EPJ Nonlinear Biomedical Physics
a SpringerOpen Journal

Reduction of calcium release site models via optimized state aggregation

Yan Hao

Correspondence: hao@hws.edu
Math and Computer Science
Department, Hobart and William
Smith Colleges, 300 Pulteney St.,
Geneva, NY 14456, USA

Abstract

Background: Markov chain models of calcium release sites in living cells exhibit stochastic dynamics reminiscent of the experimentally observed phenomenon of calcium puffs and sparks. Such models often take the form of stochastic automata networks in which the transition probabilities for each of a large number of intercellular channel models depend on the local calcium concentration and thus the state of nearby channels. The state-space size in such compositionally defined calcium release site models increases exponentially as the number of channels increases, which is referred to as "state-space explosion".

Methods: In order to overcome the state-space explosion problem, we utilized the idea of "coarse graining" and implemented an automated procedure that reduces the state space by aggregating and lumping states of the full release site model. For a given state aggregation scheme, the transition rates between reduced states are chosen consistent with the conditional probability distribution among states within each group. A genetic algorithm-based approach is then applied to select the state aggregation schemes that lead to reduced models that approximate the observable behaviors of the full model.

Results: The genetic algorithm-based approach is implemented in Matlab® and applied to two different release site models. The approach found reduced models that approximate the full model in the number of open channels, spark statistics, and the jump probability matrix as a function of time.

Conclusions: A novel automated genetic algorithm-based searching technique is implemented to find reduced calcium release site models that approximate observable behaviors of the full Markov chain models that possess intractable state-spaces. As compared to the full model, the reduced models produce quantitatively similar results using significantly less computational resources.

Keywords: State space explosion, Genetic algorithms, Calcium signaling, Stochastic automata network, Set partition, Coarse graining strategies

Background

Coarse-graining methods as a model reduction strategy

Increasingly over the years, mathematical models and computer simulation have been used in the natural sciences and the social sciences. As more and more scientific discoveries from experiments are included, models and simulation approaches have been developed to be more and more accurate. Unfortunately, the corresponding computational cost also gets higher. When a problem to be modeled is large in scale or possesses multiple

scales, the complexity and dimensionality of the model may increase to unmanageable levels of storage and computational capacity. How to reduce the computational cost without losing important properties of mathematical/computational models is therefore an important scientific question. Molecular dynamics (MD) simulation for example, is a type of N-body simulation that allows atoms and molecules to interact for a fixed period of time [1]. While the trajectories of atoms and molecules can be quite accurately determined by numerically solving Newton's equations of motion for a system of interacting particles, the computations are highly expensive [2]. To reduce the computational cost for MD simulations, the field of coarse-grained modeling and simulation has been rapidly expanding. Instead of explicitly representing every atom of a system, one partitions the system into a number of groups of atoms, and then uses "pseudo-atoms" to represent each group. The coarse-grained modeling significantly reduces the time and computational storage requirement of MD simulations. Simulations of soft matter systems, polymer dynamics, protein folding and many other physical or biological systems that feature the spatiotemporal coupling between scales thus become possible [3, 4].

As another example, continuous time Markov chains have become an important modeling approach over the past a few decades. Because it well describes the systems that undergoes transitions from one state to another, it was used extensively in modeling ion channels in cell membranes [5, 6]. However, when the collaborative behavior of a cluster of these ion channels are of interest, the state space of the cluster model increases exponentially as the number of channels increases. This combinatorial state-space explosion causes some modeling approaches to become intractable. In this paper, we present a novel approach to reduce large Markov chain models that shares the spirit of coarse-graining methods: the state-space of the Markov chain model is partitioned into groups, then each group is represented by a "mega-state" so that the full model is compressed into a tractable size. The partition scheme is selected by a genetic algorithm-based approach to preserve selected features of the full model.

Markov chain models of Ca^{2+} signaling

As a second messenger, calcium ions (Ca^{2+}) play an important role in many physiological activities. Signaling occurs when the cell is stimulated to release calcium ions (Ca^{2+}) from the endoplasmic/sarcoplasmic reticulum (ER/SR), the intracellular Ca^{2+} reservoir, or when Ca^{2+} enters the cell through plasma membrane ion channels [7]. The intracellular Ca^{2+} release which causes localized Ca^{2+} elevations known as puffs and sparks arises from concerted gating of clusters of inositol 1,4,5-trisphosphate receptors (IP3Rs) or ryanodine receptors (RyRs) on the surface of ER/SR [8–10]. For example, in cardiac myocyte excitation-contraction coupling (ECC), the cell membrane depolarizes causing L-type Ca^{2+} channels to open and the Ca^{2+} influx further activates RyRs located on the SR, known as Ca^{2+}-induced Ca^{2+} release (CICR) [11].

The spatial organization of IP3Rs and RyRs has been shown to be the basis of intracellular Ca^{2+} signaling activities that are observed via confocal microfluorimetry in cardiomyocytes, oocytes, and other cell types [10, 12, 13]. The spacing of receptor clusters was identified as a decisive parameter for the occurrence of collective behaviors [14, 15].

From the literature, the behavior of single IP3R/RyR channel gating is often modeled by continuous-time discrete-state Markov chains (CTMCs) [16, 17]. When Markov chain models of these channels are coupled via a Ca^{2+} microdomain in which the transition

rates between the states of each channel become dependent on the states of other channels, the simulated Ca^{2+} channel clusters (release sites) may exhibit stochastic excitability that is reminiscent of Ca^{2+} puff/sparks [6, 14, 18]. However, the number of states possessed by these compositionally defined Ca^{2+} release site models increases exponentially as the number of channels increases. This combinatorial state-space explosion causes some modeling approaches to become intractable.

While the dynamics of any individual Ca^{2+} release site can theoretically be obtained by Monte Carlo simulation regardless of model complexity, in practice these simulations are prohibitively computationally intensive due to the large state spaces. Moreover, because cells usually possess a large number of release sites, compositionally defined Ca^{2+} release site models have often been excluded from multiscale whole cell simulations. On the other hand, many recently developed approaches that accelerate whole cell simulations, such as probability density and moment closure approaches [19, 20], require release site models to be as compact as possible while retaining the physiological realism of collective channel gating.

For these reasons, we developed several automated approaches based on fast/slow analysis [21]. To reduce Markov chain Ca^{2+} release site models where the rate constants in release site models are categorized as either fast or slow, groups of states that are connected by fast transitions are lumped so that the full model is compressed into a tractable size while the physiological gating and interaction properties of the channels are preserved. However, when the time scale separation between transition rates that is necessary for fast/slow analysis is absent, the manner in which the full model states should be partitioned and aggregated for optimal reduction is difficult to determine a priori. Naively enumerating all partitions for a Markov chain Ca^{2+} release site model and choosing the one with the smallest error is not possible because the number of valid partitions is too large. Actually, finding out how many ways one can divide a graph that possesses n vertices into k smaller components is know to be an NP-hard problem [22]. For example, a release site model composed of merely five three-state Ca^{2+} channels (15 states) can be partitioned in approximately 10^{10} distinct ways. In this paper we discuss the implementation of a genetic algorithm that is able to automatically and rapidly select partition schemes that reduce the corresponding Markov chain model to a tractable size while keeping the reduction error in control.

Genetic algorithms

Developed in the 1970s by John Holland [23], genetic algorithms are widely used as computational schemes to find exact or approximate solutions for optimization and search problems. Genetic algorithms have been applied to various aspects of biological research, such as the profiling of gene expression in bacteria [24, 25] and phylogenetic analysis of proteins [26]. More recently, it has been used to identify parameters for cell-specific electrophysiology models [27]. Nevertheless, the application of genetic algorithms in the context of the automated reduction of Ca^{2+} release site models is novel. In our implementation, each individual in a population corresponds to a potential scheme for state aggregation. The program "evolves" the population by selecting the partitions that lead to reduced models that approximate the full model behavior.

Unlike the fast/slow analysis that assumes fixed ER/SR [Ca^{2+}] and instantaneous coupling between the channels [21], we motivate a whole cell homeostasis formulation that

takes both local and global Ca^{2+} signaling into consideration. Consequently the reduced models selected by the genetic algorithm must generate small errors for a wide range of ER/SR $[Ca^{2+}]$. In the implementation of the genetic algorithm, a population of set partitions is randomly generated, where each partition corresponds to a potential scheme for state aggregation. The states of the full model is aggregated and reduced following each individual set partition scheme. Then the difference between the full model and the corresponding reduced model on the behaviors of interest, spark statistics for example, are measured and called the "error" of the reduced model. The error is used as the key factor for deciding the fitness of each individual partition scheme. Since partition schemes that produce low errors are preferred, the fitness is chosen to be a decreasing function of error. The program selects survivors from the current population in a manner that favors better fitness (i.e. low error), then each survivor will generate one child (a new set partition scheme) by mutation. The set partition population thus evolves toward the partitions that have better fitness and lead to reduced models that better approximate the full model.

The remainder of this paper is organized as follows. In the 'Model' section we motivate the network reduction process by partitioning a minimal whole cell model of Ca^{2+} homeostasis where both localized (subcellular) and global (cellular) aspects of Ca^{2+} signaling are modeled. In the 'Methods' section we introduce genetic algorithms and detail their implementation in the context of reducing a minimal whole cell model of Ca^{2+} homeostasis that features bidirectional local and global Ca^{2+} signaling. In the 'Results' section, we demonstrate that the reduced model approximates the full model with regard to several important steady-state responses observed in the minimal whole cell environment. To show that the reduction technique is applicable to more realistic Ca^{2+} release site models, we also present Ca^{2+} release site reduction results using a single channel model that includes both cytosolic and luminal Ca^{2+} regulation.
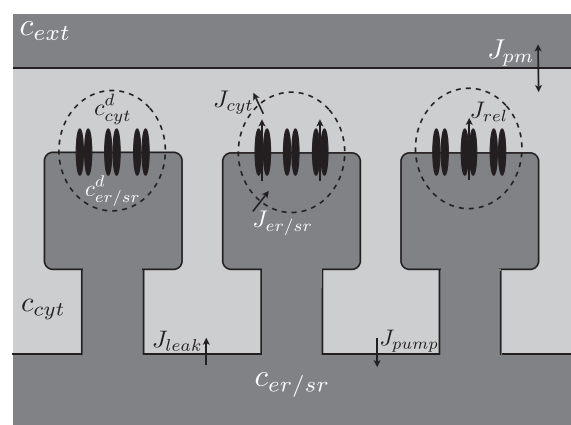
Fig. 1 Diagram of model components and fluxes. The bulk endoplasmic/sarcoplasmic reticulum $[Ca^{2+}]$ is represented by $c_{er/sr}$, the bulk cytosolic and external $[Ca^{2+}]$ is c_{cyt} and c_{ext} respectively. Ca^{2+} channels are located on the ER/SR membrane forming release sites. The domain $[Ca^{2+}]$ (c_{cyt}^d and $c_{er/sr}^d$) are rapidly changed by the release currents (J_{rel}) when the number of open channels changes. Other fluxes considered in this model are: diffusion from cytosolic domain to the bulk cytosol (J_{cyt}), diffusion from the bulk ER/SR to the luminal side domains ($J_{er/sr}$), a passive leak from the ER/SR to the cytosol (J_{leak}), the SERCA pump flux that re-sequesters Ca^{2+} in to the ER/SR (J_{pump}) and fluxes across the plasma membrane (J_{pm})

Model

A minimal whole cell model

We will demonstrate and validate our Ca^{2+} release site model reduction approach using a whole cell model of a quiescent cytosolic environment that takes Ca^{2+} homeostasis into account (Fig. 1). Similar to previous work by Hartman and colleagues [28], this minimal whole cell model considers both local and global Ca^{2+} responses to the stochastic gating of Ca^{2+} channels. Further, release and reuptake fluxes are balanced in this model. Figure 1 shows the components and fluxes of the model. A large number of Ca^{2+} release sites are coupled to the bulk cytosolic and ER/SR $[Ca^{2+}]$. Each Ca^{2+} release site is composed of 10–30 Ca^{2+} channels. In this formulation, release sites may experience different "domain" $[Ca^{2+}]$, but all channels in a given release site experience the same local cytosolic and luminal $[Ca^{2+}]$. Consistent with prior work by Hinch and colleagues [29–31], when the number of open channels in a Ca^{2+} release site changes, the local $[Ca^{2+}]$ is assumed to rapidly reach a new equilibrium in the spatially restricted domain. The change in the balance of the leak and reuptake by the endo(sarco)plasmic reticulum Ca^{2+}-ATPase (SERCA) pumps caused by this change in the domain $[Ca^{2+}]$ will influence the bulk $[Ca^{2+}]$ and further affect the puff/spark dynamics.

Steady-state of domain concentration

Figure 1 demonstrates the fluxes in this whole cell formulation. The domain $[Ca^{2+}]$ for the release sites, c_{cyt}^{d} and $c_{er/sr}^{d}$, are coupled to each other via the release flux J_{rel} when one or more channels open. As mentioned above, the domain $[Ca^{2+}]$ associated with each release site is distinct, and all domains are coupled to the bulk cytosolic and luminal compartments via the fluxes J_{cyt} and $J_{er/sr}$. Under these assumptions, the domain fluxes are given by:

$$J_{rel}^{n} = v_{rel}\gamma_{n}\left(c_{er/sr}^{d,n} - c_{cyt}^{d,n}\right) \tag{1}$$

$$J_{cyt}^{n} = v_{cyt}\left(c_{cyt}^{d,n} - c_{cyt}\right) \tag{2}$$

$$J_{er/sr}^{n} = v_{er/sr}\left(c_{er/sr} - c_{er/sr}^{d,n}\right) \tag{3}$$

where v_{rel} is the maximum release rate through a release site, c_{cyt} and $c_{er/sr}$ are the bulk cytosolic and ER/SR concentrations, and $\gamma_{n} = n/N$ is the fraction of open channels at an N-channel per release site. The rate constants v_{cyt} and $v_{er/sr}$ determine the time required for the decay and refilling of the cytosolic and luminal microdomains, respectively [32, 33].

Because the dynamics of domain Ca^{2+} is fast compared to the stochastic gating of Ca^{2+} channels, the domain fluxes associated with each release site must balance for any specific release site:

$$J_{rel}^{n} = J_{cyt}^{n} = J_{er/sr}^{n}. \tag{4}$$

The domain $[Ca^{2+}]$ of any release site with n channels open can be obtained by directly solving Eq. 4 as a function of the bulk cytosolic and luminal $[Ca^{2+}]$ (c_{cyt} and $c_{er/sr}$), that is,

$$c_{cyt}^{d,n} = \frac{v_{cyt}}{v_{cyt} + \tilde{v}_{er/sr}}c_{cyt} + \frac{\tilde{v}_{er/sr}}{v_{cyt} + \tilde{v}_{er/sr}}c_{er/sr} \tag{5}$$

$$c_{er/sr}^{d,n} = \frac{\tilde{v}_{cyt}}{\tilde{v}_{cyt} + v_{er/sr}}c_{cyt} + \frac{v_{er/sr}}{\tilde{v}_{cyt} + v_{er/sr}}c_{er/sr} \tag{6}$$

where

$$\tilde{v}_{cyt} = \frac{\gamma_n v_{rel} v_{cyt}}{\gamma_n v_{rel} + v_{cyt}}, \quad \text{and} \quad \tilde{v}_{er/sr} = \frac{\gamma_n v_{rel} v_{er/sr}}{\gamma_n v_{rel} + v_{er/sr}}. \tag{7}$$

Notice that for a release site with N channels, the number of open channels takes integer values from 0 to N. Consequently, there are $N + 1$ pairs of cytosolic and luminal domain $[Ca^{2+}]$ values for any given value of the bulk concentration (c_{cyt} and $c_{er/sr}$).

Concentration balance equations for the bulk cytosol and ER

As shown in Fig. 1, the bulk cytosolic and luminal $[Ca^{2+}]$ are both influenced by the Ca^{2+} fluxes to and from their associated microdomains, J_{cyt}^n and J_{er}^n. The bulk concentrations also interact via a SERCA pump flux that takes the form:

$$J_{pump} = \frac{v_{pump} c_{cyt}^2}{v_{pump}^2 + c_{cyt}^2} \tag{8}$$

and a passive leak from the ER/SR to the cytosol of the form,

$$J_{leak} = v_{leak}(c_{er/sr} - c_{cyt}). \tag{9}$$

Following previous work [28] by Hartman and colleagues, our model formulation assumes a permeabilized cell, and the plasma membrane flux J_{pm} is

$$J_{pm} = k_{pm}(c_{ext} - c_{cyt}) \tag{10}$$

where k_{pm} is chosen large enough so that the bulk cytosolic $[Ca^{2+}]$ is "clamped" to the extracellular bath ($c_{ext} = 0.1\mu$ M).

Now that all Ca^{2+} fluxes are defined, the concentration balance equations for the bulk cytosolic and ER compartments are given by:

$$\frac{dc_{cyt}}{dt} = J_{cyt}^T + J_{leak} - J_{pump} + J_{pm} \tag{11}$$

$$\frac{dc_{er/sr}}{dt} = \frac{1}{\lambda_{er/sr}} \left(J_{er/sr}^T - J_{leak} + J_{pump} \right), \tag{12}$$

where $\lambda_{er/sr} = V_{er/sr}/V_{cyt}$, V_{cyt} and $V_{er/sr}$ are the effective cytosolic and ER/SR volumes, i.e. taking Ca^{2+} buffering into account. J_{cyt}^T and $J_{er/sr}^T$ are the sums of fluxes over all release sites. Notice that under the fast domain Ca^{2+} assumption, there are only $N + 1$ pairs of possible domain $[Ca^{2+}]$ values and consequently J_{cyt}^T and $J_{er/sr}^T$ can be expressed as

$$J_{cyt}^T = \sum_{n=0}^{N} f_n v_{cyt}^T \left(c_{cyt}^{d,n} - c_{cyt} \right) \tag{13}$$

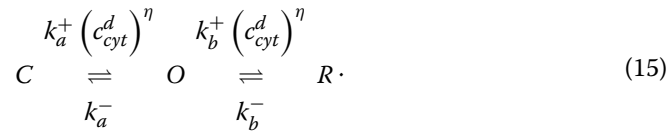$$J_{er/sr}^T = \sum_{n=0}^{N} f_n v_{er/sr}^T \left(c_{er/sr} - c_{er/sr}^{d,n} \right), \tag{14}$$

where f_n is the fraction of release sites with n open channels.

The Markov chain model of single channel gating

The stochastic gating of single channels is studied by a continuous-time discrete-state Markov chain model. This single channel model has three states, C(closed), O(open)

and R(refractory), featuring both Ca^{2+} activation and Ca^{2+} inactivation. The transition diagram of this model is given by

$$C \quad \underset{k_a^-}{\overset{k_a^+ \left(c_{cyt}^d\right)^\eta}{\rightleftharpoons}} \quad O \quad \underset{k_b^-}{\overset{k_b^+ \left(c_{cyt}^d\right)^\eta}{\rightleftharpoons}} \quad R \cdot \tag{15}$$

In this transition diagram $k_i^+ \left(c_{cyt}^d\right)^\eta$ and k_i^-, where $i \in \{a, b\}$, are transition rates with units of reciprocal time. k_i^+ is an association rate constant with units of $conc^{-\eta}$ $time^{-1}$ where η is the cooperativity of Ca^{2+} binding, and c_{cyt}^d is the domain $[Ca^{2+}]$ experienced by the release site on the cytosol side. Under the assumption that the formation and collapse of local Ca^{2+} is fast compared to channel gating, when the local Ca^{2+} concentrations are specified, the transition-state diagram Eq. 15 defines a continuous time Markov chain with the corresponding infinitesimal generator matrix $Q = (q_{ij})$ given by:

$$Q = \begin{bmatrix} \diamond & k_a^+ \left(c_{cyt}^d\right)^\eta & 0 \\ k_a^- & \diamond & k_b^+ \left(c_{cyt}^d\right)^\eta \\ 0 & k_b^- & \diamond \end{bmatrix} \tag{16}$$

The off-diagonal entries of the Q-matrix for this irreducible and time-homogeneous Markov chain are transition rates, or hazards, from state i to state j, defined by

$$q_{ij} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr[\, S(t + \Delta t) = j | S(t) = i\,], \tag{17}$$

where $i \neq j$ and the diamonds ($\diamond$) on the diagonal entries are negative values leading to row sums of zero.

All of the statistical properties of the Ca^{2+} channel can be calculated from its Q-matrix (Eq. 16). Importantly, the time evolution of the probability distribution over all three states of this model can be calculated by solving the ordinary differential equation (ODE) system:

$$\frac{d\pi}{dt} = \pi Q, \tag{18}$$

where $\pi(t) = (\pi_C, \pi_O, \pi_R)$ is a row vector indicating the probability of finding the channel in each state at time t, given the initial condition $\pi(0)$. Notice that the limiting probability distribution π_s of Markov chains (the steady state of Eq. 18) does not depend on the initial condition $\pi(0)$, and can be obtained by solving

$$\pi_s Q = 0 \quad \text{subject to} \quad \pi_s e = 1, \tag{19}$$

where e is a commensurate column vector of ones.

Compositionally defined Ca^{2+} release site models

The Ca^{2+} release site models that are used to demonstrate the implementation of the reduction approach involve N identical Ca^{2+} channels interacting via changes in local $[Ca^{2+}]$ under the assumption of "instantaneous mean-field coupling". The local $[Ca^{2+}]$ experienced by the Ca^{2+} regulatory site of each channel is assumed to depend only on the number of open channels at the Ca^{2+} release site N_O. Because identical channels coupled

in this manner are indistinguishable, a release site composed of NM-state channels includes

$$\beta(N, M) = \binom{N + M - 1}{N} = \frac{(N + M - 1)!}{N! (M - 1)!} \tag{20}$$

distinct states. In this three-state single channel model case, the N-channel release site has $\beta(N, 3) = (N+2)(N+1)/2$ states. Each state can be written in the form of an ordered three-tuple (N_C, N_O, N_R), where $N_i = k, (i \in \{C, O, R\})$ indicates k channels are in state i, and $\sum_i N_i = N$. With this notation, the states of any release site form a well ordered set and can be conveniently ranked anti-lexicographically. Figure 2 enumerates all the states and illustrates the topology of the three-state single channel model (Fig. 2a) and a minimal release site composed of two three-state channels under the mean-field coupling assumption (Fig. 2b).

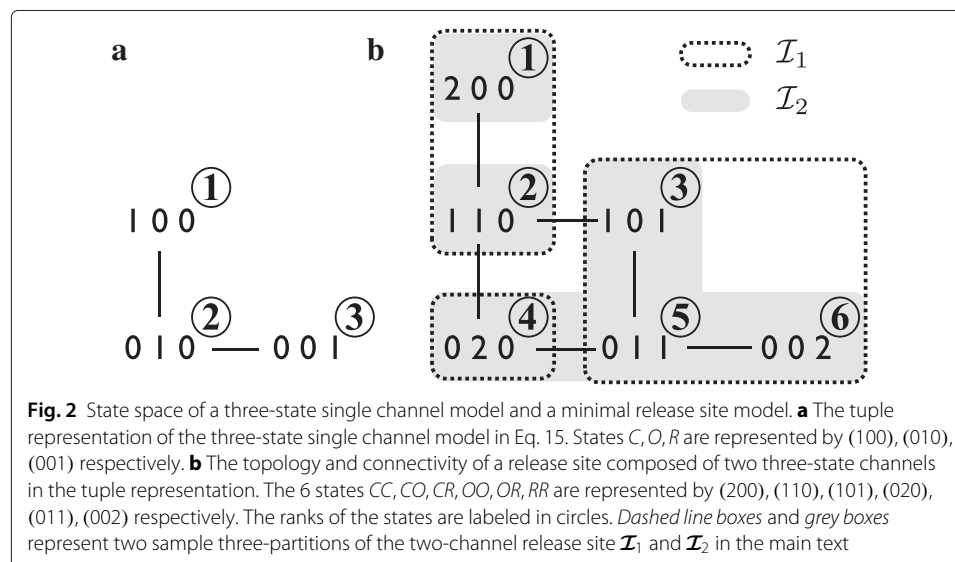## Methods

Reduction technique

Our basic strategy of reducing a Ca^{2+} release site model to a smaller model with pre-determined size $\hat{b}$ includes three major steps:

Step 1 Partition the full model into $\hat{b}$ groups.
Step 2 Lump the states within each group.
Step 3 Calculate proper transition rates between groups.

In previous work [21], *Step* 1 was achieved automatically based on the separation of time scales, where states that are connected by fast transitions were aggregated. In this paper, we employ a genetic algorithm to search for partition schemes that generate small reduction errors for general Ca^{2+} release site models, especially those without time scale difference. In the future, a partition that divides the states of a release site model into $\hat{b}$ groups will be referred to as a "$\hat{b}$-partition". *Step* 2 and *Step* 3 in this reduction technique are carried on from [21].

Fig. 2 State space of a three-state single channel model and a minimal release site model. **a** The tuple representation of the three-state single channel model in Eq. 15. States C, O, R are represented by (100), (010), (001) respectively. **b** The topology and connectivity of a release site composed of two three-state channels in the tuple representation. The 6 states CC, CO, CR, OO, OR, RR are represented by (200), (110), (101), (020), (011), (002) respectively. The ranks of the states are labeled in circles. *Dashed line boxes* and *grey boxes* represent two sample three-partitions of the two-channel release site $\mathcal{I}_1$ and $\mathcal{I}_2$ in the main text
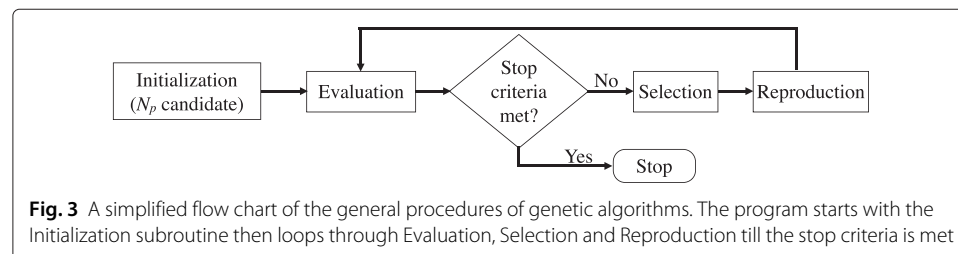
Theoretically, this genetic algorithm-based technique can be used to reduce any Ca^{2+} release site models of size $\beta(N, M)$ to any pre-determined size $\hat{b}$ ($\hat{b} < \beta(N, M)$). Furthermore, the partition scheme can be restricted by certain rules, such as "each group must be connected on the transition diagram" or "all states in each group must present the same number of open channels", etc. Hereafter, we will demonstrate the implementation of the genetic algorithm using a minimal example where the 6-state release site model illustrated in Fig. 2b is partitioned into 3 groups with the constraint "states in each group must be connected". Results from reducing a release site that is composed of 10 three-state channels will be presented in the 'Results' section. Other partition restrictions can be implemented by restricting the reproduction procedure or modifying the objective function.

Genetic algorithms

Genetic algorithms are probabilistic search algorithms that were introduced by John Holland in the 1970s [23]. Based on the mechanics of natural selection, genetic algorithms have been used to find exact or approximate solutions to optimization and search problems whose objective functions are discontinuous, nonlinear, difficult to calculate, etc. [34, 35]. These algorithms manipulate a population of solutions to the objective function and implement a "survival of the fittest" strategy in their search for better solutions. In general the methodology of genetic algorithm s can be displayed in a flowchart as shown in Fig. 3.

The program starts with ***initialization***, where a number of "*individuals*" (solution candidates) are randomly generated to form an initial "*population*". The size N_p of the *population* is usually kept constant throughout the entire searching procedure. After initialization, this *population* goes through the ***evaluation*** procedure, where each *individual* is evaluated by the objective function and its *fitness* is assigned according to the objective function value. The program then checks whether the termination criteria are satisfied; usually either the desired objective function value is attained or a predetermined number of generations is reached. If none of the termination criteria is satisfied, the program will move on to the ***selection*** process, which is usually stochastic and designed so that the *individuals* with better *fitness* have a higher probability to be selected as compared to those who are less fit. Only a fraction of the current population (N_s *individual*s, where $N_s < N_p$) can survive and enter the ***reproduction*** process as the "*parent*" solutions. To generate each "*child*" (a new solution candidate), one or more *parents* are selected, recombined (crossover) and/or varied (mutate). The ***reproduction*** process continues till N_p *individual*s are generated and thus a new generation of *population* is formed. The new generation will then go through an ***evaluation*** process to have its *fitness* evaluated and the entire program continues until one or more of the termination criteria are satisfied.

Fig. 3 A simplified flow chart of the general procedures of genetic algorithms. The program starts with the Initialization subroutine then loops through Evaluation, Selection and Reproduction till the stop criteria is met

Initialization

Our purpose in using genetic algorithm s is to find partition schemes of full Ca^{2+} release site models so that the resulting reduced models better approximate the full models. In this context, each *individual* ($\mathcal{I}_i$) is a set partition scheme which divides the $\beta(N, M)$ states of the full model into $\hat{b}$ groups. To make physical sense, the requirement of the partition process is that each group must be internally connected, that is, there is a path from any state to any other state. The dashed line boxes and grey boxes in Fig. 2 give two samples valid three-partitions that divide the 6-state release site model into 3 groups:

$$\mathcal{I}_1 = (\{1, 2\}, \{3, 5, 6\}, \{4\})$$
$$\mathcal{I}_2 = (\{1\}, \{2\}, \{3, 4, 5, 6\}).$$

In the ***initialization*** process, N_p distinct three-partitions are randomly generated.
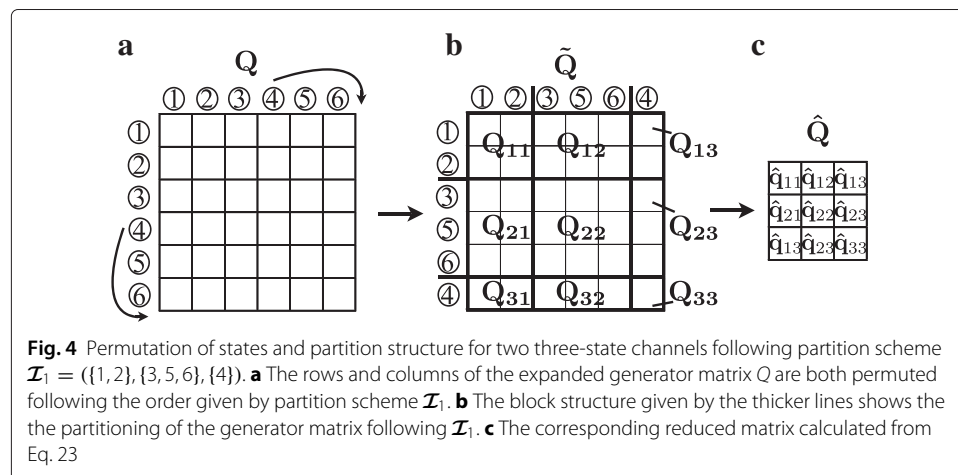
Evaluation

In the ***evaluation*** process each of the N_p *individuals* (partition schemes) must be applied to the full model and have its corresponding reduced model compared to the full model. The *fitness* of each *individual* is then assigned in a manner that favors those that produce less error.

We demonstrate this procedure in more detail with the following individual cited above:

$$\mathcal{I}_1 = (\{1, 2\}, \{3, 5, 6\}, \{4\}).$$

Firstly, the generator matrix associated with the two-channel release site model $\boldsymbol{Q}$ is permuted to reflect the order of $\mathcal{I}$. The permutation for $\mathcal{I}_1$ is shown in Fig. 4a. The new generator matrix $\tilde{\boldsymbol{Q}}$ is then partitioned into a $\hat{b} \times \hat{b}$ block matrix ($\hat{b} = 3$ for $\mathcal{I}_1$) following the scheme given by $\mathcal{I}$ (Fig. 4b). $\tilde{\boldsymbol{\pi}}$, the stationary distribution of $\tilde{\boldsymbol{Q}}$ is conformally partitioned as

$$\tilde{\boldsymbol{\pi}} = \left[\tilde{\boldsymbol{\pi}}_1, \tilde{\boldsymbol{\pi}}_2, \ldots, \tilde{\boldsymbol{\pi}}_{\hat{b}} \right]. \tag{21}$$

Fig. 4 Permutation of states and partition structure for two three-state channels following partition scheme $\mathcal{I}_1 = (\{1, 2\}, \{3, 5, 6\}, \{4\})$. **a** The rows and columns of the expanded generator matrix Q are both permuted following the order given by partition scheme $\mathcal{I}_1$. **b** The block structure given by the thicker lines shows the the partitioning of the generator matrix following $\mathcal{I}_1$. **c** The corresponding reduced matrix calculated from Eq. 23

The generator matrix $\hat{Q}$ of the target reduced Ca^{2+} release site model is a $\hat{b} \times \hat{b}$ matrix

$$\hat{Q} = \begin{bmatrix} \hat{q}_{11} & \hat{q}_{12} & \cdots & \hat{q}_{1\hat{b}} \\ \hat{q}_{21} & \hat{q}_{22} & \cdots & \hat{q}_{2\hat{b}} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{q}_{\hat{b}1} & \hat{q}_{\hat{b}2} & \cdots & \hat{q}_{\hat{b}\hat{b}} \end{bmatrix}, \tag{22}$$

where

$$\hat{q}_{ij} = \bar{\pi}_i \tilde{Q}_{ij} e_j \tag{23}$$

for $i \neq j$, and $\hat{q}_{ii} = \sum_{j \neq i} -\hat{q}_{ij}$. $\bar{\pi}_i$ is the conditional probability distribution of the states within group i:

$$\bar{\pi}_i = \frac{\tilde{\pi}_i}{\tilde{\pi}_i e_i} \tag{24}$$

and e_i are the commensurate column vectors of ones.

When the reduced matrix is generated, the transition probability matrix (jump matrix) of the corresponding reduced model ($\hat{P} = e^{t\hat{Q}}$) is compared to the transition probability matrix of the full model ($P = e^{tQ}$). Assuming the full model has b states, we write

$$\hat{E}(t) = \hat{P}(t) - UP(t)V \tag{25}$$

where V is a $b \times \hat{b}$ collector matrix [36],

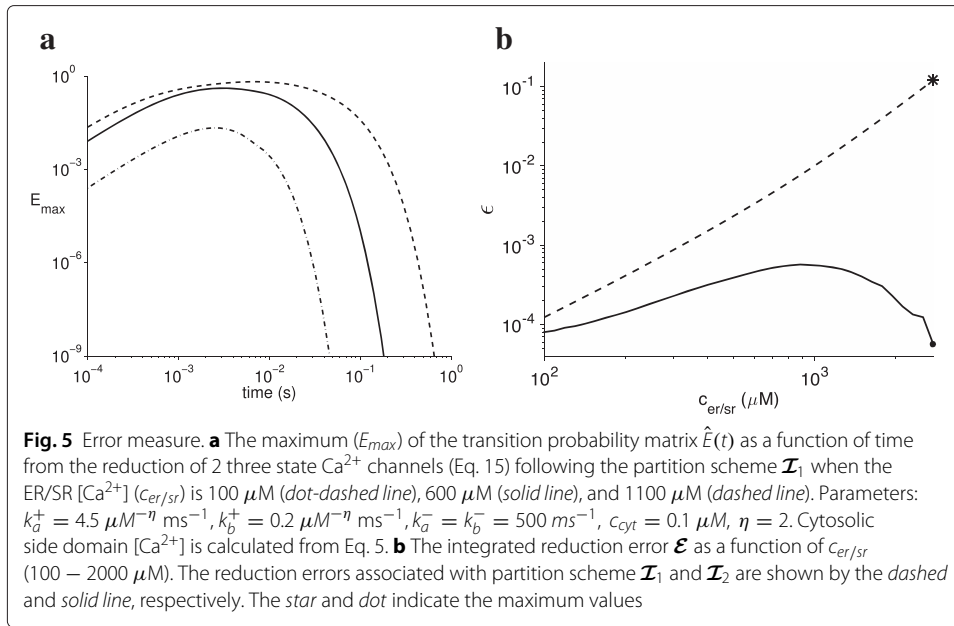$$V = \begin{bmatrix} e_1 & 0 & \cdots & 0 \\ 0 & e_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_{\hat{b}} \end{bmatrix},$$

the e_i are column vectors of ones with lengths commensurate with Q_{ii}, and U is a $\hat{b} \times b$ distributor matrix given by

$$U = \begin{bmatrix} \bar{\pi}_1 & 0 & \cdots & 0 \\ 0 & \bar{\pi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{\pi}_{\hat{b}} \end{bmatrix}. \tag{26}$$

Notice that, similar to Eqs. 23 and 24, the conditional probability distribution $\bar{\pi}_i$ of the states within group i, is calculated from the stationary distribution of the full model. The transition probabilities of the reduced model and the full model agree with each other exactly in the limit. As shown in Fig. 5a, the maximum difference of the transition probabilities falls below 10^{-9} within 1 second.

$\hat{E}(t)$ is a $b \times b$ matrix and is cumbersome to use for evaluation. Consequently, we define $E_{\max}(t) = \max_{ij} \left| \hat{E}_{ij}(t) \right|$, the element of $\hat{E}(t)$ with largest absolute value at time t. Note that $E_{\max}$ is a function of both time and $c_{er/sr}$ because the transition rates of the full Ca^{2+} release site model are functions of the luminal $[Ca^{2+}]$. Figure 5a shows $E_{\max}(t; c_{er/sr})$ for the 6-state Ca^{2+} release site model (Fig. 2b) reduced to a three-state model following the partition scheme given by $\mathcal{I}_1$. As validated in [21], the reduced model better approximates the full model as $E_{\max}$ gets smaller. The maximum transition error decreased significantly as the luminal $[Ca^{2+}]$ is dropped from 1100 μM (dashed line) to 100 μM (dot-dash-dot

Fig. 5 Error measure. **a** The maximum (E_{max}) of the transition probability matrix $\hat{E}(t)$ as a function of time from the reduction of 2 three state Ca^{2+} channels (Eq. 15) following the partition scheme $\mathcal{I}_1$ when the ER/SR [Ca^{2+}] ($c_{er/sr}$) is 100 μM (*dot-dashed line*), 600 μM (*solid line*), and 1100 μM (*dashed line*). Parameters: $k_a^+ = 4.5\ \mu M^{-\eta}\ ms^{-1}$, $k_b^+ = 0.2\ \mu M^{-\eta}\ ms^{-1}$, $k_a^- = k_b^- = 500\ ms^{-1}$, $c_{cyt} = 0.1\ \mu M$, $\eta = 2$. Cytosolic side domain [Ca^{2+}] is calculated from Eq. 5. **b** The integrated reduction error $\mathcal{E}$ as a function of $c_{er/sr}$ (100 − 2000 μM). The reduction errors associated with partition scheme $\mathcal{I}_1$ and $\mathcal{I}_2$ are shown by the *dashed* and *solid line*, respectively. The *star* and *dot* indicate the maximum values

line), indicating reductions following partition scheme $\mathcal{I}_1$ make better approximations of the full model at relatively low levels of luminal [Ca^{2+}].

In order to have the reduced model applicable in the whole cell simulation described previously, the objective we want to achieve through the genetic algorithm is to pick partitions that produce small reduction errors for all possible ER/SR [Ca^{2+}] values at all times. Consequently, we define

$$\mathcal{E}(c_{er/sr}) = \int E_{\max}(t, c_{er/sr})dt, \tag{27}$$

the area under each curve in Fig. 5a. Then, for any partition scheme $\mathcal{I}$, the integrated error $\mathcal{E}$ can be calculated as a function of $c_{er/sr}$, and the maximum $\mathcal{E}(c_{er/sr})$ is selected as the global reduction error of scheme $\mathcal{I}$. In Fig. 5b the dashed and solid lines show the integrated error $\mathcal{E}$ associated with $\mathcal{I}_1$ and $\mathcal{I}_2$, respectively, as a function of $c_{er/sr}$ (100 − 2000 μM).

Because partitions that result in lower reduction errors are preferred, the *fitness* of a given partition scheme $\mathcal{F}$ is defined by

$$\mathcal{F}_{\mathcal{I}} = \frac{1}{\max_{c_{er/sr}} \mathcal{E}_{\mathcal{I}}(c_{er/sr})}. \tag{28}$$

As shown in Fig. 5b, when the full model is partitioned and lumped following $\mathcal{I}_1$ (dashed line), the maximum possible error (star) generated by the reduced model $\mathcal{E}_1$ is approximately 210 times larger than the maximum error (dot) generated by using $\mathcal{I}_2$. The *fitness* of $\mathcal{I}_1$ is consequently 210 times less than the fitness of $\mathcal{I}_2$.

Selection and reproduction

This section introduces how the genetic algorithm implementation forms the "*next generation*" from the current *population*. The conventional **reproduction** process in genetic algorithms usually consists of **selection**, **crossover** and **mutation**.

We start with building a discrete probability distribution that is used to select *parents* of the next generation of set partitions. The probability mass function (PMF), which indicates the probability $\mathcal{P}_i$ for each "*individual*" $\mathcal{I}_i$ to be selected, is

$$\mathcal{P}_i = \frac{\mathcal{F}_i}{\sum_i \mathcal{F}_i} \qquad (1 \leq i \leq N_p), \tag{29}$$

that is, the probabilities of selection are proportional to the *fitness*, $\mathcal{F}_i$. To generate each *child*, we start by randomly selecting a pair of *parents*, one at a time, from the current *population* following the corresponding PMF. For example, the probability of $\mathcal{I}_1$ being selected is 210 times smaller than the probability that $\mathcal{I}_2$ is selected.

After a pair of *parents* are selected, a *child* is generated through the **crossover** process, which takes the permutation of states from one *parent* and the group sizes from the other *parent*. For example, if the permutation of states is taken from $\mathcal{I}_2$ ({1}, {2}, {3, 4, 5, 6}) and the group sizes are taken from $\mathcal{I}_1$ ({1, 2}, {3, 5, 6}, {4}) then the *child* would be

$$(\{1, 2\}, \{3, 4, 5\}, \{6\}).$$

After a *child* is produced, with probability p, a **mutation** process begins by randomly selecting and joining two groups of states in its partition scheme, and then randomly splitting this aggregated group of states into two new internally connected groups. For example, a possible mutation process of the *child* could be joining the second and third group:

$$(\{1, 2\}, \{3, 4, 5, 6\}),$$

then randomly split the aggregated group in to two new connected groups and a valid mutation is:

$$(\{1, 2\}, \{3\}, \{4, 5, 6\}).$$

In our implementation of the genetic algorithm, a *child* may undergo multiple mutations of this kind. A geometric distribution is assigned to the number of mutations n_m for each *child*:

$$Pr(n_m = k) = (1 - p)^{k-1} p, \tag{30}$$

where $k = 1, 2, 3, \ldots$ and $p = 0.8$.

Note that the **reproduction** process generates one *child* at a time and continues until the number of *children* reaches N_p. These *children* then forms a new generation of *individuals* and are sent to the **evaluation** process, starting a new iteration in the genetic algorithm. The genetic algorithm is executed until either a set partition has a *fitness* $\mathcal{F} \geq 1000$ (reduction error $\mathcal{E}$ less than 0.1 % for all luminal [Ca^{2+}] values) is found, or the maximum number of allowed iterations (2000) is reached.

Results

In this section we first validate the genetic algorithm implemented in the previous section by showing that the algorithm converges and produces set partition schemes that generate small reduction errors. To demonstrate that the genetic algorithm can be applied to general Ca^{2+} channels, we use this approach to reduce a release site that is composed of several four-state channels (Fig. 7) under the constraint that each group of states must be connected in the state transition diagram. This four-state model features activation by

cytosolic Ca^{2+} and luminal $[Ca^{2+}]$ regulation of the activation affinity. The reduced Ca^{2+} release site model is integrated into the whole cell model, and simulation results from the reduced and the full model are compared.
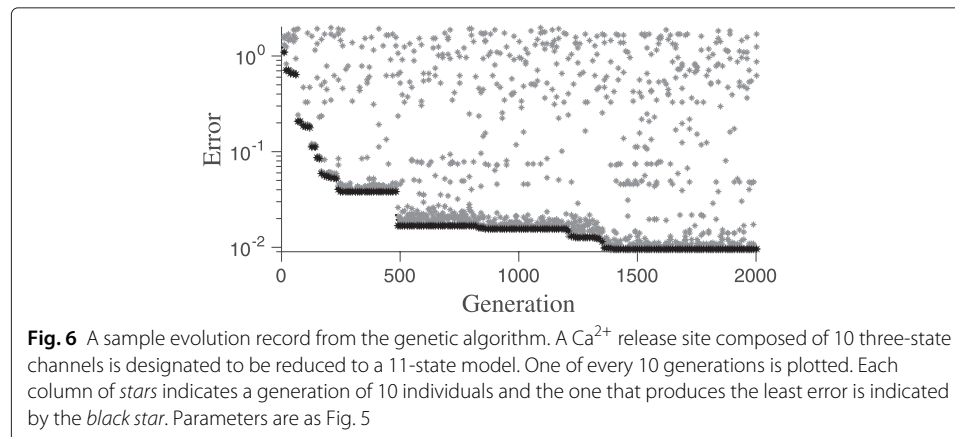
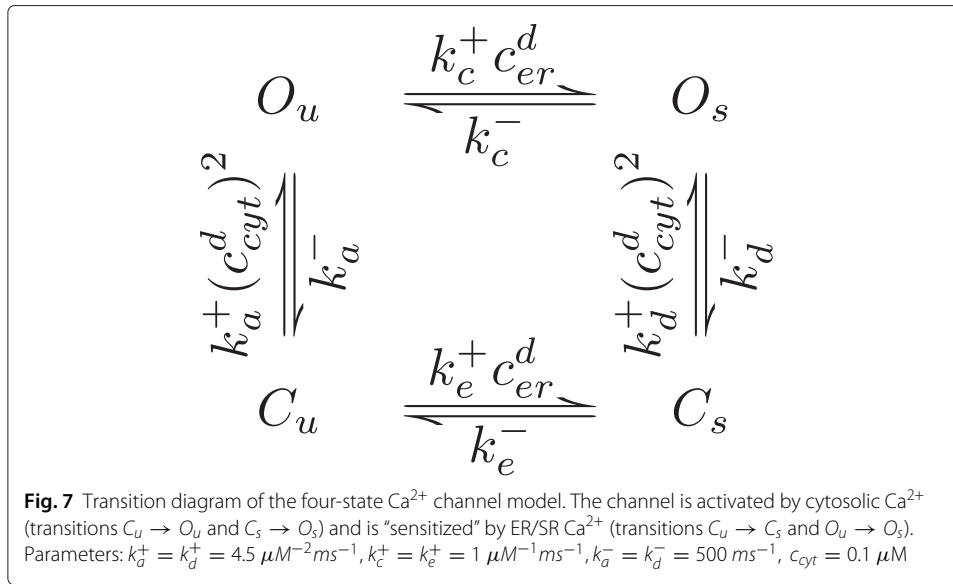### Reducing Ca^{2+} release site models that are composed of three-state channels

Figure 6 shows an example of the convergence of the genetic algorithm. We applied the genetic algorithm to reduce a Ca^{2+} release site model that is composed of 10 three-state channels (66 states) to an 11-state model. The population size $N_p = 10$ and the reduction error $\mathcal{E}$ was measured for 50 log-spaced $c_{er/sr}$ values ranging from 100 μM to 2000μM. Each column of stars represents the 10 *individuals* of a *generation*. The black stars indicate the *individual* (partition) that produces the smallest error in its *generation*. The criterion that ends the program is defined as $\mathcal{E} < 10^{-3}$ or 2000 generations are generated, whichever is satisfied first. In this specific reduction experiment, the program did generate 2000 generations and the minimum $\mathcal{E}$ was 0.0043.

Reducing Ca^{2+} release site models that are composed of four-state channels with luminal regulation

Here we demonstrate that reduced Ca^{2+} release models can replace the full model in the whole cell Ca^{2+} homeostasis model with good accuracy. To validate that the reduction procedure fits a wide variety of models, we introduce a four-state model (Fig. 7) which is activated by cytosolic Ca^{2+} and the activation affinity is regulated by the luminal $[Ca^{2+}]$.

The four-state Ca^{2+} channel model is assumed to have a regular or "unsensitized" mode (states C_u, O_u) in which the activation dissociation constant ($K_a = \sqrt{k_a^-/k_a^+}$) is higher than the activation dissociation constant ($K_d = \sqrt{k_d^-/k_d^+}$) of the "sensitized" mode (C_s, O_s). We also assume that the channel is more likely to be in the "sensitized" mode when the ER/SR $[Ca^{2+}]$ is high. $k_i^+ \left(c_{cyt}^d\right)^2$, $k_j^+ c_{er/sr}^d$ and k_i^-, for $i, j \in \{a, ..., e\}$, are transition rates with units of reciprocal time. k_i^+ is an association rate constant with units of conc$^{-\eta}$ time^{-1}, where η is the cooperativity of Ca^{2+} binding while c_{cyt}^d and $c_{er/sr}^d$ are the domain $[Ca^{2+}]$ experienced by the release site on the cytosol and ER/SR side respectively. Notice that we assume that the Ca^{2+} binding cooperativity ($\eta = 1$) of the channel sensitization (luminal regulation) process is different from the binding cooperativity of the activation process ($\eta = 2$).

Fig. 6 A sample evolution record from the genetic algorithm. A Ca^{2+} release site composed of 10 three-state channels is designated to be reduced to a 11-state model. One of every 10 generations is plotted. Each column of *stars* indicates a generation of 10 individuals and the one that produces the least error is indicated by the *black star*. Parameters are as Fig. 5

Fig. 7 Transition diagram of the four-state Ca^{2+} channel model. The channel is activated by cytosolic Ca^{2+} (transitions $C_u \to O_u$ and $C_s \to O_s$) and is "sensitized" by ER/SR Ca^{2+} (transitions $C_u \to C_s$ and $O_u \to O_s$). Parameters: $k_a^+ = k_d^+ = 4.5 \, \mu M^{-2} ms^{-1}$, $k_c^+ = k_e^+ = 1 \, \mu M^{-1} ms^{-1}$, $k_a^- = k_d^- = 500 \, ms^{-1}$, $c_{cyt} = 0.1 \, \mu M$

An important motivation in using this four-state model is that luminal regulation of RyRs is observed in many experiments [37, 38] but the detailed mechanism is yet not clear. In this paper, we are interested in how the "sensitization" of the activation of each individual Ca^{2+} channel affects the cooperative gating of the Ca^{2+} release site. Consequently we experiment on different sensitized activation rates as well as the dissociation constant $K_c = k_c^-/k_c^+$ (Fig. 7) of the sensitization process. On the other hand, the parameters of the regular or "unsensitized" Ca^{2+} activation were chosen to be consistent with the parameters in [28], where many puff/spark statistics of a group of 10 two-state Ca^{2+}-activated channels were studied.

The number of Ca^{2+} release sites is assumed to be large so that the distribution of release site states can be well approximated by $\pi(t)$ (solved from Eq. 18) instead of using Monte Carlo simulation. However, simply substituting π_n for f_n in Eqs. 13 and 14 will fail because the "fast domain" assumption is a singular limit of the ODE system. Consequently, instead of using Eqs. 11 and 12 we consider the total cytosolic ($\hat{c}_{cyt}$) and ER/SR [Ca^{2+}] ($\hat{c}_{er/sr}$), which are sums of the bulk and domain concentrations weighted by effective volume ratios,

$$\hat{c}_{cyt} = c_{cyt} + \Lambda_{cyt}^d \bar{c}_{cyt}^d \tag{31}$$

$$\hat{c}_{er/sr} = c_{er/sr} + \frac{\Lambda_{sr}^d}{\lambda_{sr}} \bar{c}_{er/sr}^d. \tag{32}$$

where $\bar{c}_{cyt}^d$ and $\bar{c}_{er/sr}^d$ are the given by

$$\bar{c}_{cyt}^d = \sum_{n=0}^{N} \pi_n c_{cyt}^{d,n} \tag{33}$$

$$\bar{c}_{er/sr}^d = \sum_{n=0}^{N} \pi_n c_{er/sr}^{d,n}, \tag{34}$$

which are the mean values of the cytosolic and SR domain Ca^{2+} concentrations. The effective volume ratios in Eqs. 31 and 32 are given by

$$\Lambda^d_{cyt} = \frac{V^{d,T}_{cyt}}{V_{cyt}} \tag{35}$$

$$\Lambda^d_{sr} = \frac{V^{d,T}_{sr}}{V_{cyt}}, \tag{36}$$

where $V^{d,T}_{cyt}$ and $V^{d,T}_{sr}$ are the effective volumes of the aggregated cytosolic and SR domains, respectively. The equations that balance $\hat{c}_{cyt}$ and $[Ca^{2+}]\,\hat{c}_{er/sr}$ are given by:

$$\frac{d\hat{c}_{cyt}}{t} = J^T_{rel} + J_{leak} - J_{pump} + J_{pm} \tag{37}$$

$$\frac{d\hat{c}_{er/sr}}{t} = \frac{1}{\lambda_{er/sr}}\left(-J^T_{rel} - J_{leak} + J_{pump}\right). \tag{38}$$

The total release flux J^T_{rel} is given by

$$J^T_{rel} = \sum_{n=0}^{N} \pi_n \gamma_n v^T_{rel}\left(c^{d,n}_{er/sr} - c^{d,n}_{cyt}\right), \tag{39}$$

where $\gamma_n = n/N$, $c^{d,n}_{cyt}$ and $c^{d,n}_{er/sr}$ are given by Eqs. 5, 6 and 7, and π_n is the probability that a randomly sampled release site has n open channels, which can be found from $\boldsymbol{\pi} = (\pi_0, \pi_1, \cdots, \pi_N)$ by integrating Eq. 18.

Figure 8 shows a comparison of 20 numerical calculations of the stationary dynamics of a Ca^{2+} release site composed of 10 four-state RyRs (286 states, lines) and the corresponding reduced 34 state model (circles and crosses) using different values of the disassociation rate of sensitization K_C. The filled circles and triangles show the results of a release site composed of 10 two-state RyRs. When the disassociation rate of sensitization K_C is high enough, the sensitized states are rarely visited and consequently the four-state model results should approach the two-state model results. As shown in Fig. 8 the four-state model well approximates the two-state model when K_C is approximately 1000 μM.

In Fig. 8, panel a shows that decreasing the K_C will decrease the bulk SR $[Ca^{2+}]$ and the results calculated from the reduced Ca^{2+} release site model is a close approximation to the full model. Figure 8b shows the open probability of a single four-state channel (dashed line) as a function of K_C. The solid line in Fig. 8b shows the fraction of open channels, f_O of the 10-channel release site, where
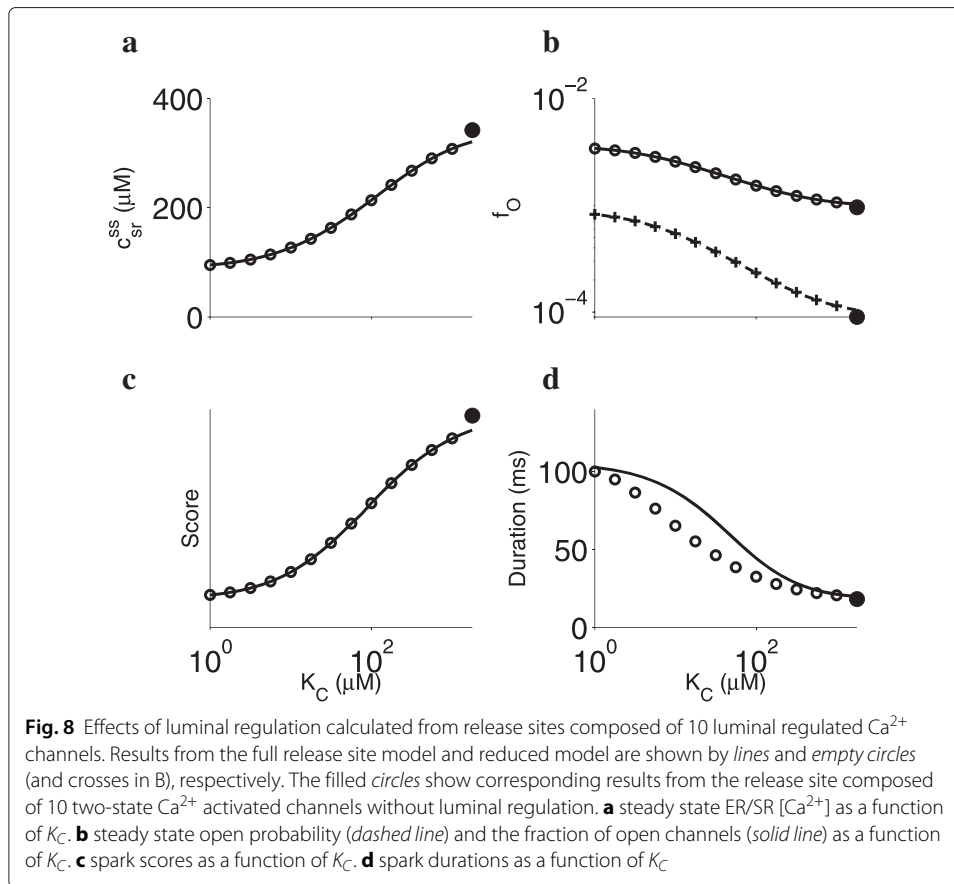
$$f_O = E\left[N_O\right]/N, \tag{40}$$

and

$$E\left[N_O\right] = \sum_{n=0}^{N} n\pi_n \tag{41}$$

is the average number of open channels per release site. The reduced model gives a good approximation for both parameters of the full model (empty circles and crosses). As K_C decreases, both parameters increase, which indicates adding the sensitized states increases the open probability of the channels. The increased open probability further

Fig. 8 Effects of luminal regulation calculated from release sites composed of 10 luminal regulated Ca^{2+} channels. Results from the full release site model and reduced model are shown by *lines* and *empty circles* (and crosses in B), respectively. The filled *circles* show corresponding results from the release site composed of 10 two-state Ca^{2+} activated channels without luminal regulation. **a** steady state ER/SR $[Ca^{2+}]$ as a function of K_C. **b** steady state open probability (*dashed line*) and the fraction of open channels (*solid line*) as a function of K_C. **c** spark scores as a function of K_C. **d** spark durations as a function of K_C

causes a lower steady state SR $[Ca^{2+}]$, which is consistent with Fig. 8a. In prior work by Nguyen and colleagues [6], a puff/spark *Score* was defined as

$$Score = \frac{\text{Var}\left[f_O\right]}{\text{E}\left[f_O\right]} = \frac{1}{N}\frac{\text{Var}\left[N_O\right]}{\text{E}\left[N_O\right]} \tag{42}$$

from which the presence or absence of puff/spark can be assessed. This measure ranges between 0 and 1, and values that are larger than 0.2 indicate the presence of robust Ca^{2+} puffs/sparks. Figure 8c, shows the *Score*s of the full model and the reduced model as a function of K_C. The reduced model *Score*s give a close approximation to the full model results.

Notice that the *Score* values are above 0.35 for all K_C values, indicating robust Ca^{2+} puffs/sparks present in both the full and reduced model. We further studied the mean duration of spontaneous Ca^{2+} puffs/sparks occurring as a function of K_C in the whole cell formulation, shown in Fig. 8d. We assume that a transition from $N_O = 4$ to $N_O = 5$ is considered to initialize a puff/spark and a transition from $N_O = 1$ to $N_O = 0$ (all channels closed) terminates the puffs/spark. The mean puffs/spark duration was calculated using the matrix analytic method described in [39]. As K_C decreases, the channels are more likely to be sensitized, the puff/spark duration increases, indicating that the luminal regulation of the channel might lead to longer puffs/sparks, which is consistant with experimental observations [38, 40]. Compared to the Ca^{2+} release site model composed of 10 two-state channels (filled circle), the average puff/spark duration of a release site composed of the same number of four-state channels can be up to four times

(when $K_C = 1$) longer. While in prior work [21], similar comparisons to a Ca^{2+} release site composed of Keizer-Levine model [41] and its corresponding reduced model gave good agreement, in this new study the reduced model tends to slightly underestimate the puff/spark durations.

Conclusions and discussion

A brief summary

We have implemented and validated a novel genetic algorithm-based searching technique to find reduced models that produce moderate errors for Ca^{2+} release site models that are compositionally defined from single channel Markov models. Given a full model and the designated size of the reduced model, this algorithm samples and evolves a population of set partitions, each corresponding to a potential scheme for state aggregation, that leads to the partitions that lead to reduced models which approximate the full model on the behaviors of interest. A Ca^{2+} release site composed of 10 four-state channels that are activated by the cytosolic Ca^{2+} and regulated by luminal Ca^{2+} is reduced by this technique and the steady state responses of the reduced model well approximate the full model in the minimal whole cell homeostasis environment (Fig. 8).

When a Ca^{2+} release site model is reduced, the resulting models are designated to have significantly fewer states, which is inevitably accompanied by losing some transition information. Different state aggregation schemes may preserve different information. A main benefit from using genetic algorithms is that the **evaluation** function is flexible enough to pick state aggregation schemes that maximize any information that is of specific interest to the user. In this report, for example, we are interested in how luminal regulation affects the spark behavior of the Ca^{2+} release site, and the **evaluation** function is consequently designed to assign higher *fitness* to the partitions which generate small errors in a wide range of ER/SR [Ca^{2+}]. As another example, if the spark frequency is crucial in some study, we can conveniently edit the **evaluation** function to calculate the spark frequency of each reduced model generated from partition $\mathcal{I}$ and assign higher *fitness* to the ones that better approximate the full model spark frequency. When focusing on a single release site, behaviors of interest that could be implemented include, but not limited to: the number of open channels, transition probabilities among specific states, puff/spark amplitude, puff/spark durations, and inter-puff/spark-intervals. When considering Ca^{2+} diffusion and homeostasis, hybrid stochastic and deterministic simulations as used by Rückl and colleagues [42, 43] can be implemented in the **evaluation** function such that reduced models approximate the [Ca^{2+}] wave and oscillation statistics of the full model.

Comparison to previous work

As compared to the fast/slow reduction technique [21], the genetic algorithm-based approach does not require time scale differences and allows users to choose the size of the reduced model freely. More importantly, the genetic algorithm-based approach often finds partition schemes that produce less error than the fast/slow technique. It is also important to note that this procedure performs better on models whose parameters are of the same scale. When time scale differences are present, like in the Keizer-Levine model and the De Young-Keizer model, because for every *Individual* $\mathcal{I}$, we must reduce the full model following the aggregation scheme, calculating the reduction error by computing matrix exponentials. Thus, it is recommended to fine tune the genetic algorithm

parameters to achieve faster convergence. The choice of crossover and/or mutation probabilities, selection techniques, and even population size per generation could affect the performance of genetic algorithms [44]. For example, when reducing a full model consisting of 10 four-state channels, if using 10 as the population size and 2000 generations are generated until the program terminates, the total time consumed is approximately 2000 times that of the fast/slow procedure. If the population size was increased to 200 individuals per generation, the genetic algorithm can find equally good reductions approximately 10 times faster. Fortunately, for any specific objective assigned, the reduction procedures need to execute very few times and the reduced release site models are potentially able to save significantly more time in the whole cell simulations. Furthermore, this genetic algorithm-based reduction technique can be hybridized with other stochastic optimization methods. For example, we implemented one version of the genetic algorithm with a simulated annealing twist, where the **crossover** procedure was skipped and every individual that survived selection would generate a child through mutation. This version converges faster as compared to the traditional genetic algorithm when applied to some release site models.

Future work

As discussed in the 'Background' section, how to control computational cost of large mathematical/computational models has become an increasingly important research topic. The recent work by Cao and colleagues [45] that proposed a deterministic model of IP3Rs that qualitatively predicts some stochastic Ca^{2+} oscillation properties is very encouraging. This two-state model was constructed by reducing a 6-state stochastic IP3R model [46], where the six states were partitioned into two groups assuming time scale differences and the experimentally observed "two mode" property. Then each group was lumped to a single state according to the steady state probability distribution of the full model. Not only can this deterministic model quantitatively reproduce Ca^{2+} puffs and stochastic oscillations, but their model is also approximately 10 times faster than the comparable stochastic simulations. It would be very beneficial if we can find simple deterministic models that can replace other stochastic calcium channel models under certain circumstances. Since not all models possess time scale differences, we can potentially use the genetic algorithm approach to search for partition schemes of stochastic IP3R or RyR models that allow the reduced deterministic models to well approximate the stochastic behaviors of these channels.

Another important project for the near future is to search for common features in the partition schemes that produce small errors. Should the aggregated states in the reduction follow certain topological pattern or possess similar functional feature (having similar number of open/refractory channels for example) the model reduction approach can be significantly accelerated by using a biased initial population. So far, an interesting phenomenon observed while reducing Ca^{2+} release site models using the genetic algorithm based approach is that the state aggregation schemes which result in small reduction errors tend to be "heavy headed". That is, these low-error partitions usually feature one large group that contains more than 50 % of the states while other groups contain significantly fewer (sometimes only one or two) states. Moreover, the states aggregated in the small groups are highly likely to be the states that are less often visited in the full model, and this phenomenon exists in all Ca^{2+} release site reduction procedures.

This observation is a good explanation for the fact that the generator matrices associated with the reduced model $\hat{Q}$ tend to be ill-conditioned. This observation suggests that it may be possible to generate a biased initial population to accelerate the evolution procedure.

References
1. Alder BJ, Wainwright TE. Studies in molecular dynamics. i. general method. J Chem Phys. 1959;31(2):459.
2. Plimpton SJ. Computational limits of classical molecular-dynamics simulations. Comput Mater Sci. 1995;4:361–4.
3. Brini E, Algaer EA, Ganguly P, Li C, Rodríguez-Ropero F, van der Vegt NFA. Systematic coarse-graining methods for soft matter simulations - a review. Soft Matter. 2013;7:2108–119.
4. Saunders MG, Voth GA. Coarse-graining methods for computational biology. Annu Rev Biophys. 2013;42(2):73–93.
5. Zheng J, Vankataramanan L, Sigwortha FJ. Hidden markov model analysis of intermediate gating steps associated with the pore gate of shaker potassium channels. J Gen Physiol. 2001;118(5):547–64.
6. Nguyen V, Mathias R, Smith GD. A stochastic automata network descriptor for markov chain models of instantaneously-coupled intracellular Ca^{2+} channels. Bull Math Biol. 2005;67(3):393–432.
7. Clapham DE. Calcium signaling. Cell. 1995;80(2):259–68.
8. Berridge MJ. Elementary and global aspects of calcium signalling. J Physiol (Lond). 1997;499(Pt 2):291–306.
9. Cheng H, Lederer MR, Lederer WJ, Cannell MB. Ca^+ sparks and $[Ca^{2+}]_i$ waves in cardiac myocytes. Am J Physiol. 1996;270(1 Pt 1):148–59.
10. Yao Y, Choi J, Parker I. Quantal puffs of intracellular Ca^{2+} evoked by inositol trisphosphate in Xenopus oocytes. J Physiol. 1995;482(Pt 3):533–3.
11. Endo M. Calcium release from the sarcoplasmic reticulum. Phys Rev. 1977;57(1):71–108.
12. Cheng H, Lederer WJ, Cannell MB. Calcium sparks: elementary events underlying excitation-contraction coupling in heart muscle. Science. 1993;262(5134):740–4.
13. Parker I, Choi J, Yao Y. Elementary events of IP3-induced Ca^{2+} liberation in Xenopus oocytes: hot spots, puffs and blips. Cell Calcium. 1996;20(2):105–21.
14. Shuai JW, Jung P. Optimal ion channel clustering for intracellular calcium signaling. Proc Natl Acad Sci USA. 2003;100(2):506–10.
15. Cannell MB, Cheng H, Lederer WJ. Spatial non-uniformities in $[Ca^{2+}]_i$ during excitation-contraction coupling in cardiac myocytes. Biophys J. 1994;67(5):1942–56.
16. Colquhoun D, Hawkes A. A Q-matrix cookbook: how to write only one program to calculate the sigle-channel and macroscopic predictions for any kinetic mechanism In: Sakmann B, Neher E, editors. Single-Channel Recording. New York: Plenum Press; 1995. p. 589–633.
17. Smith GD. Modeling the stochastic gating of ion channels In: Fall C, Marland E, Wagner J, Tyson J, editors. Computational Cell Biology. New York: Springer; 2002. p. 291–325.
18. DeRemigio H, Smith GD. Calcium release site ultrastructure and the dynamics of puffs and sparks. Math Med Biol. 2008;25(1):65–85.
19. Williams GSB, Huertas MA, Sobie EA, Jafri MS, Smith GD. A probability density approach to modeling local control of Ca^{2+}-induced Ca^{2+} release in cardiac myocytes. Biophys J. 2007;92(7):2311–28.
20. Williams GSB, Huertas MA, Sobie EA, Jafri MS, Smith GD. Moment closure for local control models of Ca^{2+}-induced Ca^{2+} release in cardiac myocytes. Biophys J. 2008;95(4):1689–703.
21. Hao Y, Kemper P, Smith GD. Reduction of calcium release site models via fast/slow analysis and iterative aggregation/disaggregation. Chaos. 2009;5(19):037107.
22. Feldmann AE. Fast balanced partitioning is hard even on grids and trees In: Rovan B, Sassone V, Widmayer P, editors. Mathematical Foundations of Computer Science 2012. Berlin Heidelberg: Springer; 2012. p. 372–82.
23. Holland JH. Adaptation in Natural and Artificial Systems. Ann Arbor: The U. of Michigan Press; 1975.
24. Gesú VD, Giancarlo R, Bosco GL, Raimondi A, Scaturro D. Genclust: a genetic algorithm for clustering gene expression data. BMC Bioinformatics. 2005;6:289.
25. To C, Vohradsky J. A parallel genetic algorithm for single class pattern classification and its application for gene expression profiling in streptomyces coelicolor. BMC Genomics. 2007;8:49.
26. Hill T, Lundgren A, Fredriksson R, Schioth H. Genetic algorithm for large-scale maximum parsimony phylogenetic analysis of proteins. Biochim Biophys Acta. 2005;1725(1):19–29.
27. Groenendaal W, Ortega FA, Kherlopian AR, Zygmunt AC, Krogh-Madsen T, Christini DJ. Cell-specific cardiac electrophysiology models. PLoS Comput Biol. 2015;11(4):1004242.

28. Hartman JA, Sobie EA, Smith GD. Calcium sparks and homeostasis in a minimal model of local and global calcium responses in quiescent ventricular myocytes. AJP: Heart Circ Physiol. 2010. doi:10.1152/ajpheart.00293.2010.
29. Hinch R, Greenstein JL, Tanskanen AJ, Xu L, Winslow RL. A simplified local control model of calcium-induced calcium release in cardiac ventricular myocytes. Biophys J. 2004;87(6):3723–6.
30. Hinch R, Greenstein JL, Winslow RL. Multi-scale models of local control of calcium induced calcium release. Prog Biophys Mol Biol. 2006;90(1-3):136–50.
31. Greenstein JL, Hinch R, Winslow RL. Mechanisms of excitation-contraction coupling in an integrative model of the cardiac ventricular myocyte. Biophys J. 2006;90(1):77–91.
32. Mazzag B, Tignanelli C, Smith GD. The effect of residual Ca^{2+} on the stochastic gating of Ca^{2+}-regulated Ca^{2+} channels. J Theor Biol. 2005;235(1):121–50.
33. Huertas MA, Smith GD. The dynamics of luminal depletion and the stochastic gating of Ca^{2+}-activated Ca^{2+} channels and release sites. J Theor Biol. 2007;246(2):332–54.
34. Davis L. Handbook Of Genetic Algorithms. New York: Van Nostrand Reingold; 1991.
35. Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs. New York: Springer; 1994.
36. Nicola V. Lumping in markov reward processes. Technical report, RC14719, IBM Thomas Watson Research Centre, PO Box 704, Yorktown Heights, NY 10598;1998.
37. Shannon TR, Wang F, Puglisi J, Weber C, Bers DM. A mathematical treatment of integrated Ca^{2+} dynamics within the ventricular myocyte. Biophys J. 2004;87(5):3351–71.
38. Stevens SC, Terentyev D, Kalyanasundaram A, Periasamy M, Györke S. Intra-sarcoplasmic reticulum Ca^{2+} oscillations are driven by dynamic regulation of ryanodine receptor function by luminal Ca^{2+} in cardiomyocytes. J Physiol (Lond). 2009;587(20):4863–72. published in October 2009.
39. Groff JR, Smith GD. Calcium-dependent inactivation and the dynamics of calcium puffs and sparks. J Theor Biol. 2008;253(3):483–99.
40. Györke I, Györke S. Regulation of the cardiac ryanodine receptor channel by luminal Ca^{2+} involves luminal Ca^{2+} sensing sites. Biophys J. 1998;75(6):2801–10.
41. Keizer J, Levine L. Ryanodine receptor adaptation and Ca^{2+}(-)induced Ca^{2+} release-dependent Ca^{2+} oscillations. Biophys J. 1996;71(6):3477–487.
42. Rückl M, Parker I, Marchant JS, Nagaiah C, Johenning FW, Rüdiger S. Modulation of elementary calcium release mediates a transition from puffs to waves in an IP3R cluster model. PLoS Comput Biol. 2015;11(1):1003965.
43. Rüdiger S, Shuai JW, Huisinga W, Nagaiah C, Warnecke G, Parker I, Falcke M. Hybrid stochastic and deterministic simulations of calcium blips. Biophys J. 2007;93:1847–57.
44. Karafotias G, Hoogendoorn M, Eiben AE. Parameter control in evolutionary algorithms: Trends and challenges. IEEE Trans Evol. 2015;19(2):167–87.
45. Cao P, Tan X, Donovan G, Sanderson MJ, Sneyd J. A deterministic model predicts the properties of stochastic calcium oscillations in airway smooth muscle cells. PLoS Comput Biol. 2014;10(8):1003783.
46. Cao P, Donovan G, Falcke M, Sneyd J. A stochastic model of calcium puffs based on single-channel data. Biophys J. 2013;105:1133–42.