	$O(b s \log_2 n)$
ODE (Euler)	$O(b s t)$	$O(b s \log_2 n)$

n (total number of molecules), s (total number of states), b (number of branching states), t (number of timepoints).

nearly identical implementation complexity $O(bst)$ (section 3.3). Hence we see the Markov chain abstractions of reaction kinetics not only as a means to approach biophysical realism in modeling molecular cellular dynamics without molecular-scale representation, but also as a means toward efficient neuromorphic hardware without biophysical compromise. The key point is that the computational complexity of implementing our multinomial Markov model is essentially identical to that of stochastic ODEs (see **Table 2**), whether in software executing serially on a von Neumann programmable digital computer or in massively parallel digital or analog hardware. Hence, the neuromorphic circuit designer tasked to implement brain-inspired models of information processing faces an easy choice: more bio-realistic models that account for detailed stochasticity in reaction kinetics incur the same resource utilization and energy costs, and use similar design principles, as their stochastic ODE approximations.

In addition to field-programmable gate array (FPGA) reconfigurable (Pedroni et al., 2020) or custom-integrated neuromorphic programmable (Davies et al., 2018) instantiations in digital hardware, we envision physically neuromorphic instantiations in micropower analog continuous-time compute-in-memory hardware that obviate sampling from posterior distributions and directly implement Markov state transitions through parallel implementation of sum-product rules with self-normalizing probabilities (Chakrabarty and Cauwenberghs, 2004, 2005), at throughput density and energy efficiency that are orders of magnitude higher than today's most advanced general-programmable computational platforms.

4. DISCUSSION

The goal of this work was to create a more computationally efficient model of biologically realistic synaptic transmission for use in large-scale neuromorphic systems. We created a multinomial MCMC sampling strategy for capturing the internal states of vital molecules in the system in response to stimulus where transition probabilities could be voltage- or concentration-dependent, and the next timestep could be predicted solely using the current timestep. This scheme was implemented to capture the dynamics of the stochastic opening and closing of VDCCs through multiple internal states as well as the resulting calcium influx into the presynaptic bouton through the open VDCCs. Once calcium has entered the presynaptic terminal, we also simulated calcium binding to the calbindin buffer which

modulates calcium levels in the bouton, directly impacting the amount of calcium that reaches the calcium sensors in the active zone. This amount impacts the neurotransmitter release from the presynaptic side and the resulting effects on the postsynaptic side.

All simulations were modeled using the multinomial Markov sampling method as well as a typical Markov sampling method and compared to highly detailed 3D geometric stochastic reaction-diffusion simulations done using MCell. The Markov simulations show agreement with the MCell simulations for the system dynamics including the number of open VDCCs and calcium influx in response to an action potential stimulus as well as the binding of calcium to the calbindin buffer. Differences are observed from the deterministic solution to the stochastic simulations implying the importance of stochasticity in these simulations to capture more biologically-realistic systems.

Exemplified by runtime and total number of operations, the multinomial MCMC method of simulations was shown to be more efficient than the standard Markov model while also being faster than the MCell equivalents. This is hopeful for scaling these biologically-realistic models to large-scale systems while maintaining biological tunability.

Next steps involve modeling the remaining kinetics in a similar fashion including the binding and removal of calcium by the plasma membrane calcium (PMCA) pumps as well as binding to the calcium sensors. In addition, to capture the diffusion of calcium through the presynaptic terminal but specifically to the calcium sensors at the active zone, a diffusive kernel must be included to the system. Upon inclusion of these elements, the entire process from stimulus to neurotransmitter release can be captured as a series of Markov chains leading to powerful implications for synaptic transmission modeling. The whole synapse can be included as well with the inclusion of a diffusive kernel across the synaptic cleft as well as downstream effects on the postsynaptic side, of which many mirror similar kinetics and dynamics as the presynaptic side leading to a natural extension of this modeling framework. The resulting system would be a biologically tunable model of synaptic transmission for any stimulus input in a highly efficient manner. This opens the door for large-scale implementations of synaptic transmission and learning readily implementable into neuromorphic architectures with strong biological realism.

Through the utilization of Markov-based abstractions applied to biophysically realistic 3D reaction-diffusion models of a chemical synapse, we have created a compact and efficient internal state space representation of synaptic transmission. This is in response to the challenge presented by the high dimensionality and complex nature of molecular-scale interactions in synapses and across scales making implementation in very large-scale systems previously unattainable. The model is directly amenable to efficient emulation in parallel neuromorphic hardware systems while maintaining biophysically relevant and interpretable parameters that are readily tunable. This opens the door toward neuromorphic circuits and systems on very large scale that strike a greater balance between integration density and biophysical accuracy in modeling neural function at the whole-brain level.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/margotwagner/markov-mcell-data>.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

- Armstrong, C. M. (1971). Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axons. *J. Gen. Physiol.* 58, 413–437. doi: 10.1085/jgp.58.4.413
- Bartol, T. M., Keller, D. X., Kinney, J. P., Bajaj, C. L., Harris, K. M., Sejnowski, T. J., et al. (2015). Computational reconstitution of spine calcium transients from individual proteins. *Front. Synaptic Neurosci.* 7:17. doi: 10.3389/fnsyn.2015.00017
- Bartolozzi, C., and Indiveri, G. (2007). Synaptic dynamics in analog VLSI. *Neural Comput.* 19, 2581–2603. doi: 10.1162/neco.2007.19.10.2581
- Bischofberger, J., Geiger, J. R. P., and Jonas, P. (2002). Timing and efficacy of Ca^{2+} channel activation in hippocampal mossy fiber boutons. *J. Neurosci.* 22:0593–602. doi: 10.1523/JNEUROSCI.22-24-105.93.2002
- Boybat, I., Gallo, M. L., Nandakumar, S. R., Moraitis, T., Parnell, T., Tuma, T., et al. (2018). Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* 9:2514. doi: 10.1038/s41467-018-04933-y
- Broccard, F. D., Joshi, S., Wang, J., and Cauwenberghs, G. (2017). Neuromorphic neural interfaces: from neurophysiological inspiration to biohybrid coupling with nervous systems. *J. Neural Eng.* 14, 041002. doi: 10.1088/1741-2552/aa67a9
- Castillo, J., and Katz, B. (1954). Quantal components of the end-plate potential. *J. Physiol.* 124, 560–573. doi: 10.1113/jphysiol.1954.sp005129
- Cauwenberghs, G. (2013). Reverse engineering the cognitive brain. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15512–15513. doi: 10.1073/pnas.1313114110
- Chakrabartty, S., and Cauwenberghs, G. (2004). “Margin normalization and propagation in analog VLSI,” in *2004 IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. 1 (Vancouver, BC: IEEE), 1–901. doi: 10.1109/ISCAS.2004.1328341
- Chakrabartty, S., and Cauwenberghs, G. (2005). “Sub-microwatt analog VLSI support vector machine for pattern classification and sequence estimation,” in *Advances in Neural Information Processing Systems*, Vol. 17, eds L. Saul, Y. Weiss, and L. Bottou (Cambridge, MA: MIT Press).
- Chow, C. C., and White, J. A. (1996). Spontaneous action potentials due to channel fluctuation. *Biophys. J.* 71, 3013–3021. doi: 10.1016/S0006-3495(96)79494-8
- Church, P., and Stanley, E. (1996). Single L-type calcium channel conductance with physiological levels of calcium in chick ciliary ganglion neurons. *J. Physiol.* 496 (Pt 1):59–68. doi: 10.1113/jphysiol.1996.sp021665
- Churchland, P. S., and Sejnowski, T. J. (1992). The Computational Brain. *The MIT Press*. doi: 10.7551/mitpress/2010.001.0001
- Colquhoun, D. (1973). *The Relation Between Classical and Cooperative Models for Drug Action. Drug Receptors*. (London: Macmillan Press).
- Davies, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., et al. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* 38, 82–99. doi: 10.1109/MM.2018.112130359
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008

FUNDING

This work was supported by a NSF Graduate Research Fellowship to MW, the National Science Foundation, and the Office of Naval Research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2021.698635/full#supplementary-material>

- Hodgkin, A., and Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764
- Indiveri, G., Linares-Barranco, B., Hamilton, T. J., van Schaik, A., Etienne-Cummings, R., Delbruck, T., et al. (2011). Neuromorphic silicon neuron circuits. *Front. Neurosci.* 5:73. doi: 10.3389/fnins.2011.00073
- Koch, C. (1999). *Biophysics of Computation*. (New York: Oxford University Press).
- Korn, H., and Faber, D. (1991). Quantal analysis and synaptic efficacy in the CNS. *Trends Neurosci.* 14, 439–445. doi: 10.1016/0166-2236(91)90042-S
- Malagon, G., Miki, T., Llano, I., Neher, E., and Marty, A. (2016). Counting vesicular release events reveals binomial release statistics at single glutamatergic synapses. *J. Neurosci.* 36, 4010–4025. doi: 10.1523/JNEUROSCI.4352-15.2016
- McDonnell, M. D., Goldwyn, J. H., and Lindner, B. (2016). Neuronal stochastic variability: Influences on spiking dynamics and network activity. *Front. Comput. Neurosci.* 10:38. doi: 10.3389/fncom.2016.00038
- Mead, C. (1989). *Analog VLSI and Neural Systems*. (Boston, MA: Addison-Wesley).
- Mead, C. (1990). Neuromorphic electronic systems. *Proc. IEEE* 78, 1629–1636. doi: 10.1109/5.58356
- Moezzi, B., Iannella, N., and McDonnell, M. D. (2014). Modeling the influence of short term depression in vesicle release and stochastic calcium channel gating on auditory nerve spontaneous firing statistics. *Front. Comput. Neurosci.* 8:163. doi: 10.3389/fncom.2014.00163
- Nadkarni, S., Bartol, T. M., Stevens, C. F., Sejnowski, T. J., and Levine, H. (2010). Modelling vesicular release at hippocampal synapses. *PLoS Comput. Biol.* 6:e1000983. doi: 10.1371/journal.pcbi.1000983
- Nagerl, U. B., Novo, D., Mody, I., and Vergara, J. L. (2000). Binding kinetics of calbindin-D28k determined by flash photolysis of caged Ca^{2+} . *Biophys. J.* 79, 3009–3018. doi: 10.1016/S0006-3495(00)76537-4
- Pedroni, B. U., Deiss, S. R., Mysore, N., and Cauwenberghs, G. (2020). “Design principles of large-scale neuromorphic systems centered on high bandwidth memory,” in *2020 International Conference on Rebooting Computing (ICRC)*, (Atlanta, GA) 90–94.
- Pedroni, B. U., Joshi, S., Deiss, S. R., Sheik, S., Detorakis, G., Paul, S., et al. (2019). Memory-efficient synaptic connectivity for spike-timing-dependent plasticity. *Front. Neurosci.* 13:357. doi: 10.3389/fnins.2019.00357
- Sejnowski, T. J. (2020). *The Deep Learning Revolution*. (Cambridge, MA: The MIT Press).
- Skaugen, E., and Walloe, L. (1979). Firing behavior in a stochastic nerve membrane model based upon the Hodgkin-Huxley equations. *Acta Physiol. Scand.* 107, 343–363. doi: 10.1111/j.1748-1716.1979.tb06486.x
- van Rossum, M. C., O’Brien, B., and Smith, R. (2003). Effects of noise on the spike timing precision of retinal ganglion cells. *Am. Physiol. Soc.* 89, 2406–2419. doi: 10.1152/jn.01106.2002
- von Neumann, J. (1958). *The Computer and the Brain*. Yale University Press.
- Wang, J., Cauwenberghs, G., and Broccard, F. D. (2020). Neuromorphic dynamical synapses with reconfigurable voltage-gated kinetics.

- IEEE Trans. Biomed. Eng.* 67, 1831–1840. doi: 10.1109/TBME.2019.2948809
- Winslow, R. L., Rice, J., Jafri, S., Marbán, E., and O'Rourke, B. (1999). Mechanisms of altered excitation-contraction coupling in canine tachycardia-induced heart failure, ii. *Circ. Res.* 84, 571–586. doi: 10.1161/01.RES.84.5.571
- Woo, S. S., Kim, J., and Sarpeshkar, R. (2018). A digitally programmable cytomorphic chip for simulation of arbitrary biochemical reaction networks. *IEEE Trans. Biomed. Circ. Syst.* 12, 360–378. doi: 10.1109/TBCAS.2017.2781253

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wagner, Bartol, Sejnowski and Cauwenberghs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.