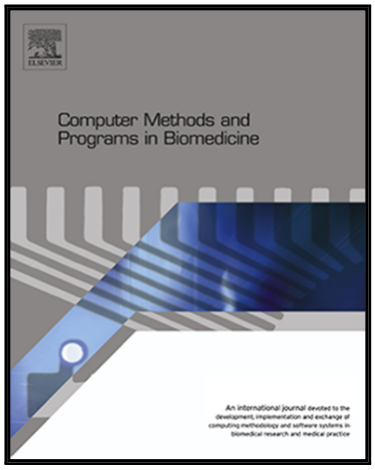
Journal Pre-proof



Model-driven optimal experimental design for calibrating cardiac electrophysiology models

Chon Lok Lei, Michael Clerx, David J. Gavaghan, Gary R. Mirams

PII: S0169-2607(23)00355-3
DOI: https://doi.org/10.1016/j.cmpb.2023.107690
Reference: COMM 107690

To appear in: *Computer Methods and Programs in Biomedicine*

Received date: 14 March 2023
Revised date: 9 June 2023
Accepted date: 22 June 2023

Please cite this article as: Chon Lok Lei, Michael Clerx, David J. Gavaghan, Gary R. Mirams, Model-driven optimal experimental design for calibrating cardiac electrophysiology models, *Computer Methods and Programs in Biomedicine* (2023), doi: https://doi.org/10.1016/j.cmpb.2023.107690

Highlights

- Current models of cardiac action potentials reflect only an average cell behaviour and ignore cell-to-cell variability.
- Optimal single-cell experiments are designed to reduce calibration uncertainty and improve predictive power in this study.
- These optimal designs perform better than many of the commonly used designs, with shorter experiment durations.
- These optimal designs perform better than many of the commonly used designs, with shorter experiment durations.

Model-driven optimal experimental design for calibrating cardiac electrophysiology models

Chon Lok Lei[*,1,2], Michael Clerx[3], David J. Gavaghan[4,5], and Gary R. Mirams[*,3]

[1] *Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau, China.*
[2] *Department of Biomedical Sciences, Faculty of Health Sciences, University of Macau, Macau, China.*
[3] *Centre for Mathematical Medicine & Biology, School of Mathematical Sciences, University of Nottingham, Nottingham, United Kingdom.*
[4] *Department of Computer Science, University of Oxford, Oxford, United Kingdom.*
[5] *Doctoral Training Centre, University of Oxford, Oxford, United Kingdom.*

Abstract

Background and Objective: Models of the cardiomyocyte action potential have contributed immensely to the understanding of heart function, pathophysiology, and the origin of heart rhythm disturbances. However, action potential models are highly nonlinear, making them difficult to parameterise and limiting to describing 'average cell' dynamics, when cell-specific models would be ideal to uncover inter-cell variability but are too experimentally challenging to be achieved. Here, we focus on automatically designing experimental protocols that allow us to better identify cell-specific maximum conductance values for each major current type.

Methods and Results: We developed an approach that applies optimal experimental designs to patch-clamp experiments, including both voltage-clamp and current-clamp experiments. We assessed the models calibrated to these new optimal designs by comparing them to the models calibrated to some of the commonly used designs in the literature. We showed that optimal designs are not only overall shorter in duration but also able to perform better than many of the existing experiment designs in terms of identifying model parameters and hence model predictive power.

Conclusions: For cardiac cellular electrophysiology, this approach will allow researchers to define their hypothesis of the dynamics of the system and automatically design experimental protocols that will result in theoretically optimal designs.

Keywords: optimal experimental design, mathematical modelling, model calibration, electrophysiology, patch clamp, action potential

1 Introduction

Cardiac cellular electrophysiology is the study of how cells manipulate ionic concentrations and membrane voltage to initiate and synchronise heartbeats. Having demonstrated their usefulness in fundamental research, quantitative models of the action potential (AP) are increasingly being used (Plank et al., 2021) in safety-critical applications (Mirams et al., 2012; Niederer et al., 2018) such as optimising patient treatments (Corral-Acero et al., 2020; Niederer et al., 2020) or assessing drug safety (Li et al., 2019). These models contain several parameters, which may represent chemical properties such as binding rates, but also patient and even cell-specific properties such as ion current conductances (which are determined by gene expression levels). In many cases such parameters are not measured directly, but *inferred* indirectly from experimental recordings of membrane potential or transmembrane current. To create accurate models, we need to ensure the experiments used in their creation contain sufficient information on all parameters we need to infer. However, the experiments used in this process have not usually been designed with model calibration in mind, and as a result some parameters may be poorly constrained by the available data (Clerx et al., 2019a; Whittaker et al., 2020). Optimal experimental

[*]Corresponding authors: chonloklei@um.edu.mo, gary.mirams@nottingham.ac.uk

design (OED) is a method of designing experiments which uses an objective approach to elicit data containing "optimal information" for model calibration (Lindley et al., 1956; Kiefer, 1959; Smucker et al., 2018; Seurat et al., 2021). It is extremely popular in physical sciences such as geosciences (Seidler et al., 2016; Ushijima et al., 2021), mechanical engineering (Gupta and Dhingra, 2013; Gherardini et al., 2021), chemical engineering (Gottu Mukkula and Paulen, 2017; Schenkendorf et al., 2018), etc. In this study, we apply the techniques of OED to cardiac cellular electrophysiology to expedite the development of AP models suitable for both basic research and safety-critical applications.

The electrophysiological properties of many types of isolated cells can be studied and examined via patch-clamp experiments (Hamill et al., 1981) in which the experimenter either controls ('clamps') the cell's membrane potential and measures the resulting transmembrane current (voltage-clamp mode), or controls the current while measuring potential (current-clamp mode). In either mode, a pre-determined waveform (i.e. a sequence of voltages or currents) is applied, which we shall refer to as the experimental *protocol*. OED is a methodology of designing protocols based on the intuition that a model parameter can only be estimated from a measured model output (e.g. a current or voltage trace) if that output is strongly sensitive to the parameter value during the experiment. In short, it produces protocols that *maximise* (in some exact sense to be chosen by the experimenter) the model output's sensitivity to all parameters of interest.

We apply an OED strategy to design experimental protocols that allow us to better identify cell-specific maximum conductance values for the major current types in mathematical cardiac myocyte models. Using simulation results from a controllable, understandable cellular system, we show that OED can successfully be applied to both voltage-clamp and current-clamp experiments to obtain cell-specific models. Moreover, the results demonstrate how OED protocols perform better than many of the existing experiments in the literature, in terms of reducing uncertainty in model parameters and its theoretical predictive power. This study provides a road map for researchers in cardiac cellular electrophysiology to automatically design theoretically-optimal experimental protocols.

2 Methods

In the design of experiments, the results of OED are experimental designs that are optimal with respect to some statistical criteria—*optimal design measures*. Formally, we consider a (controlled) nonlinear differential equation model of the form

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{d}), \boldsymbol{\theta}), \tag{1}$$

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{d}), \boldsymbol{\theta}), \tag{2}$$

where $\boldsymbol{x}$ is the vector of model states, $\boldsymbol{u}$ is the vector of system inputs, $\boldsymbol{\theta}$ is the vector of model parameters. $\boldsymbol{f}$ and $\boldsymbol{h}$ are the systems of equations, where $\boldsymbol{f}$ describes the dynamical equations, and $\boldsymbol{h}$ maps the solutions of the dynamical equations to the vector of observables $\boldsymbol{y}$, i.e. the model outputs that we compare with data. We assume that the system input $\boldsymbol{u}$ can be parameterised with $\boldsymbol{d}$, i.e. $\boldsymbol{d}$ is the vector of parameters that controls the experimental conditions. For example, if $\boldsymbol{u}$ is an applied voltage-clamp protocol, $\boldsymbol{d}$ could be the voltages and durations of its steps; if $\boldsymbol{u}$ is an applied current stimulus, $\boldsymbol{d}$ could be the durations between each stimulus, as described in details in the next sections. The optimisation of the experimental design procedure is then defined as

$$\hat{\boldsymbol{d}} = \underset{\boldsymbol{d}}{\operatorname{argmin}}\, \Phi(\boldsymbol{d}), \tag{3}$$

where the function Φ represents the optimal design measure to be minimised.

All optimal design measures discussed in this paper assume the model calibration process to be least-square estimation, maximum likelihood estimation, or posterior estimation; if one considers a different calibration process, such as bounded-error parameter estimation (also known as guaranteed parameter estimation) or approximate Bayesian computation (ABC), biases may be introduced, and a different set of design measures should be used (Gottu Mukkula and Paulen, 2017). All methods are implicitly conditional on the chosen set of model equations, i.e. they are only optimal for the models used during the design; this is a general limitation of model-based OED.

2.1 Action potential models for OED

We consider cardiac cellular electrophysiology models of APs under either a voltage-clamp or a current-clamp experiment, and we aim to optimise either the voltage-clamp or the current-clamp protocol for conductance identification. In practise, we design experiments and perform the fitting to data with a proposed model which would be our current best representation of reality. A schematic overview of the method is given in Figure 1.

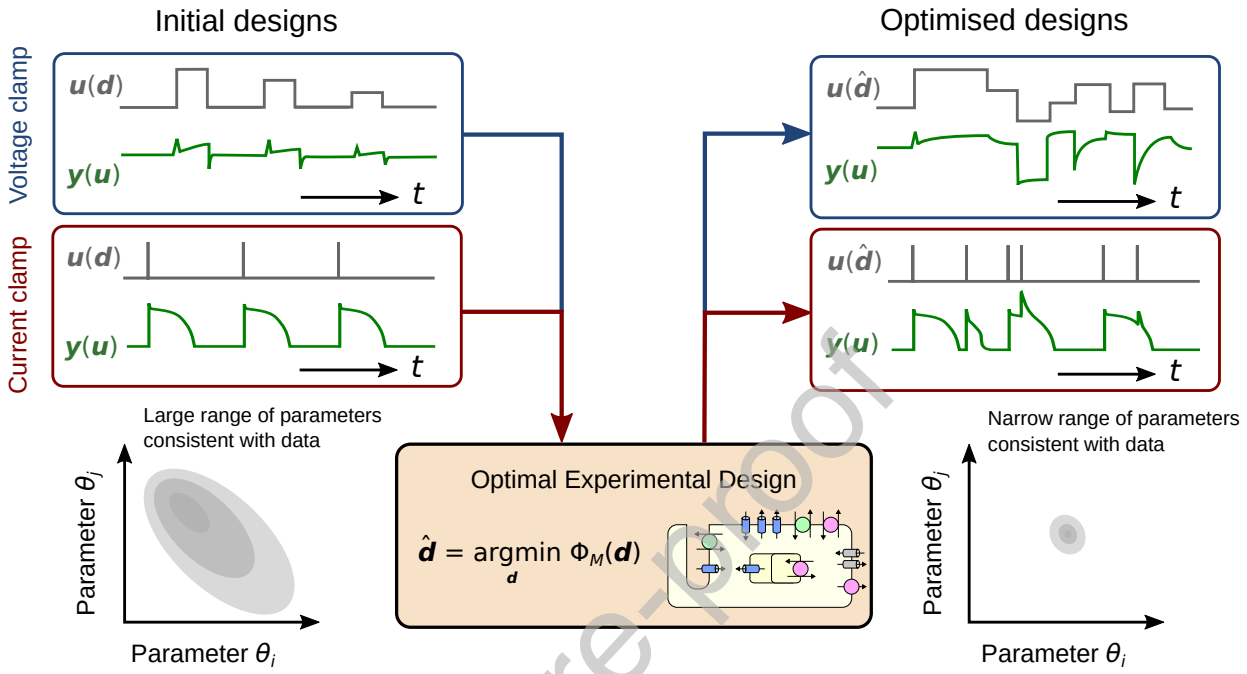


Figure 1: A schematic of optimal experimental design for patch-clamp experiments. Initial designs (left, $\boldsymbol{u}(\boldsymbol{d})$) of the voltage-clamp and current-clamp experiments, which can commonly be found in the literature or perhaps are 'intuitive' choices of the design inputs ($\boldsymbol{d}$), elicit uninformative data/observables $\boldsymbol{y}(\boldsymbol{u})$ for calibrating complex mathematical models, leading to a wide posterior for the model parameters $\boldsymbol{\theta}$ (shown for two parameters but the principle applies for higher numbers of parameters). Optimal experimental designs with the mathematical model M under a certain statistical criterion Φ_M derive optimised designs (right, $\boldsymbol{u}(\hat{\boldsymbol{d}})$) with non-intuitive choices of design inputs $\hat{\boldsymbol{d}}$ which elicit informative data for calibration, and lead to narrower parameter posteriors.

Designing experiments with a single model

Here, we first use a widely-used adult ventricular (epicardial) AP model by O'Hara et al. (2011) as a 'proposed model'. Under the voltage-clamp configuration, the observed current I_{obs} can be expressed as

$$I_{\text{obs}}(t) = \sum_j g_j \cdot x_j(V_{\text{m}}, t), \tag{4}$$

where g_j is the maximum conductance or permeability and $x_j(V)$ is some nonlinear function of the voltage V_{m} for the current of type j, and we consider the major currents only (Eq. (6)). These nonlinear functions $x_j(V)$ are the product of the 'kinetics' (describing the voltage-dependent opening and closing of ion channels in response to changes in membrane voltage) and the 'driving term' of the currents, either the Ohmic (membrane potential minus reversal potential) term or the Goldman-Hodgkin-Katz flux equation.

For our voltage-clamp experimental designs, the output of interest ($\boldsymbol{y}$ in Eq. (2)) is the observed current I_{obs}, and the vector of system inputs $\boldsymbol{u}$ describes the applied voltage-clamp protocol $V_{\text{m}}(t)$.

Alternatively, for current-clamp experiments, the observed membrane voltage V_{m} follows

$$\frac{\mathrm{d}V_{\mathrm{m}}}{\mathrm{d}t} = -\frac{1}{C_{\mathrm{m}}}\big(I_{\mathrm{obs}}(t) + I_{\mathrm{stim}}(t)\big), \tag{5}$$

where $I_{\mathrm{obs}}(t)$ is the same as Eq. (4). In this case, the membrane voltage V_{m} is the output of interest ($\boldsymbol{y}$ in Eq. (2)), and the vector of system inputs $\boldsymbol{u}$ describes the externally-applied current-clamp protocol, i.e. the stimulus current $I_{\mathrm{stim}}(t)$.

We *assume* that the x_j kinetics equations and their parameters are known (and correct). We are interested in finding all g_j, which determine the magnitude of the currents when the channels are fully open, therefore we have

$$\boldsymbol{\theta} = \{g_{\mathrm{Na}}, g_{\mathrm{CaL}}, g_{\mathrm{Kr}}, g_{\mathrm{Ks}}, g_{\mathrm{to}}, g_{\mathrm{NaCa}}, g_{\mathrm{K1}}, g_{\mathrm{NaK}}\} \tag{6}$$

to be inferred, which are the maximum conductance or permeability of the fast sodium current, the L-type calcium current, the rapid-delayed rectifier potassium current, the slow-delayed rectifier potassium current, the transient potassium current, the sodium-calcium exchanger current, the inward rectifier potassium current, and the sodium-potassium pump current, respectively. We note that this ignores some smaller background currents in the model.

Simulations were run using Myokit (Clerx et al., 2016), with tolerance settings for the CVODE solver (Hindmarsh et al., 2005) set to `abs_tol` $= 10^{-8}$ and `rel_tol` $= 10^{-10}$.

Designing experiments with an averaged measure over models

Each OED is optimal for the model used to perform the design. However, we may want to obtain a design that is optimal for a range of biological assumptions instead of focusing on one particular model, so that the same data can be used to study multiple similar models. Therefore, we also consider using multiple models for calculating the cost function; instead of using only one model to optimise the experiment while the true data generating process could potentially be more similar to another model. That is, if we have N_M proposed models, then each model output $I_{\mathrm{obs},M}$ (for voltage clamp) or $V_{\mathrm{m},M}$ (for current clamp) is used to calculate an optimal design measure Φ_M, and we optimise the experimental protocol parameters using

$$\hat{\boldsymbol{d}} = \underset{\boldsymbol{d}}{\operatorname{argmin}} \underset{M}{\mathbb{E}} \left(\Phi_M(\boldsymbol{d})\right). \tag{7}$$

It minimises the mean of the measures Φ_M calculated independently for each of the N_M proposed models. Here we use five adult ventricular (epicardial) AP models to represent an averaged behaviour: ten Tusscher et al. (2004), Fink et al. (2008), O'Hara et al. (2011), Chang et al. (2017), and Tomek et al. (2019). Note that the averaged measure (over models) is used only for the designing the experiments, i.e. obtaining $\hat{\boldsymbol{d}}$, and only one model will be used at a time during predictions.

2.2 Experiments to be optimised: Two modes of experiments

We design the experiments by varying the system input $\boldsymbol{u}(\boldsymbol{d})$. As described in the model section, we could measure the system with two modes: voltage clamp and current clamp.

Parameterising voltage-clamp experiment designs

The voltage-clamp protocol V is the system input $\boldsymbol{u}(\boldsymbol{d})$, and we choose it to be a piecewise function defined as

$$\boldsymbol{u}(\boldsymbol{d}) = V(t) = \begin{cases} V_1, & 0 \le t < T_1 \\ V_2, & T_1 \le t < T_2 \\ \vdots & \vdots \\ V_N, & T_{N-1} \le t < T_N. \end{cases} \tag{8}$$

The vector of parameters that defines the voltage-clamp protocol is

$$\boldsymbol{d} = \{V_1, T_1, V_2, T_2, \ldots, V_N, T_N\} \tag{9}$$

for a N-step protocol; equivalently we can define $\Delta T_i = T_i - T_{i-1}$ with $T_0 = 0$, then we obtain

$$\boldsymbol{d} = \{V_1, \Delta T_1, V_2, \Delta T_2, \ldots, V_N, \Delta T_N\}. \tag{10}$$

In order to ensure the resulting voltage-clamp protocol V is feasible to be run experimentally, we constrain the protocol to have $V_i \in [-120, 60]\,\text{mV}$ and $\Delta T_i \in [50, 2000]\,\text{ms}$ for all i, which defines the protocol parameter space. We choose N to be 20, giving a maximum protocol duration of 40 s. We applied both the local sensitivity analysis (LSA) and global sensitivity analysis (GSA) designs for this experimental setting.

In theory, the voltage-clamp protocol can be replaced with any arbitrary function, e.g. a sum of sinusoidal functions (Beattie et al., 2018), as long as it can be parameterised with some vector of parameters $\boldsymbol{d}$. However, the benefit of choosing such a piecewise step function is that it can be directly applied in any patch-clamp system, including some of the high-throughput automated patch-clamp systems such as the SyncroPatch from Nanion Technologies (Lei et al., 2019b,a).

Parameterising current-clamp experiment designs

In the current-clamp mode, the stimulus current I_{stim} is the system input $\boldsymbol{u}(\boldsymbol{d})$. Partly inspired by Groenendaal et al. (2015), who *randomly* applied stimulus pulses to create an informative non-uniform protocol, here we *optimise* when to apply stimulus pulses. We parameterised the current-clamp protocol as follows: the stimulus current with an amplitude $\overline{I_{\text{stim}}}$ and duration T_{stim} is applied at times $T_1, T_2, \ldots, T_N$ (and is otherwise zero), where $T_{i+1} > iT_{\text{stim}} + T_i$. The values of $\overline{I_{\text{stim}}}$ and T_{stim} are chosen to be those given in the original publication of each model. Similar to the voltage-clamp protocol parameterisation, equivalently we can define $\Delta T_i = T_{i+1} - iT_{\text{stim}} - T_i$, giving

$$\boldsymbol{d} = \{\Delta T_1, \Delta T_2, \ldots, \Delta T_N\}. \tag{11}$$

We further constrain $\Delta T_i \in [100, 2000]\,\text{ms}$, $\forall i$, to make the designed experiments practically usable. We applied only the LSA designs for this experimental setting; to apply GSA designs for current-clamp experiments, we would need to first explore the boundaries (not necessarily being a hypercube) of the parameter space that generate APs.

2.3 OED design measures

Local sensitivity analysis (LSA)-based designs

The sensitivity matrix based on LSA, $\mathbf{S}_{\text{L}}(\boldsymbol{d} \mid \boldsymbol{\theta} = \hat{\boldsymbol{\theta}})$, is defined as

$$(\mathbf{S}_{\text{L}})_{ij} = \left.\frac{\partial y_i}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{12}$$

Here the subscript i (row) in y_i runs through all model outputs in $\boldsymbol{y}$ and all sampled time points t_k*, and the subscript j (column) θ_j goes over all model parameters in $\boldsymbol{\theta}$. The *Fisher information matrix* (FIM) is given by

$$\text{FIM}(\boldsymbol{\theta}) = \mathbf{S}_{\text{L}}^T\, \boldsymbol{\Sigma}^{-1}\, \mathbf{S}_{\text{L}}, \tag{13}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the measurement data (noise). Here we assume that data are from a single output (a single observed time series trace) with independent and identically distributed (i.i.d.) zero mean and σ^2 variance Gaussian noise through time such that $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ with σ constant and $\mathbf{I}$ the identity matrix.

Many local design criteria are based on the parameter (co)variance matrix defined as

$$\mathbf{C}_{\boldsymbol{\theta}} = \mathbb{E}\left[(\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta}])(\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta}])^T\right] \geq \text{FIM}(\boldsymbol{\theta})^{-1}, \tag{14}$$

*That is, for N_o model outputs and N_t time points, $i = 1, 2, \ldots, N_o, N_o + 1, N_o + 2, \ldots, N_o \times N_t$.

where $\mathbb{E}[\cdot]$ denotes the expected value. This describes the covariance of the estimates of a fixed true parameter set[†]. For any unbiased estimator, where the expected value is equal to the true value of the parameter set, a lower bound for the variance is given by the inverse of the FIM (the inequality in Eq. (14)), known as the Cramér-Rao bound (Walter and Pronzato, 1997). Therefore, to evaluate the local design criteria based on the covariance matrix, we use the the Cramér-Rao bound and calculate $\mathrm{FIM}(\boldsymbol{\theta})^{-1}$ instead (Walter and Pronzato, 1997; Schenkendorf et al., 2018). In fact, equality with the Cramér-Rao bound would be achieved as K (number of observations) approaches infinity

$$\lim_{K\to\infty} \mathbf{C}_{\boldsymbol{\theta}} = \mathrm{FIM}(\boldsymbol{\theta})^{-1}. \tag{15}$$

For proof of Eq. (15) see, for example, Pant (2018). Note, $\mathbf{C}_{\boldsymbol{\theta}}$, FIM, and $\mathbf{S}_\mathrm{L}$ all depend on the choice of the system input $\boldsymbol{d}$.

Common criteria for local design are the '*alphabetic family*' (Kiefer, 1959; Walter and Pronzato, 1997). These criteria each represent a cost function to be minimised, a subset of them (A-optimal, D-optimal, and E*-optimal designs, Atkinson and Donev, 1992) are given in Table 1, where $\lambda_{\max}(\mathbf{X})$ and $\lambda_{\min}(\mathbf{X})$, respectively, denote the maximum and minimum eigenvalues of a matrix $\mathbf{X}$.

Local design criteria	Cost functions
A-optimal design	$\Phi_\mathrm{A} = \mathrm{trace}([\mathbf{S}_\mathrm{L}^T\,\mathbf{S}_\mathrm{L}]^{-1})$
D-optimal design	$\Phi_\mathrm{D} = \det([\mathbf{S}_\mathrm{L}^T\,\mathbf{S}_\mathrm{L}]^{-1})$
E*-optimal design	$\Phi_{\mathrm{E}^*} = \lambda_{\max}([\mathbf{S}_\mathrm{L}^T\,\mathbf{S}_\mathrm{L}]^{-1})/\lambda_{\min}([\mathbf{S}_\mathrm{L}^T\,\mathbf{S}_\mathrm{L}]^{-1})$

Table 1: Local design criteria for OED. Here we have assumed the parameter covariance matrix $\mathbf{C}_{\boldsymbol{\theta}}$ for some given model parameters $\boldsymbol{\theta}$ can be approximated as $\sigma^{-2}[\mathbf{S}_\mathrm{L}^T\mathbf{S}_\mathrm{L}]^{-1}$, where σ^{-2} is dropped from the design criteria (cost functions) as it is assumed to be constant. E* is the modified E criterion.

The design criteria in Table 1 are parameter based criteria (Atkinson and Donev, 1992): these cost functions (as functions of the FIM) can be interpreted as properties (the shape and size) of the *confidence hyper-ellipsoid* (a generalisation of the confidence interval for multivariate statistics) for the parameters $\boldsymbol{\theta}$ (Vanrolleghem and Dochain, 1998; Banga and Balsa-Canto, 2008). Therefore, by improving certain properties of the confidence ellipsoid, the uncertainty of the inferred parameters using the data measured under an optimally designed protocols should (in theory) be reduced.

One obvious limitation of the local designs is that they are *local*, i.e. they depend on a particular choice of the model parameters ($\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ in the equations above). Hence the design is only guaranteed to be optimal for that particular choice of parameters. However, we do not know the true model parameters in the experiments by definition. This issue could be alleviated by replacing the local sensitivity with an averaged sensitivity matrix

$$\mathbb{E}[(\mathbf{S}_\mathrm{L})_{ij}] = \int_\Theta (\mathbf{S}_\mathrm{L})_{ij}\, p(\theta_j)\mathrm{d}\theta_j, \tag{16}$$

where $p(\theta_j)$ is the probability density function of the j^{th} parameter, defined over a domain Θ (a subspace of the parameter space). The integral in Eq. (16) can be approximated using Monte Carlo simulations. Although this averaged sensitivity matrix has been shown to give a more robust design (Chu and Hahn, 2013), it does not take the multivariate interaction into account (Schenkendorf et al., 2018), and hence the framework of global sensitivity comes into play.

Global sensitivity analysis (GSA)-based designs

One of the most straight forward ways of doing GSA-based design is by replacing the $\mathbf{S}_\mathrm{L}$ in Table 1 with a sensitivity matrix based on GSA methods (Rodriguez-Fernandez et al., 2007; Kucerová et al., 2016;

[†] Note that under this framework, the true parameters do not vary, as opposed to the Bayesian framework where the true parameters can be treated as random variables.

Schenkendorf et al., 2018). We can use variance-based methods such as the *Sobol method* (also known as the *Sobol indices*, Sobol, 2001). The Sobol method decomposes the variance of the model output into fractions, and attributes these variances to parameters or sets of parameters, usually referred to as the *order*. The first-order Sobol index attributes these variances to each parameter θ_j alone; whilst higher order Sobol indices attribute the variances to different combinations of parameters. These variances are calculated using a set of parameters sampled within a domain Θ, a subspace of the parameter space. These parameters can be sampled using various methods, for example Monte Carlo methods, or some low-discrepancy sequences (including Sobol sequences or Latin hypercube sampling); we used the extension of the Sobol sequence by Saltelli (2002) and Saltelli et al. (2010) here. The first-order Sobol sensitivity index is given by

$$(\mathbf{S}_\mathrm{G})_{ij} = \frac{1}{\mathrm{Var}\,(y_i)} \underset{\theta_j}{\mathrm{Var}} \left(\underset{\boldsymbol{\theta}_{-j}}{\mathbb{E}} \left[y_i \mid \theta_j \right] \right), \tag{17}$$

where $\mathrm{Var}\,(y_i)$ is the total variance of the model output y_i, $\mathrm{Var}_{\theta_j}(X) = \mathbb{E}_{\theta_j}(X^2) - (\mathbb{E}_{\theta_j}(X))^2$ denotes the conditional variance only over parameter θ_j, and $\mathbb{E}_{\boldsymbol{\theta}_{-j}}[\cdot]$ indicates the expected value taken over all parameters $\boldsymbol{\theta}$ except parameter θ_j. Criteria using $\mathbf{S}_\mathrm{G}$ are given in Table 2. Note that one may further include higher-order Sobol indices for analysing multivariate parameter interactions and dependencies (Kucerová et al., 2016), but since we would like to maximise the independent sensitivity of the parameters for the experimental design, using the first-order sensitivities is sufficient.

GSA-based design criteria	Cost functions
A-optimal design	$\Phi_\mathrm{A}^\mathrm{GSA} = \mathrm{trace}([\mathbf{S}_\mathrm{G}^T\,\mathbf{S}_\mathrm{G}]^{-1})$
D-optimal design	$\Phi_\mathrm{D}^\mathrm{GSA} = \det([\mathbf{S}_\mathrm{G}^T\,\mathbf{S}_\mathrm{G}]^{-1})$
E*-optimal design	$\Phi_{\mathrm{E}^*}^\mathrm{GSA} = \lambda_\mathrm{max}([\mathbf{S}_\mathrm{G}^T\,\mathbf{S}_\mathrm{G}]^{-1})/\lambda_\mathrm{min}([\mathbf{S}_\mathrm{G}^T\,\mathbf{S}_\mathrm{G}]^{-1})$

Table 2: Global design criteria for OED based on a generalisation of the local design criteria in Table 1 using a GSA method.

2.4 Optimisation of protocol designs

To perform the optimisation of the protocols as defined by Eq. (3), we optimise the design measures (cost functions) by varying the protocol parameters ($\boldsymbol{d}$ in Eq. (10) for voltage clamp and in Eq. (11) for current clamp). The protocol was initialised with parameters $\boldsymbol{d} = \boldsymbol{d}_0$ randomly sampled uniformly within the boundaries (the protocol parameter space) defined in Section 2.2. Optimisations were performed using the covariance matrix adaptation evolution strategy (Hansen, 2006) via our open source Python package, PINTS (Clerx et al., 2019b). This was repeated 10 times with different initial protocol parameters, and the best result out of all repeats was used and presented in the results.

To compute the Sobol indices, we define a hypercube for the model parameter space Θ, where each conductance value g_j in Eq. (6) can be scaled from $e^{-2} \approx 0.14$ to $e^2 \approx 7.39$ times its original value. We use the Saltelli et al. (2010) extension of a Sobol sequence to generate 512 parameter samples within this parameter space Θ for computing the Sobol indices.

To numerically calculate the local sensitivity matrix $\mathbf{S}_\mathrm{L}$, we use the first-order central difference to approximate the local derivatives with a step-size of 0.1% of the parameter values. It may be worth noting that although LSA-based criteria require local derivatives, we are not truly interested in (local) infinitesimal changes when designing the protocols, hence a reasonable approximation of the derivatives typically sufficies.

2.5 Comparing the performance of OEDs

Cross-measure evaluations

We defined a cross-measure matrix $\mathbf{X}$ for each optimised protocol, such that each entry $X_{i,j}$ was the criterion j score for this protocol with currents generated by model i, with $1 < j < N_\mathrm{measure}$ and

$1 < i < N_{\text{model}}$. The models indexed by i were the ten Tusscher et al. (2004) model, the Fink et al. (2008) model, the O'Hara et al. (2011) model, the Chang et al. (2017) model, the Tomek et al. (2019) model, and the averaged model. The criteria indexed by j were the LSA A-, D-, E*-designs, and the GSA A-, D-, E*-designs, respectively, for the voltage-clamp mode; j were the LSA A-, D-, E*-designs, respectively, for the current-clamp mode.

The numerical values for different cost functions were not directly comparable, so we normalised the values. Normalisation should be better than using a simple ranking as it should highlight outliers better. Outliers concerned us here, because we expected the best protocols to be robust to the model (possible current kinetics in reality) and to perform well under all measures. Ideally, we would like to have a protocol that is good under all criteria and all possible models of kinetics. The normalisation for each entry was calculated using $\overline{X}_{i,j} = (X_{i,j} - a_{i,j})/(b_{i,j} - a_{i,j}) \times 100\%$, where $a_{i,j}$ and $b_{i,j}$ were the minimum (best) and maximum (worst) scores seen across all the protocols (including both voltage-clamp and current-clamp experiments). Note the normalisation (percentile) was calculated with only the optimised protocols, i.e. without the benchmark protocols, for clarity. Therefore the percentiles shown in Figures 6 and 7 were not necessarily within 0 to 100%. at the entry i, j (i.e. with this criterion and current model).

Practical evaluations

We further evaluated the optimised protocols from a practical point of view. We generated synthetic data under these protocols with synthetic noise, and we assessed identifiability of the parameters in Eq. (6).

We generated synthetic data using the O'Hara et al. (2011) model with $\sim \mathcal{N}(0, \sigma^2)$ as the synthetic noise, where $\sigma = 0.15\,\text{A}\,\text{F}^{-1}$ for voltage clamp and $\sigma = 0.15\,\text{mV}$ for current clamp. Then the same model was used to fit the synthetic data, to test the ideal case where we know the ground truth kinetics. We reparameterised the models with a scaling factor s_j for the maximum conductance g_j, where $s_j = 1$ was the original literature value. The likelihood of observing the data $\mathbf{Y}$ with K time points given the model parameters $\boldsymbol{\theta}$ was

$$p(\mathbf{Y} \mid \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{K/2}} \exp\left(-\frac{\sum_{i=1}^{K} (Y_i - y_i)^2}{2\sigma^2}\right). \tag{18}$$

The posterior distribution of the parameters was

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{Y})} \tag{19}$$

$$\propto p(\mathbf{Y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{20}$$

where $p(\boldsymbol{\theta})$ is the prior, and $p(\mathbf{Y})$ is the marginal likelihood which is a constant. Uniform priors were used $\mathcal{U}(0.04, 25)$ for the scaling of the conductance values, which was wider than the subspace we used for computing the GSA to test its robustness for parameters outside the design parameter space.

The posterior distributions of the parameters were estimated using a Monte-Carlo based sampling scheme—a population MCMC (Jasra et al., 2007) algorithm with adaptive Metropolis as the base sampler—via our open source Python package, PINTS (Clerx et al., 2019b). We ran four chains for the population MCMC, each with 4×10^4 samples, and discarded the first 10^4 samples as warm-up period.

3 Results

In this section we briefly motivate our OED approach before presenting 12 new protocols derived for voltage-clamp and 6 new protocols for current-clamp experiments. Various ways to evaluate the 'quality' of these optimised protocols are explored and we discuss some of the properties that we would like or expect them to have. To check the designs, we evaluate them using both theoretical optimality measures and practical tests in which we infer parameters from simulated data. Finally, we compare the OED performance against designs found in the literature.

3.1 Optimal experimental designs for patch-clamp experiments

OED provides a framework for obtaining high-quality, statistically grounded designs $\boldsymbol{u}$ that are optimal with respect to a statistical criterion Φ. It is based on the assumption that we can parameterise the designs $\boldsymbol{u}$ with a *control* or *design variable vector* $\boldsymbol{d}$ such that it can be optimised with the criterion Φ that depends on the choice of a model M. Two types of experimental protocols ($\boldsymbol{u}$) were optimised: voltage clamp and current clamp (see § 2.2). We defined voltage-clamp protocols as sequences of steps with a magnitude and duration that could be varied by OED as the design variables $\boldsymbol{d}$. Our current-clamp protocols consisted of short bursts of current injection with a fixed magnitude and duration, here the interval between burst was used as design variable (see Figure 1). OED can also work within a variety of constraints, such as restrictions in duration of the experiment or applied voltage ranges, to ensure the optimal designs are practically feasible.

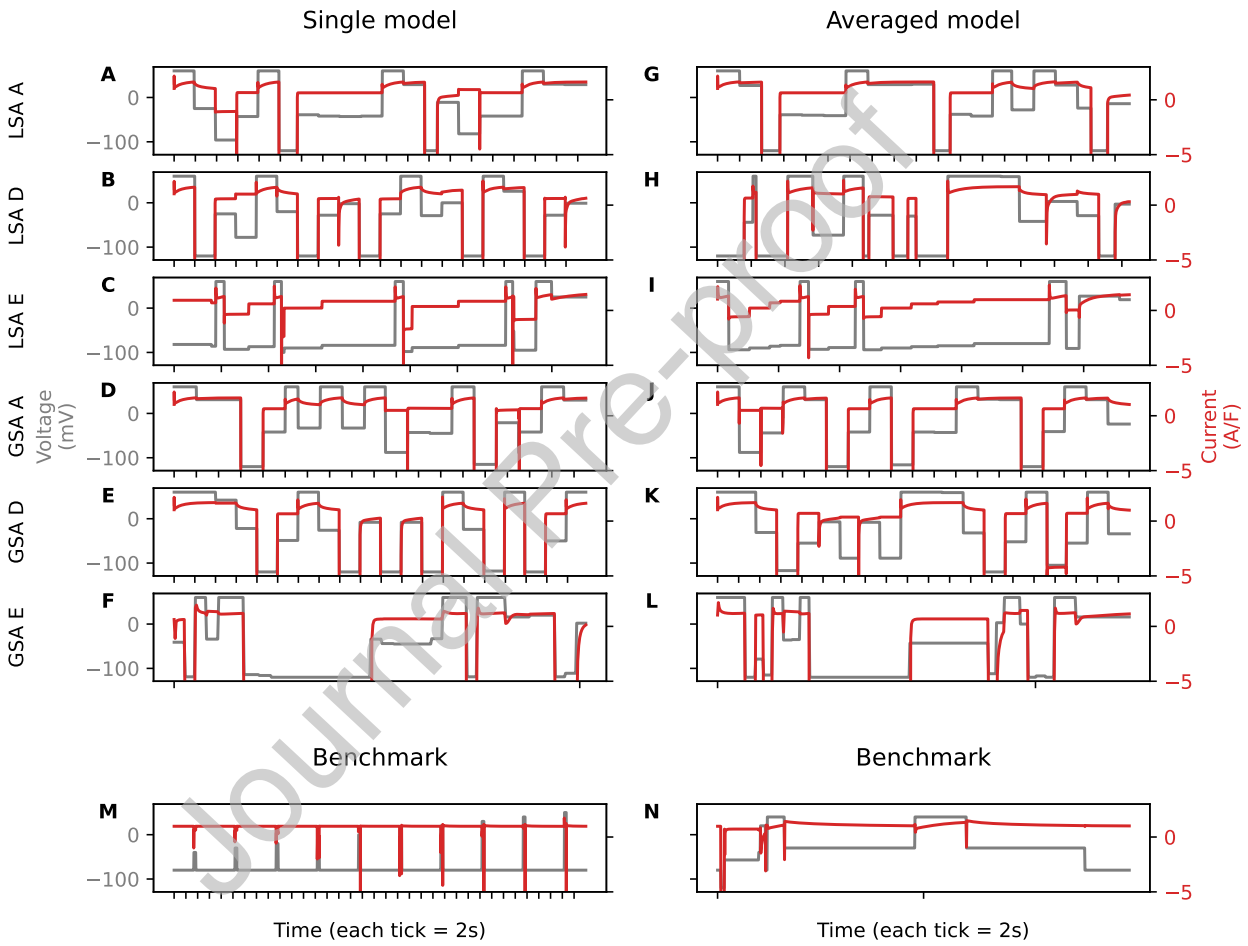


Figure 2: **(A–L)** The optimised protocols (shown in grey, values on left axis) using different criteria and the corresponding I_{obs} (shown in red, values on right axis) simulated using the O'Hara et al. (2011) model. Rows are protocols optimised with different optimal design criteria: the LSA A-, D-, E*-designs, and the GSA A-, D-, E*-designs. Columns are protocols optimised based on different types of models: the single model criterion, and the averaged model criterion. **(M, N)** The benchmark protocols from Lei et al. (2017) and from Groenendaal et al. (2015).

Experiment designs $\boldsymbol{u}(\boldsymbol{d})$ were optimised based on a design measure which consists of a choice of the optimal design criterion and a choice of model used to perform the design, using a global optimisation algorithm. For voltage-clamp experiments, we used six design criteria (LSA A-, D-, E*-designs, and GSA A-, D-, E*-designs) and two types of model, giving us in total 12 optimal design strategies (see

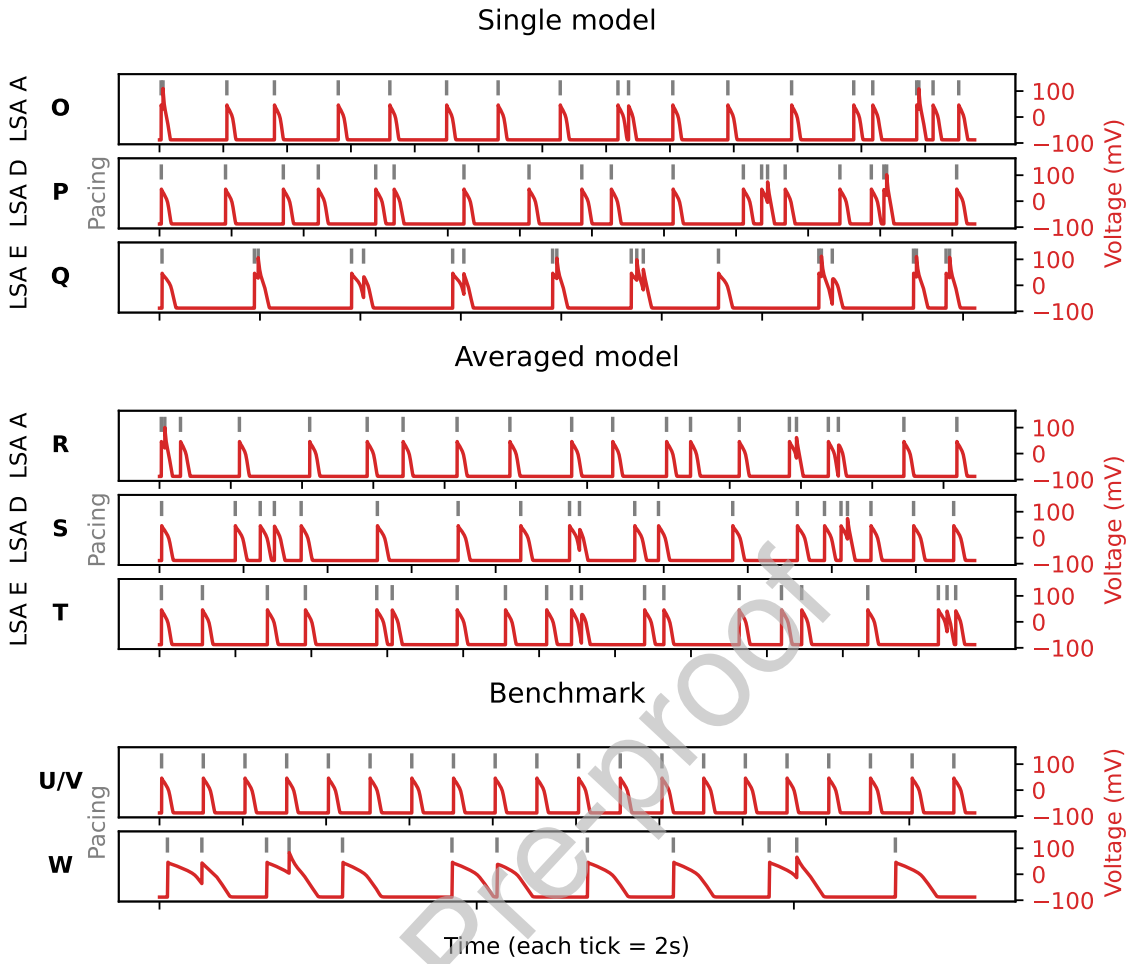


Figure 3: The optimised protocols (shown in grey, the positions at which stimuli were applied) using different criteria and the corresponding membrane voltage V_{m} (shown in red, values on right axis) simulated using the O'Hara et al. (2011) model. Rows are protocols optimised with different optimal design criteria: the LSA A-, D-, E*-designs. Protocols were optimised based on **(O–Q)** the single model criterion, and **(R–T)** the averaged model criterion. **(U/V)** The benchmark protocols with 1 Hz pacing, where Protocol U uses only the AP biomarkers for parameter inference, and Protocol V uses the full time series. **(W)** Another benchmark protocol from Groenendaal et al. (2015).

§ 2.3). These design criteria can be interpreted as properties (the shape and size) of the *confidence hyper-ellipsoid* for the calibrated model parameters $\boldsymbol{\theta}$ (Vanrolleghem and Dochain, 1998; Banga and Balsa-Canto, 2008). The A-optimal design can be interpreted as minimising the average variance of the confidence ellipsoid. The D-optimal design minimises the volume of the confidence ellipsoid. The E*-optimal design minimises the ratio of the length of the largest to the smallest axis, making the ellipsoid as spherical as possible. The resulting optimised protocols using different strategies are shown in Figure 2 (Protocols A–L), with the corresponding current ($\boldsymbol{y} = I_{\mathrm{obs}}$) simulated using the O'Hara et al. (2011) model (the corresponding eigenvalue spectra are shown in Supplementary Figures). For current-clamp experiments, we used three design criteria (the LSA A-, D-, E*-designs) and two types of model, giving us in total 6 optimal design strategies. The resulting optimised protocols using different strategies are shown in Figure 3 (Protocols O–T) with the corresponding membrane voltage ($\boldsymbol{y} = V_{\mathrm{m}}$) simulated using the O'Hara et al. (2011) model.

Each of the protocols appears to elicit rich and varied dynamics. For voltage-clamp protocols, these optimal protocols are also shorter in duration than most of the conventional protocols, such as Protocol M, where long holding steps are usually required for bringing the cells back to quasi-

steady state. Since these optimal protocols were designed to maximise the information for all model parameters for inference, we evaluate the performance of these optimised protocols and compare them against some protocols from the literature.

3.2 Do OEDs help us find model parameters?

Taking a practical approach, we generated synthetic (simulated) data under these optimised protocols $\boldsymbol{u}(\hat{\boldsymbol{d}})$, then asked the following questions: (i) can we identify the model parameters $\boldsymbol{\theta}$ correctly using these synthetic data? (ii) how does the error, if any, and uncertainty in the parameters vary between these optimised protocols?

We generated synthetic data using the O'Hara et al. (2011) model with i.i.d. Gaussian synthetic noise $\mathcal{N}(0, \sigma^2)$ with zero mean and standard deviation $\sigma = 0.15\,\mathrm{A\,F^{-1}}$ for voltage clamp and $\sigma = 1\,\mathrm{mV}$ for current clamp. Then the same model was used to fit the synthetic data, to test whether we can identify the maximum conductance parameters $\boldsymbol{\theta}$ when we have a good estimate of the kinetics. Posterior distributions of the parameters were estimated using a Markov chain Monte Carlo (MCMC) based sampling scheme (Jasra et al., 2007), where the final chains converged with $\widehat{R} < 1.1$ for all parameters (Gelman et al., 2013).

In terms of reducing the uncertainty and error in the inferred parameters, the two most successful optimal designs $\hat{\boldsymbol{d}}$ for voltage-clamp experiments were the LSA D-design and GSA A-design for the O'Hara et al. (2011) model (Protocols B and D), as shown in Figure 4A. Each chain returned a posterior distribution over the parameters, since all chains converge and they are all very similar to each other, we use all three chains to compute the marginal parameter distributions (per parameter). The mean (over all the parameters) root-mean-square error (RMSE) of these distributions to the true values is shown in Figure 4B for all of the voltage-clamp protocols. The posterior RMSEs of the individual parameter scaling factors are shown in Supplementary Table S1, which shows that the dominant contribution to the error was from s_{NaCa}. We expect this is because I_{NaCa} is reasonably small and has fairly linear dynamics under voltage clamp which happens to allow it to compensate for the sum of the errors in the other inferred conductances.

For current-clamp experiments, all the OED results had a similar posterior width, with the most successful optimal design to be the LSA A-design for the single model (Protocol O) as shown in Figure 5A. The mean (over all the parameters) RMSE of these distributions is shown in Figure 5B for all of the current-clamp protocols, and the individual parameters are shown in Supplementary Table S2. The same analyses were also repeated for the ten Tusscher et al. (2004) model, shown in Supplementary Figures S3 & S4, and Supplementary Tables S3 & S4. Similar results as for the O'Hara et al. (2011) model were observed, demonstrating the conclusions may hold across further models, especially those with similar dynamics.

3.3 Cross-measure assessments of OEDs

We may also expect the best protocol from those OEDs would be a robust good 'all-round' protocol. Each of the OED is by definition optimal under a given design measure, whilst a robust good all-round design would not be the worst under another design measure and, on average, should be good under all measures.

The performance of all the optimised protocols under a particular experimental mode were compared across all the design measures; a design measure is defined by the choices of the model and the design criterion. We used the ten Tusscher et al. (2004) model, the Fink et al. (2008) model, the O'Hara et al. (2011) model, the Dutta et al. (2017) model, the Tomek et al. (2019) model, and the averaged model to assess the OED performance; the O'Hara et al. (2011) model and the averaged model were used to optimise the protocols, and the remaining four were included to assess how these optimised protocols perform should data really arise from cells that are better described by these other models. We used the same criteria to assess the OEDs, and their combinations form what we refer to as the cross-measure matrix. Each design measure entry of the cross-measure matrix was normalised to 0–100 % with the best and the worst values seen across all the OED protocols such that it can be compared with other design measures.

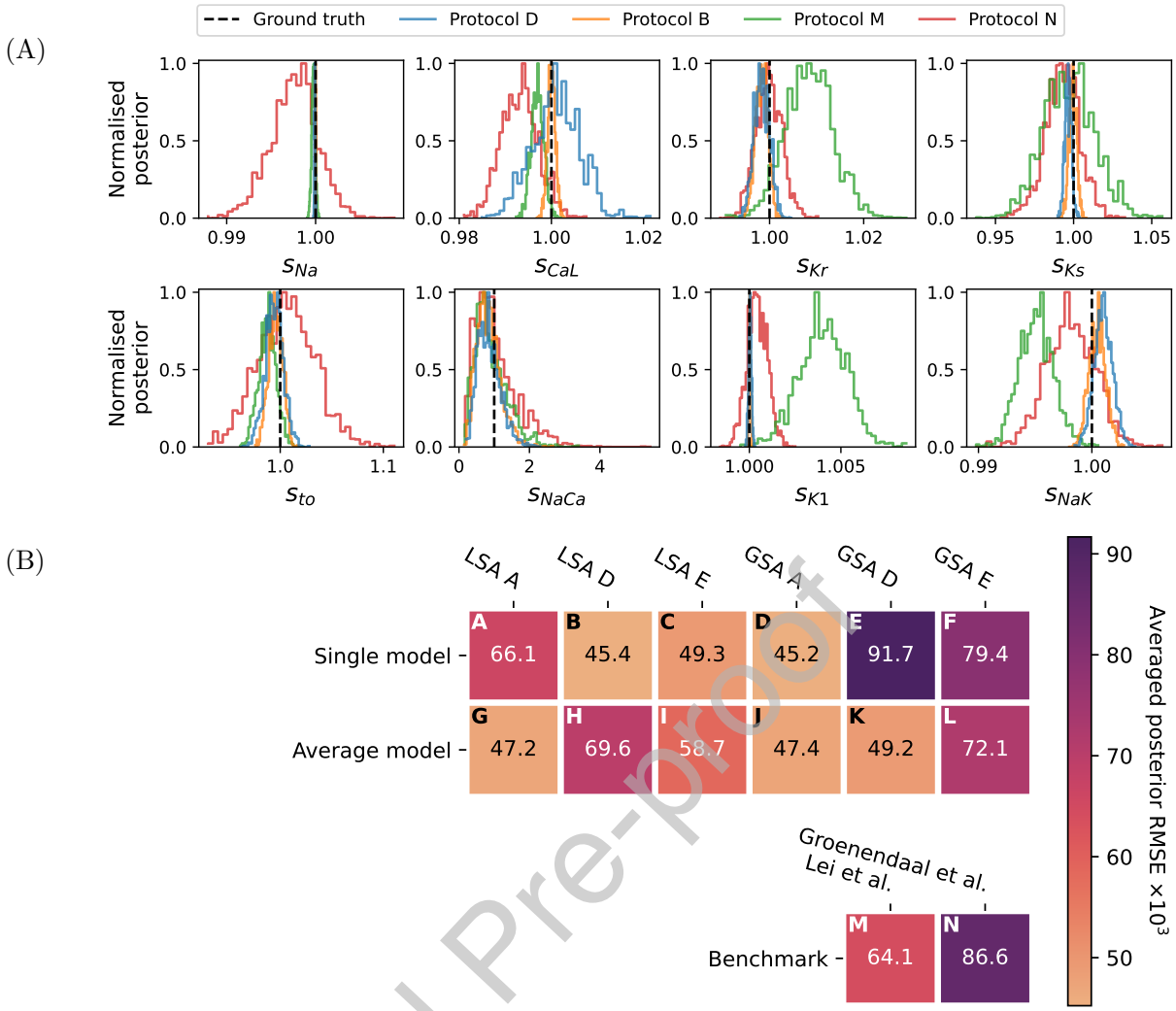


Figure 4: The posterior distribution of the parameters for the synthetic study with voltage clamp, generated with the O'Hara et al. (2011) model and fitted with the same model. **(A)** The marginal posterior for the best two protocols: LSA D- and GSA A-designs for the O'Hara et al. (2011) model (Protocols B and D), according to the mean RMSE of the marginal posterior to the true values, and for the two benchmark protocols. **(B)** The mean RMSE of the marginal posterior for all of the optimised protocols, as well as the benchmark protocols.

Examples of cross-measure assessments for the OEDs are shown in Figure 6A for voltage clamp and Figure 7A for current clamp. Figure 6A shows our best voltage-clamp protocol design, the OED with the 'averaged model score' under GSA D-criterion (Protocol K), which performs well under all measures and model currents. Figure 7A shows our best current-clamp protocol design, the OED with the 'averaged model score' under LSA A-criterion (Protocol R), which performs well under all measures and model currents. The 'best' protocols in this assessment were decided based simply on the mean of all entries in the cross-measure matrix. The averaged score for each optimised protocol is shown in Figure 6C for voltage clamp and Figure 7C for current clamp.

Figure 6C and Figure 7C show that the protocols based on multiple models (bottom row) are similar to those based on only the O'Hara et al. (2011) model (top row), although the best protocols for each experimental mode is based on multiple models. This may be expected as the protocols based on only one model may perform relatively badly under other models. We also noticed that the protocols based on the LSA E criterion are worse than the other LSA criteria. Finally, in Figure 6C, the protocols based on the GSA criteria are better than (or similar to) their LSA counterparts, apart from the E*-design using the O'Hara et al. (2011) model.

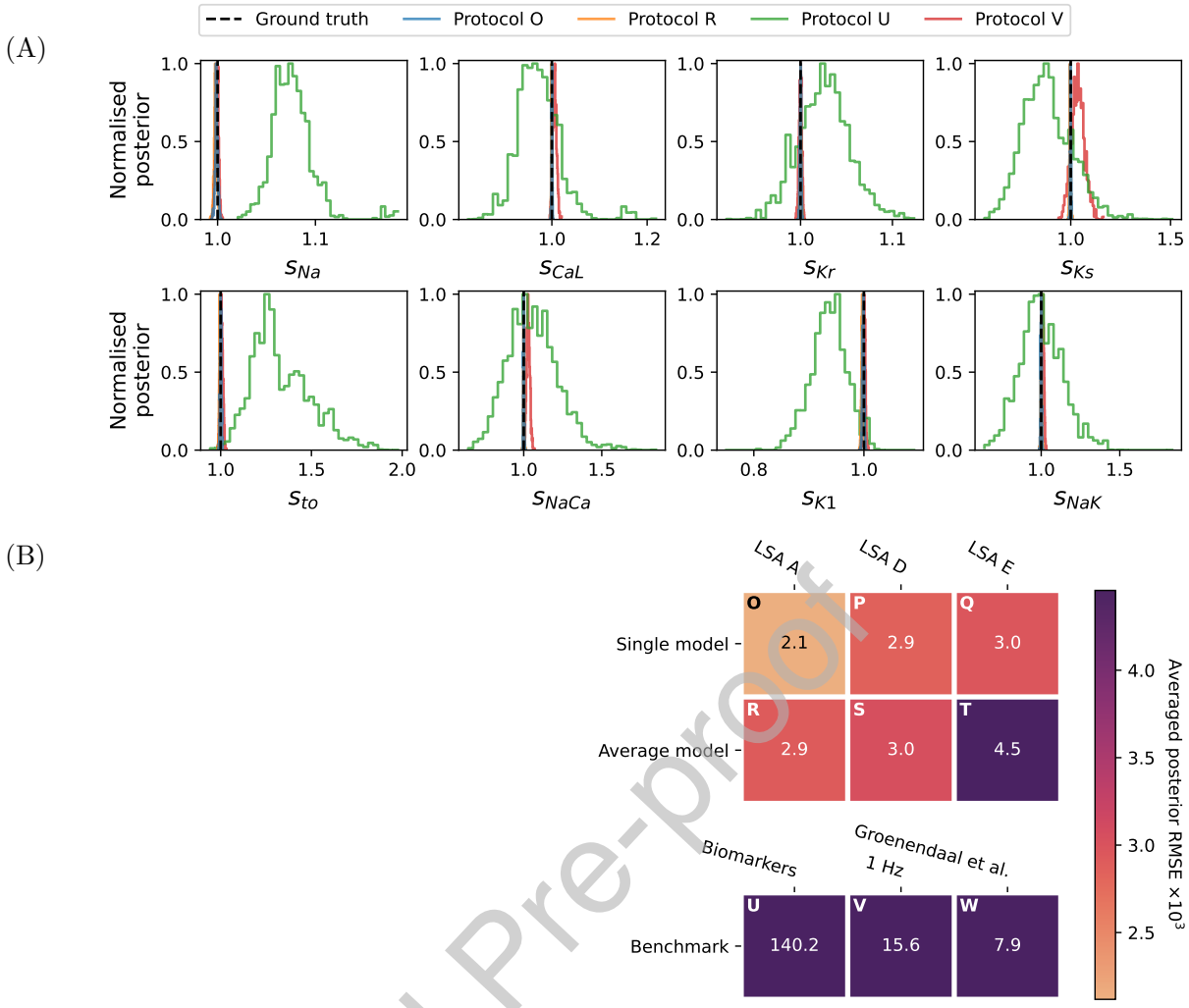


Figure 5: The posterior distribution of the parameters for the synthetic study with current clamp, generated with the O'Hara et al. (2011) model and fitted with the same model. **(A)** The marginal posterior for the best two protocols: LSA A-design (Protocols O and R), according to the mean RMSE of the marginal posterior to the true values, and for the two benchmark protocols. Note that y-axes are rescaled to unity for visualisation purpose. Magnification of this figure around true value is provided in Supplementary Figures. **(B)** The mean RMSE of the marginal posterior for all of the optimised protocols, as well as the benchmark protocols.

3.4 OEDs outperform literature experimental designs

Finally, we assessed the OED results by comparing their performance against some of the designs of experiments in the literature—a benchmark comparison. For voltage-clamp experiments, we used (1) Protocol M: a simple step protocol from Lei et al. (2017) and (2) Protocol N: an expertly-designed protocol from Groenendaal et al. (2015) to benchmark the performance. For current-clamp experiments, we used (1) Protocol W: the random pacing protocol from Groenendaal et al. (2015); (2) Protocol V: a simple 1 Hz constant pacing protocol; and (3) Protocol U: the derived summary statistics (biomarkers) based on 1 Hz pacing, which are commonly used in the literature (Bartolucci et al., 2020; Britton et al., 2013; Jæger et al., 2019, 2020, 2021; Paci et al., 2017, 2018; Tveito et al., 2018).

For the voltage-clamp protocols, the benchmark protocol from Lei et al. (2017) (Protocol M) shown in Figure 6B performed badly for most of the design measures, which may be expected as it is not designed using any of these measures. Similarly, the protocol from Groenendaal et al. (2015) scored badly compared to OED protocols, although its score was, on average, just twice that of the worst OED protocol (Protocol F) which is impressive given its manual design. Figure 4 also shows that the

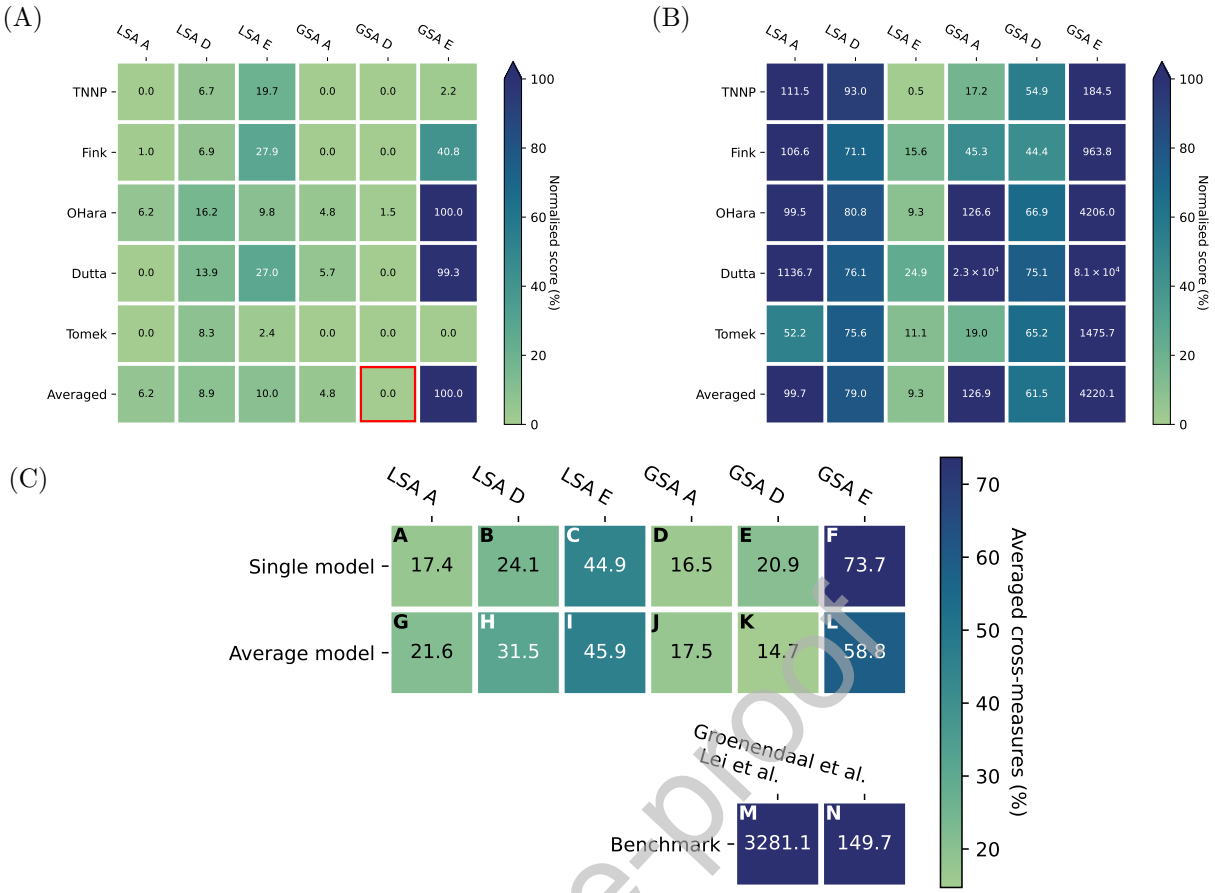


Figure 6: Cross-measure evaluations for the optimised voltage-clamp protocols, compared against the benchmark protocols. **(A)** The normalised cross-measure matrix for the best protocol, Protocol K, optimised using the GSA D-criterion and the averaged model (indicated with the red box). **(B)** The matrix for the worst protocol, Protocol M, from Lei et al. (2017). **(C)** The mean of the normalised cross-measure evaluations for all of the optimised protocols, as well as the benchmark protocols.

benchmark protocols resulted in considerably wider posterior distributions than most of the optimised protocols. Additionally, Supplementary Table S1 shows that Protocol B improves the posterior RMSEs for all of the parameters compared to the benchmark protocols; in Figure 4B, where Protocol E appeared to be the worst, it actually helped identify all the other parameters better than the benchmark protocols apart from s_{NaCa}. Often the number of data samples (or data points) in the i.i.d. Gaussian likelihood that we use is a major factor determining the width of the posterior. Here even for the protocol from Lei et al. (2017) (Protocol M) which is much longer in duration (hence more data samples) than any of the optimised protocols, the width of its posterior is still bigger than most of the optimised protocols. Interestingly, even though the protocol by Groenendaal et al. (2015) (Protocol N) was designed specifically to tease out different dynamics of the ionic currents, the width of the posterior is similar to that of the protocol from Lei et al. (2017). Overall, this shows that the OED methods were successful in reducing the parameter uncertainty for voltage-clamp experiments.

For current-clamp experiments, the biomarkers that we used were action potential duration (APD) at 10 % repolarisation (APD_{10}), APD_{20}, APD_{30}, APD_{40}, APD_{50}, APD_{60}, APD_{70}, APD_{80}, APD_{90}, triangularisation (defined as $\mathrm{APD}_{90} - \mathrm{APD}_{40}$), plateau duration, area under the curve of voltage trace (at APD_{30}), maximum upstroke rate/velocity (MUR), resting potential, maximum voltage (V_{max}), dome peak, and action potential amplitude (APA) which were used in Bartolucci et al. (2020); Britton et al. (2013); Jæger et al. (2019, 2020, 2021); Paci et al. (2017, 2018); Tveito et al. (2018). To perform practical evaluations for biomarkers, we used a percentage difference for calculating the likelihood as it was done similarly in e.g. Tveito et al. (2018); Jæger et al. (2019). We first note that the same

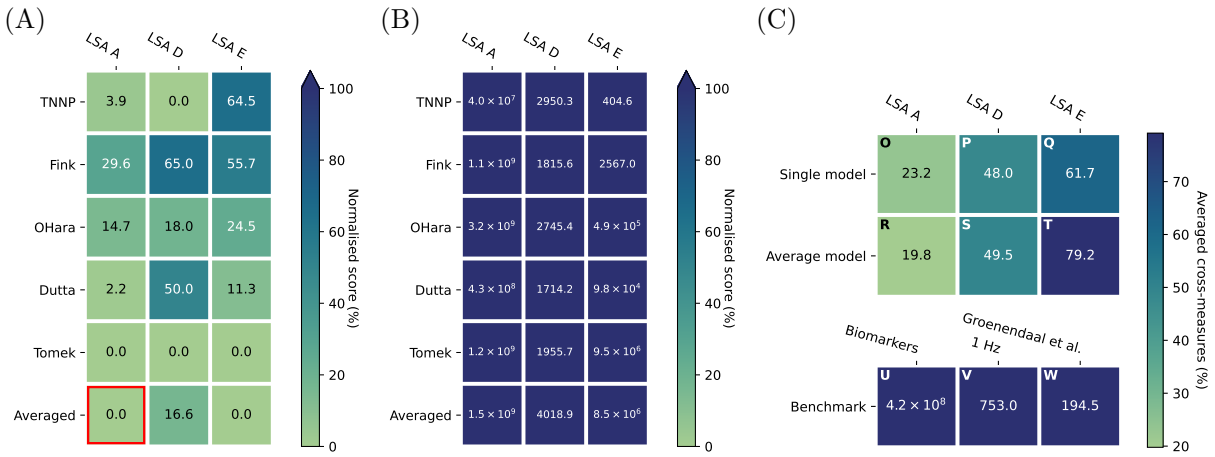


Figure 7: Cross-measure evaluations for the optimised current-clamp protocols, compared against the benchmark protocols. **(A)** The *normalised* cross-measure matrix for the best protocol, Protocol R, optimised using the LSA A-criterion and the averaged model (indicated with the red box). **(B)** The matrix for the worst protocol, Protocol U, biomarker estimation with 1 Hz pacing. **(C)** The mean of the *normalised* cross-measure evaluations for all of the optimised protocols, as well as the benchmark protocols.

was observed for the 1 Hz pacing protocol (Protocol V) and the Groenendaal et al. (2015) protocol (Protocol W): not only is the cross evaluation matrix consistently worse than the OED protocols (Figure 7B), but the posterior obtained in the practical assessment is also wider than all of the OED protocols (Figure 5). The results from using the biomarkers (Protocol U, Figure 5A) were not only more uncertain than the other approaches but also biased in some of the parameter estimates, e.g. I_{Na} and I_{to}. Since Protocol W was designed to improve the information gain by randomly pacing the system, it outperformed Protocols U and V. However, we showed that the OED protocols which maximised the information gain were better in estimating the model parameters than those in the literature, including Protocol W.

4 Discussion

We have applied OED to various AP models to design voltage-clamp and current-clamp protocols for fitting maximum conductances. This is, to our knowledge, the first time such an approach has been taken in cardiac cellular electrophysiology. Some studies have looked into designing experiments for model calibration from *expertly* designed protocols, ranging from calibrating models of individual ionic current dynamics (Fink and Noble, 2009; Zhou et al., 2009; Beattie et al., 2018; Lei et al., 2019b,a) to whole-cell models of APs of cardiomyocytes (Groenendaal et al., 2015); similarly in neuroscience (Hobbs and Hooper, 2008; Tomaiuolo et al., 2012) too. However, these protocols are not truly 'optimal' under a certain statistical or objective measure; they are designed through (subjective) expert views. Our (objectively) optimised protocols successfully reduced the uncertainty in the inferred parameters, compared to some of the conventional protocols.

We explored various designs adapted from the OED literature (Schenkendorf et al., 2018; Gherardini et al., 2021; Seurat et al., 2021), with the aim of designing experiments that help parameter reconstruction to be more accurate and reliable. Most of them were able to improve the uncertainty in the inferred parameters compared to the benchmark protocols, as assessed via our practical evaluations. We tried to seek the best protocol amongst the OED generated protocols. However, we were not able to identify the best one, as each of them perform slightly differently under different situations, although the E*-design appeared to balance the minimum and maximum eigenvalues of FIM by reducing the highest sensitivity which made the resulting protocols relatively short and resulted in one of the worst performances. For example Protocol K was the best in the cross-measures compar-

ison, whilst Protocol D was better in reducing parameter uncertainty and error. We found that the protocols based on averaged model were consistently (slightly) better than the others, which may be a better way for designing patch-clamp protocols, which implicitly makes some allowance for differences in ion current kinetics. Although one could easily extend the presented methodology to create optimal protocols for more parameters, the OED approach does not automatically solve all potential issues, e.g. a priori unidentifiability problem, or computational challenges such as the curse of dimensionality in large-scale parameter inference.

Fitting conductances to whole-cell recordings while using ion channel kinetics models taken (or slightly adapted) from the literature, has been commonly used (Whittaker et al., 2020). Many inference approaches and biomarkers have been adopted, including multivariate regression (Sarkar and Sobie, 2010), history matching (Coveney and Clayton, 2018), "population of models" (Muszkiewicz et al., 2016), and moment-matching (Tixier et al., 2017), and all of them assumed "out of the box" or only slightly modified kinetics models when fitting maximum conductances. Similarly Johnstone et al. (2016) adapted a Bayesian approach with MCMC to infer maximum conductances using AP recordings. Groenendaal et al. (2015) proposed to use (cell-specific) voltage-clamp and current-clamp data with a genetic algorithm to find maximum conductances in an AP model. Some of these methods were included in our benchmark comparison for our approach; we have shown that optimal designs outperform these literature methods in terms of recovering parameters from simulated data.

When inspecting the error (and uncertainty) of the inferred parameters from voltage- and current-clamp experiments (Figures 4 & 5), current clamping appears to be the superior methodology for parameter reconstruction for most of the protocols (but not simply with biomarker-equivalent summary statistics as discussed below). This could be due to the highly nonlinear interaction between all of the currents in current clamping, whilst voltage clamping by definition unties these and turns the system into independent current components. These nonlinear interactions in current-clamp mode may actually help pin down the parameters, because even though some of the currents are small in some parts of the data traces, we may gain information about these currents by the help of the nonlinear interactions with the remaining currents. However, this phenomenon is based on the assumption that there are no model discrepancies. In practice we believe voltage clamping may provide important information by studying each current more 'independently', i.e. less vulnerable to discrepancy in one current affecting inferences about another, therefore we would not advocate to do just one type of experiment but not the other.

Amongst all the methods that we compared, the biomarker approach performed the worst, even though it has been one of the most popular approaches. We further note that some of the biases in the biomarker posteriors were due to the biases introduced in estimating the biomarkers with noisy data. Such an issue is illustrated in Figure 8A which is a *consequence* of using biomarker estimations from noisy AP traces. In Figure 8A, the biomarkers were estimated for 1000 noise realisations and compared against the biomarkers taken from idealised noise-free data (dashed line). For example, there are no simple ways of estimating the unbiased (i.e. neither over- nor under-estimating) maximum voltage or the AP amplitude with noisy AP traces; if we simply define the maximum voltage as the maximum voltage value of the noisy data, as done here, then we will overestimate the actual maximum voltage most of the time. This is because we have a number of data samples (sampling time points) around the underlying maximum voltage where the noise is centred, then the noisy data will have values bigger (as well as smaller) than the mean most of the time, giving an overestimated maximum voltage value. However, if we define the maximum voltage as the averaged value over a (short) period of time around the maximum voltage value of the noisy data, then the maximum voltage will be biased differently depending on the duration of the averaging period and the shape of the AP. Therefore, depending on the data postprocessing procedure, the biomarkers, and hence model parameters inferred from these, can be biased differently.

The biases introduced during biomarker extraction (Figure 8A) therefore affect posterior predictions. Figure 8B shows such posterior predictions using the parameter posteriors shown in Figure 5, where grey and black (dashed) lines show the data and the underlying ground truth, blue and orange show the predictions from the posteriors estimated using Protocol U (biomarkers) and Protocol O (LSA A with single model). The inset of Figure 8B shows a magnification around the apex of the AP,

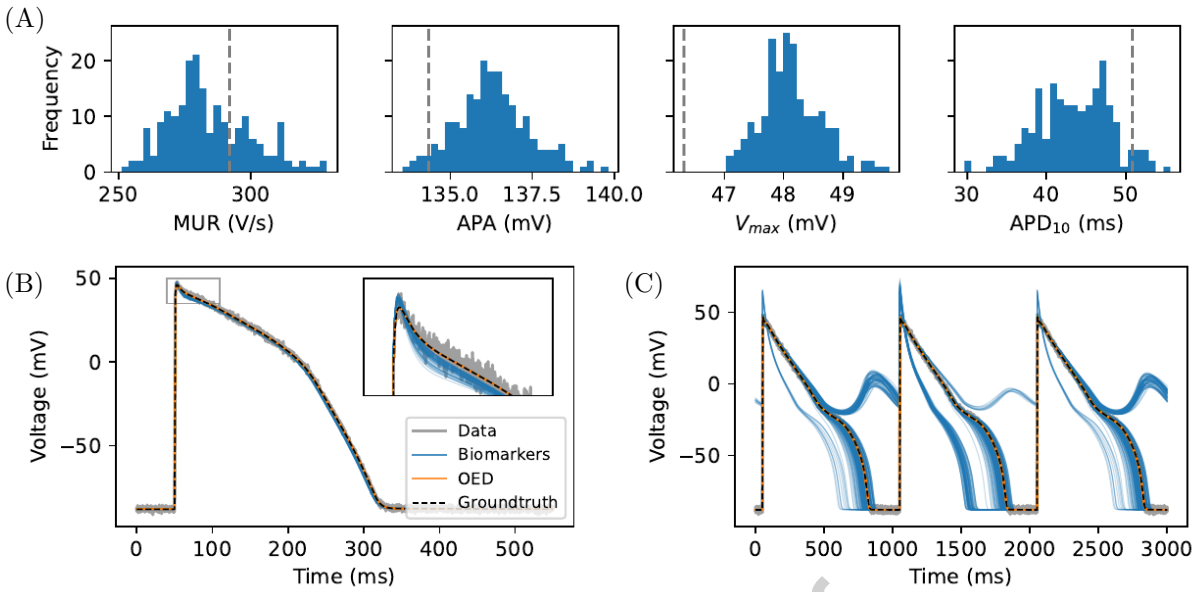


Figure 8: Undesired effects of using biomarkers (Protocol U) for AP parameter inference. **(A)** Various biomarkers were estimated for 1000 noise realisations and compared against the biomarkers estimated without noise (dashed line). **(B)** Posterior predictions using the parameter posterior samples shown in Figure 5 (blue for Protocol U, biomarkers, and orange for Protocol O, OED) for 1 Hz prepaced AP. **(C)** Posterior predictions for 1 Hz prepaced AP under a 90 % reduction of I_{Kr} together with a 50 % reduction of I_{Ks}.

highlighting the biases of the posterior predictions with biomarkers (blue) as compared to the OED results (orange). These biases may in turn cause issues during model predictions for unseen situations. We illustrate these issues in Figure 8C which attempts to predict the AP under a 90 % reduction of I_{Kr} together with a 50 % reduction of I_{Ks}. The huge variability and inaccuracy of the posterior prediction using biomarkers (blue) are consequences of both increased *parameter unidentifiability* as well as bias discussed above, as shown by comparison with the predictions with parameters inferred by fitting to the full time traces from Protocol O (orange).

Several limitations of this work suggest future studies to advance the design of experimental protocols. First, all existing OED approaches, to our knowledge, provide optimal designs only under the assumption of having a correct underlying model, meaning when there is discrepancy between the model we use and the real system (including the remaining parameters of the model that were left out for the design and inference, experimental artefacts (Sakmann and Neher, 1995; Raba et al., 2013; Lei et al., 2020a, 2021) and noise/observation model error (Simoen et al., 2013; Maier et al., 2017; Creswell et al., 2020; Lambert et al., 2023)) then there is no guarantee that the designs will work better, or even work well, as discussed in Lei et al. (2020b); Lei (2020). We believe the presented approach is the first step towards promoting and setting up optimisation of experimental designs in the field. In the future, incorporating the minimisation of the impact of additional uncertainties in the underlying mathematical descriptions as part of the OED measures will be an important next step towards improved future rounds of experimental design. The results, however, have generated experimental protocols that can readily be tested experimentally, especially since these designs can be implemented in most cellular electrophysiology laboratories, and these multiple automatically generated protocols could also potentially be used to quantify predictive uncertainty due to model discrepancy (Shuttleworth et al., 2023). A second limitation is that although we tested 18 OEDs, most of these are 'classical' design criteria that relies on certain approximation of the covariance of the estimates. In future work, we intend to determine whether more complex criteria or approaches, such as Bayesian decision theoretic approach (Huan and Marzouk, 2013; Liepe et al., 2013) and reinforcement approaches (Treloar et al., 2022), can improve the protocol design.

In summary, we have demonstrated that using mechanistic mathematical modelling and optimal experimental design can improve identification of model parameters and hence its theoretical predictive power. This work offers a methodology to automatically, objectively develop experimental protocols that can alleviate the issue of model parameter unidentifiability when collecting new experimental data.

Data Availability

All code and data are freely available under a BSD 3-clause open source licence at:
https://github.com/chonlei/action-potential-experimental-design
[to be archived on Zenodo and assigned a DOI upon acceptance].

Statements of ethical approval

This study did not include any human or animal subjects.

Conflict of Interest Statement

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.

Acknowledgments

This work was performed in part at the High Performance Computing Cluster supported by the Information and Communication Technology Office of the University of Macau.

Funding

This work was funded by the Science and Technology Development Fund, Macao S.A.R. (FDCT) [reference number 0048/2022/A]; the UK Engineering and Physical Sciences Research Council [grant number EP/S024093/1]; and the Wellcome Trust [grant number 212203/Z/18/Z]. CLL acknowledges support from the FDCT, Macao S.A.R. and support from the University of Macau via a UM Macao Fellowship. MC and GRM acknowledge support from the Wellcome Trust via a Senior Research Fellowship to GRM. DJG acknowledges support from the UK Engineering and Physical Sciences Research Council for Doctoral Training Programme.

This research was funded in whole, or in part, by the Wellcome Trust [212203/Z/18/Z]. For the purpose of open access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author Contributions

Conceptualization: CLL, MC, DJG, GRM. Methodology: CLL, GRM. Investigation: CLL. Writing—original draft: CLL, GRM. Writing—review: CLL, MC, DJG, GRM.

References

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Oxford University Press, Oxford.

Banga, J. R. and Balsa-Canto, E. (2008). Parameter estimation and optimal experimental design. *Essays in Biochemistry*, 45:195–209.

Bartolucci, C., Passini, E., Hyttinen, J., Paci, M., and Severi, S. (2020). Simulation of the effects of extracellular calcium changes leads to a novel computational model of human ventricular action potential with a revised calcium handling. *Frontiers in Physiology*, 11:314.

Beattie, K. A., Hill, A. P., Bardenet, R., Cui, Y., Vandenberg, J. I., Gavaghan, D. J., De Boer, T. P., and Mirams, G. R. (2018). Sinusoidal voltage protocols for rapid characterisation of ion channel kinetics. *The Journal of physiology*, 596(10):1813–1828.

Britton, O. J., Bueno-Orovio, A., Van Ammel, K., Lu, H. R., Towart, R., Gallacher, D. J., and Rodriguez, B. (2013). Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proceedings of the National Academy of Sciences*, 110(23):E2098–E2105.

Chang, K. C., Dutta, S., Mirams, G. R., Beattie, K. A., Sheng, J., Tran, P. N., Wu, M., Wu, W. W., Colatsky, T., Strauss, D. G., et al. (2017). Uncertainty quantification reveals the importance of data variability and experimental design considerations for in silico proarrhythmia risk assessment. *Frontiers in Physiology*, 8:917.

Chu, Y. and Hahn, J. (2013). Necessary condition for applying experimental design criteria to global sensitivity analysis results. *Computers and Chemical Engineering*, 48:280–292.

Clerx, M., Beattie, K. A., Gavaghan, D. J., and Mirams, G. R. (2019a). Four ways to fit an ion channel model. *Biophysical Journal*, 117(12):2420–2437.

Clerx, M., Collins, P., de Lange, E., and Volders, P. G. A. (2016). Myokit: A simple interface to cardiac cellular electrophysiology. *Progress in Biophysics & Molecular Biology*, 120(1–3):100–114.

Clerx, M., Robinson, M., Lambert, B., Lei, C. L., Ghosh, S., Mirams, G. R., and Gavaghan, D. J. (2019b). Probabilistic inference on noisy time series (PINTS). *Journal of Open Research Software*, 7(1):23.

Corral-Acero, J., Margara, F., Marciniak, M., Rodero, C., Loncaric, F., Feng, Y., Gilbert, A., Fernandes, J. F., Bukhari, H. A., Wajdan, A., et al. (2020). The âĂŸdigital twinâĂŹto enable the vision of precision cardiology. *European Heart Journal*, 41(48):4556–4564.

Coveney, S. and Clayton, R. H. (2018). Fitting two human atrial cell models to experimental data using Bayesian history matching. *Progress in Biophysics and Molecular Biology*, 139:43–58.

Creswell, R., Lambert, B., Lei, C. L., Robinson, M., and Gavaghan, D. (2020). Using flexible noise models to avoid noise model misspecification in inference of differential equation time series models. *arXiv preprint arXiv:2011.04854*.

Dutta, S., Chang, K. C., Beattie, K. A., Sheng, J., Tran, P. N., Wu, W. W., Wu, M., Strauss, D. G., Colatsky, T., and Li, Z. (2017). Optimization of an in silico cardiac cell model for proarrhythmia risk assessment. *Frontiers in Physiology*, 8:616.

Fink, M. and Noble, D. (2009). Markov models for ion channels: versatility versus identifiability and speed. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 367(1896):2161–79.

Fink, M., Noble, D., Virag, L., Varro, A., and Giles, W. R. (2008). Contributions of hERG K+ current to repolarization of the human ventricular action potential. *Progress in Biophysics and Molecular Biology*, 96(1-3):357–376.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Texts in Statistical Science. Chapman and Hall/CRC, Boca Raton, FL, 3rd edition.

Gherardini, M., Mannini, A., and Cipriani, C. (2021). Optimal spatial sensor design for magnetic tracking in a myokinetic control interface. *Computer Methods and Programs in Biomedicine*, 211:106407.

Gottu Mukkula, A. R. and Paulen, R. (2017). Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation. *Computers and Chemical Engineering*, 99:198–213.

Groenendaal, W., Ortega, F. A., Kherlopian, A. R., Zygmunt, A. C., Krogh-Madsen, T., and Christini, D. J. (2015). Cell-specific cardiac electrophysiology models. *PLoS Computational Biology*, 11(4):e1004242.

Gupta, D. K. and Dhingra, A. K. (2013). Input load identification from optimally placed strain gages using d-optimal design and model reduction. *Mechanical Systems and Signal Processing*, 40(2):556–570.

Hamill, O. P., Marty, A., Neher, E., Sakmann, B., and Sigworth, F. J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv*, 391(2):85–100.

Hansen, N. (2006). The CMA Evolution Strategy: a comparing review. In Lozano, J. A., Larrañaga, P., Inza, I., and Bengoetxea, E., editors, *Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms*, pages 75–102. Springer-Verlag, Heidelberg.

Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. E., and Woodward, C. S. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, 31(3):363–396.

Hobbs, K. H. and Hooper, S. L. (2008). Using complicated, wide dynamic range driving to develop models of single neurons in single recording sessions. *Journal of Neurophysiology*, 99(4):1871–1883.

Huan, X. and Marzouk, Y. M. (2013). Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317.

Jæger, K. H., Charwat, V., Charrez, B., Finsberg, H., Maleckar, M. M., Wall, S., Healy, K. E., and Tveito, A. (2020). Improved computational identification of drug response using optical measurements of human stem cell derived cardiomyocytes in microphysiological systems. *Frontiers in Pharmacology*, 10:1648.

Jæger, K. H., Charwat, V., Wall, S., Healy, K. E., and Tveito, A. (2021). Identifying drug response by combining measurements of the membrane potential, the cytosolic calcium concentration, and the extracellular potential in microphysiological systems. *Frontiers in Pharmacology*, page 2085.

Jæger, K. H., Wall, S., and Tveito, A. (2019). Detecting undetectables: Can conductances of action potential models be changed without appreciable change in the transmembrane potential? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(7):073102.

Jasra, A., Stephens, D. A., and Holmes, C. C. (2007). On population-based simulation for static inference. *Statistics and Computing*, 17(3):263–279.

Johnstone, R. H., Chang, E. T. Y., Bardenet, R., De Boer, T. P., Gavaghan, D. J., Pathmanathan, P., Clayton, R. H., and Mirams, G. R. (2016). Uncertainty and variability in models of the cardiac action potential: can we build trustworthy models? *Journal of Molecular and Cellular Cardiology*, 96:49–62.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society: Series B*, 21(2):272–304.

Kucerová, A., Sykora, J., Janouchová, E., Jarušková, D., and Chleboun, J. (2016). Acceleration of robust experiment design using Sobol indices and polynomial chaos expansion. In *Proceedings of the 7th International Workshop on Reliable Engineering Computing (REC), Bochum, Germany*, pages 15–17.

Lambert, B., Lei, C. L., Robinson, M., Clerx, M., Creswell, R., Ghosh, S., Tavener, S., and Gavaghan, D. J. (2023). Autocorrelated measurement processes and inference for ordinary differential equation models of biological systems. *Journal of The Royal Society Interface*, 20(199):20220725.

Lei, C. L. (2020). *Model-Driven Design and Uncertainty Quantification for Cardiac Electrophysiology Experiments*. PhD thesis, University of Oxford.

Lei, C. L., Clerx, M., Beattie, K. A., Melgari, D., Hancox, J. C., Gavaghan, D. J., Polonchuk, L., Wang, K., and Mirams, G. R. (2019a). Rapid characterisation of hERG channel kinetics II: temperature dependence. *Biophysical Journal*, 117:2455–2470.

Lei, C. L., Clerx, M., Gavaghan, D. J., Polonchuk, L., Mirams, G. R., and Wang, K. (2019b). Rapid characterisation of hERG channel kinetics I: using an automated high-throughput system. *Biophysical Journal*, 117:2438–2454.

Lei, C. L., Clerx, M., Whittaker, D. G., Gavaghan, D. J., De Boer, T. P., and Mirams, G. R. (2020a). Accounting for variability in ion current recordings using a mathematical model of artefacts in voltage-clamp experiments. *Philosophical Transactions of the Royal Society A*, 378(2173):20190348.

Lei, C. L., Fabbri, A., Whittaker, D. G., Clerx, M., Windley, M. J., Hill, A. P., Mirams, G. R., and de Boer, T. P. (2021). A nonlinear and time-dependent leak current in the presence of calcium fluoride patch-clamp seal enhancer [version 2; peer review: 4 approved]. *Wellcome Open Research*, 5:152.

Lei, C. L., Ghosh, S., Whittaker, D. G., Aboelkassem, Y., Beattie, K. A., Cantwell, C. D., Delhaas, T., Houston, C., Novaes, G. M., Panfilov, A. V., et al. (2020b). Considering discrepancy when calibrating a mechanistic electrophysiology model. *Philosophical Transactions of the Royal Society A*, 378(2173):20190349.

Lei, C. L., Wang, K., Clerx, M., Johnstone, R. H., Hortigon-Vinagre, M. P., Zamora, V., Allan, A., Smith, G. L., Gavaghan, D. J., Mirams, G. R., and Polonchuk, L. (2017). Tailoring mathematical models to stem-cell derived cardiomyocyte lines can improve predictions of drug-induced changes to their electrophysiology. *Frontiers in Physiology*, 8.

Li, Z., Ridder, B. J., Han, X., Wu, W. W., Sheng, J., Tran, P. N., Wu, M., Randolph, A., Johnstone, R. H., Mirams, G. R., et al. (2019). Assessment of an in silico mechanistic model for proarrhythmia risk prediction under the CiPA initiative. *Clinical Pharmacology & Therapeutics*, 105(2):466–475.

Liepe, J., Filippi, S., Komorowski, M., and Stumpf, M. P. (2013). Maximizing the information content of experiments in systems biology. *PLoS Computational Biology*, 9(1):e1002888.

Lindley, D. V. et al. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005.

Maier, C., Loos, C., and Hasenauer, J. (2017). Robust parameter estimation for dynamical systems from outlier-corrupted data. *Bioinformatics*, 33(5):718–725.

Mirams, G. R., Davies, M. R., Cui, Y., Kohl, P., and Noble, D. (2012). Application of cardiac electrophysiology simulations to pro-arrhythmic safety testing. *British Journal of Pharmacology*, 167(5):932–945.

Muszkiewicz, A., Britton, O. J., Gemmell, P., Passini, E., Sánchez, C., Zhou, X., Carusi, A., Quinn, T. A., Burrage, K., Bueno-Orovio, A., and Rodriguez, B. (2016). Variability in cardiac electrophysiology: using experimentally-calibrated populations of models to move beyond the single virtual physiological human paradigm. *Progress in Biophysics and Molecular Biology*, 120(1-3):115–127.

Niederer, S., Aboelkassem, Y., Cantwell, C. D., Corrado, C., Coveney, S., Cherry, E. M., Delhaas, T., Fenton, F. H., Panfilov, A., Pathmanathan, P., et al. (2020). Creation and application of virtual patient cohorts of heart models. *Philosophical Transactions of the Royal Society A*, 378(2173):20190558.

Niederer, S. A., Lumens, J., and Trayanova, N. A. (2018). Computational models in cardiology. *Nature Reviews Cardiology*, 16:100–111.

O'Hara, T., Virág, L., Varró, A., and Rudy, Y. (2011). Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. *PLoS Computational Biology*, 7(5):e1002061.

Paci, M., Passini, E., Severi, S., Hyttinen, J., and Rodriguez, B. (2017). Phenotypic variability in lqt3 human induced pluripotent stem cell-derived cardiomyocytes and their response to antiarrhythmic pharmacologic therapy: an in silico approach. *Heart Rhythm*, 14(11):1704–1712.

Paci, M., Pölönen, R.-P., Cori, D., Penttinen, K., Aalto-Setälä, K., Severi, S., and Hyttinen, J. (2018). Automatic optimization of an in silico model of human ipsc derived cardiomyocytes recapitulating calcium handling abnormalities. *Frontiers in Physiology*, 9:709.

Pant, S. (2018). Information sensitivity functions to assess parameter information gain and identifiability of dynamical systems. *Journal of The Royal Society Interface*, 15(142):20170871.

Plank, G., Loewe, A., Neic, A., Augustin, C., Huang, Y.-L., Gsell, M. A., Karabelas, E., Nothstein, M., Prassl, A. J., SÃąnchez, J., Seemann, G., and Vigmond, E. J. (2021). The opencarp simulation environment for cardiac electrophysiology. *Computer Methods and Programs in Biomedicine*, 208:106223.

Raba, A. E., Cordeiro, J. M., Antzelevitch, C., and Beaumont, J. (2013). Extending the conditions of application of an inversion of the hodgkin–huxley gating model. *Bulletin of mathematical biology*, 75(5):752–773.

Rodriguez-Fernandez, M., Kucherenko, S., Pantelides, C., and Shah, N. (2007). Optimal experimental design based on global sensitivity analysis. In *Computer Aided Chemical Engineering*, volume 24, pages 63–68. Elsevier.

Sakmann, B. and Neher, E., editors (1995). *Single-Channel Recording*. Springer US, 2 edition.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270.

Sarkar, A. X. and Sobie, E. A. (2010). Regression analysis for constraining free parameters in electrophysiological models of cardiac cells. *PLoS Computational Biology*, 6(9):e1000914.

Schenkendorf, R., Xie, X., Rehbein, M., Scholl, S., and Krewer, U. (2018). The impact of global sensitivities and design measures in model-based optimal experimental design. *Processes*, 6(4):27.

Seidler, R., Padalkina, K., Buecker, H. M., Ebigbo, A., Herty, M., Marquart, G., and Niederau, J. (2016). Optimal experimental design for reservoir property estimates in geothermal exploration. *Computational Geosciences*, 20(2):375–383.

Seurat, J., Tang, Y., MentrÃĺ, F., and Nguyen, T. T. (2021). Finding optimal design in nonlinear mixed effect models using multiplicative algorithms. *Computer Methods and Programs in Biomedicine*, 207:106126.

Shuttleworth, J. G., Lei, C. L., Whittaker, D. G., Windley, M. J., Hill, A. P., Preston, S. P., and Mirams, G. R. (2023). Empirical quantification of predictive uncertainty due to model discrepancy by training with an ensemble of experimental designs: an application to ion channel kinetics. *arXiv preprint arXiv:2302.02942*.

Simoen, E., Papadimitriou, C., and Lombaert, G. (2013). On prediction error correlation in bayesian model updating. *Journal of Sound and Vibration*, 332(18):4136–4152.

Smucker, B., Krzywinski, M., and Altman, N. (2018). Optimal experimental design. *Nat. Methods*, 15(8):559–560.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280.

ten Tusscher, K. H., Noble, D., Noble, P.-J., and Panfilov, A. V. (2004). A model for human ventricular tissue. *American Journal of Physiology-Heart and Circulatory Physiology*, 286(4):H1573–H1589.

Tixier, E., Lombardi, D., Rodriguez, B., and Gerbeau, J. F. (2017). Modelling variability in cardiac electrophysiology: a moment-matching approach. *Journal of the Royal Society Interface*, 14(133):20170238.

Tomaiuolo, M., Bertram, R., Leng, G., and Tabak, J. (2012). Models of electrical activity: calibration and prediction testing on the same cell. *Biophysical Journal*, 103(9):2021–2032.

Tomek, J., Bueno-Orovio, A., Passini, E., Zhou, X., Minchole, A., Britton, O., Bartolucci, C., Severi, S., Shrier, A., Virag, L., et al. (2019). Development, calibration, and validation of a novel human ventricular myocyte model in health, disease, and drug block. *eLife*, 8:e48890.

Treloar, N. J., Braniff, N., Ingalls, B., and Barnes, C. P. (2022). Deep reinforcement learning for optimal experimental design in biology. *PLoS Computational Biology*, 18(11):1–24.

Tveito, A., Jæger, K. H., Huebsch, N., Charrez, B., Edwards, A. G., Wall, S., and Healy, K. E. (2018). Inversion and computational maturation of drug response using human stem cell derived cardiomyocytes in microphysiological systems. *Scientific Reports*, 8(1):1–14.

Ushijima, T. T., Yeh, W. W., and Wong, W. K. (2021). Constructing robust and efficient experimental designs in groundwater modeling using a galerkin method, proper orthogonal decomposition, and metaheuristic algorithms. *PLoS One*, 16(8):e0254620.

Vanrolleghem, P. A. and Dochain, D. (1998). Bioprocess model identification. In Van Impe J.F.M., Vanrolleghem P.A., and Iserentant D.M., editors, *Advanced Instrumentation, Data Interpretation, and Control of Biotechnological Processes*, chapter 10, pages 251–318. Springer, Dordrecht.

Walter, E. and Pronzato, L. (1997). *Identification of Parametric Models: From Experimental Data*. Communications and Control Engineering. Springer, London.

Whittaker, D. G., Clerx, M., Lei, C. L., Christini, D. J., and Mirams, G. R. (2020). Calibration of ionic and cellular cardiac electrophysiology models. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 12(4):e1482.

Zhou, Q., Zygmunt, A. C., Cordeiro, J. M., Siso-Nadal, F., Miller, R. E., Buzzard, G. T., and Fox, J. J. (2009). Identification of i kr kinetics and drug binding in native myocytes. *Annals of Biomedical Engineering*, 37(7):1294–1309.

S1 Supplementary Figures

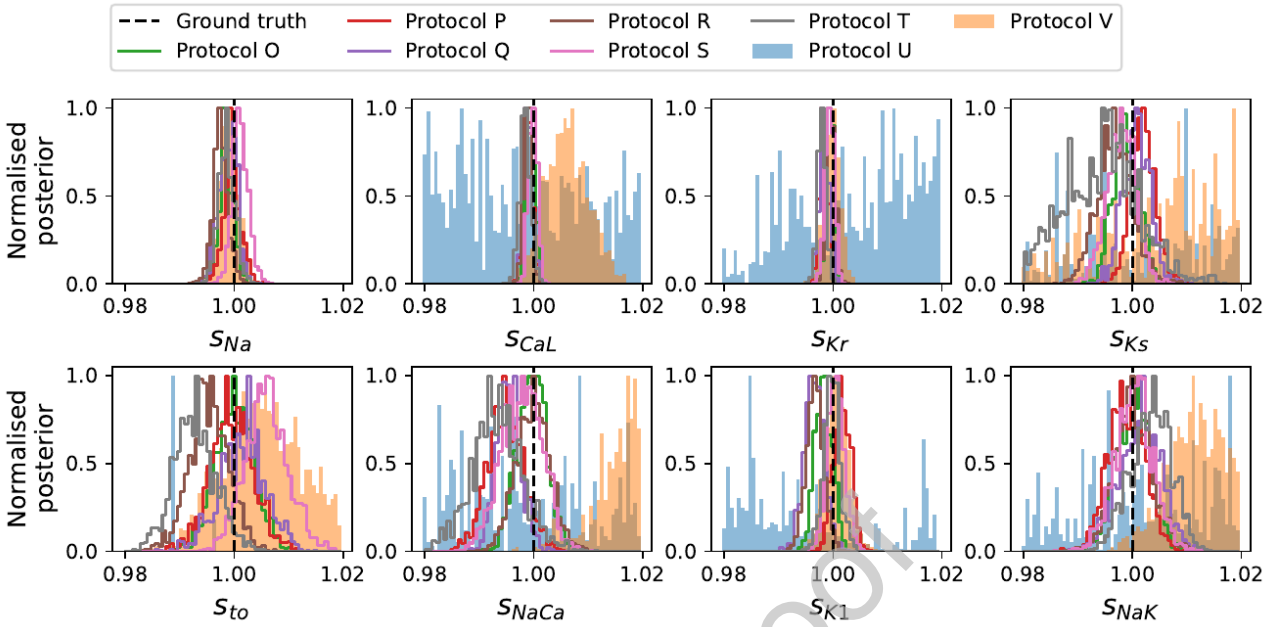


Figure S1: The marginal posterior distribution of the parameters for the synthetic study with current clamp, generated with the O'Hara et al. (2011) model and fitted with the same model. A magnification version of Figure 5.

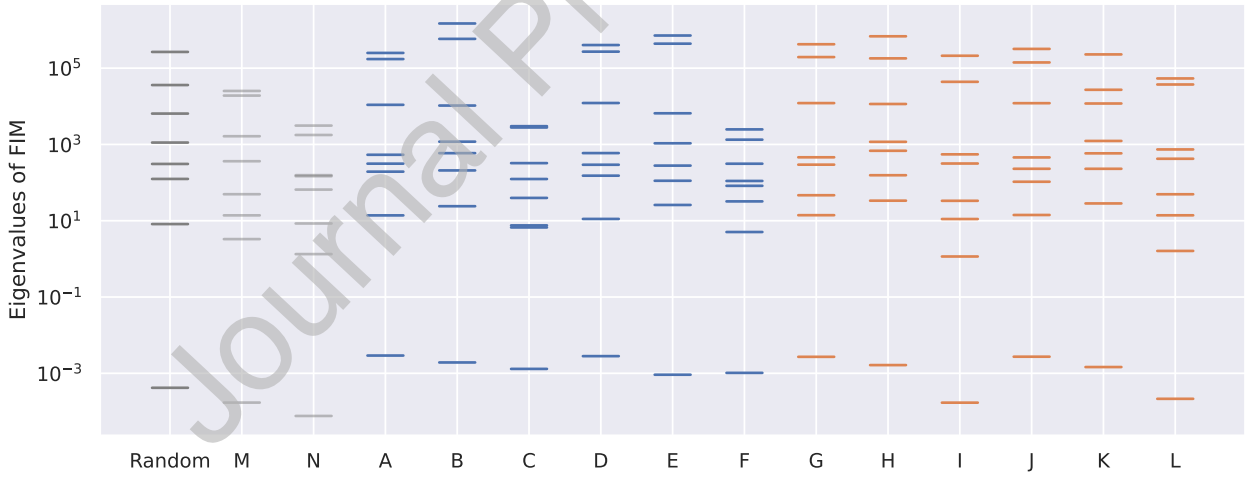


Figure S2: The eigenvalue spectra of the Fisher information matrix (FIM) for the voltage-clamp protocols presented in the main text.

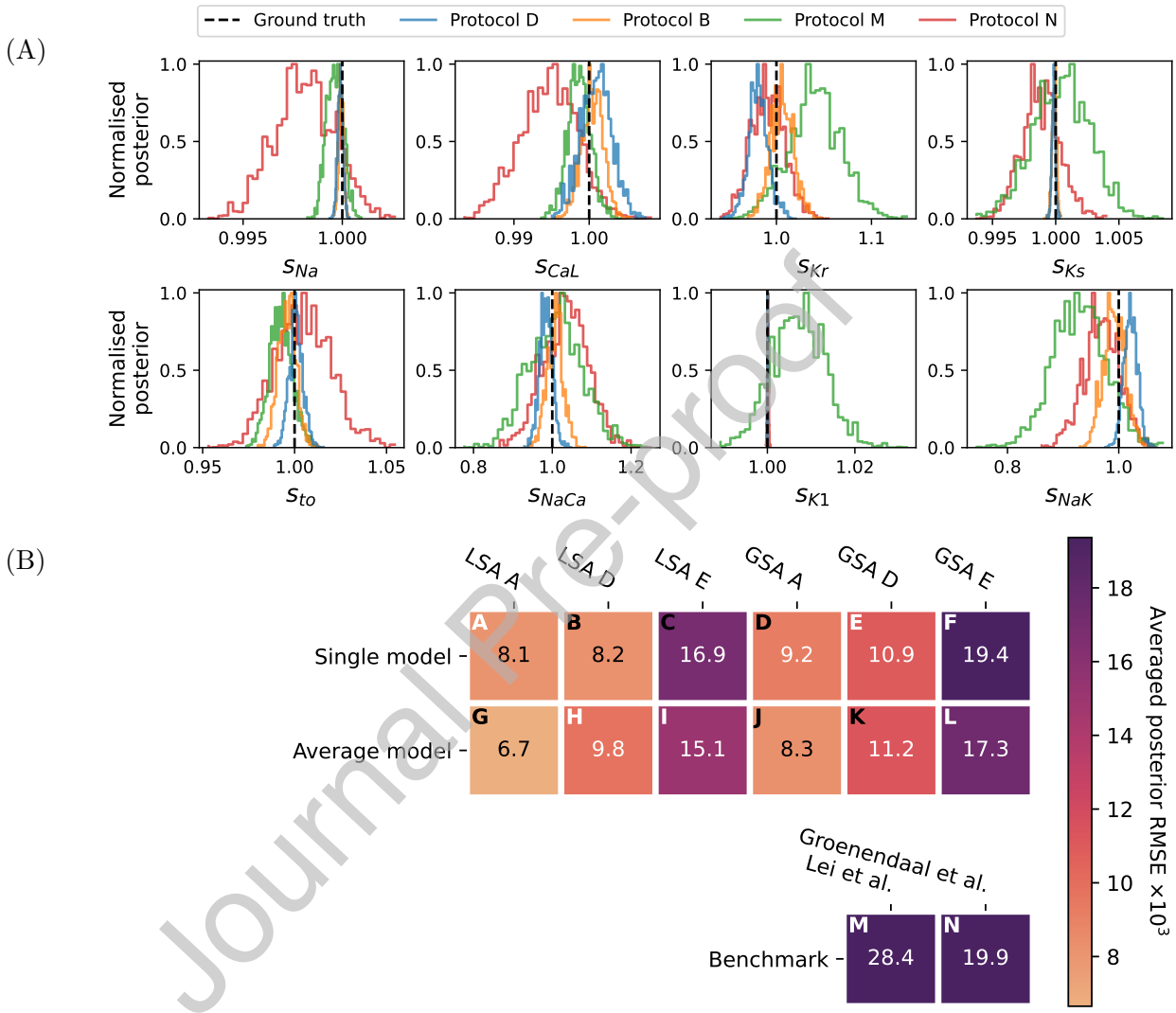


Figure S3: The posterior distribution of the parameters for the synthetic study with voltage clamp, generated with the ten Tusscher et al. (2004) model and fitted with the same model. **(A)** The marginal posterior for LSA D- and GSA A-designs for the O'Hara et al. (2011) model (Protocols B and D), and for the two benchmark protocols. **(B)** The mean RMSE of the marginal posterior for all of the optimised protocols, as well as the benchmark protocols.

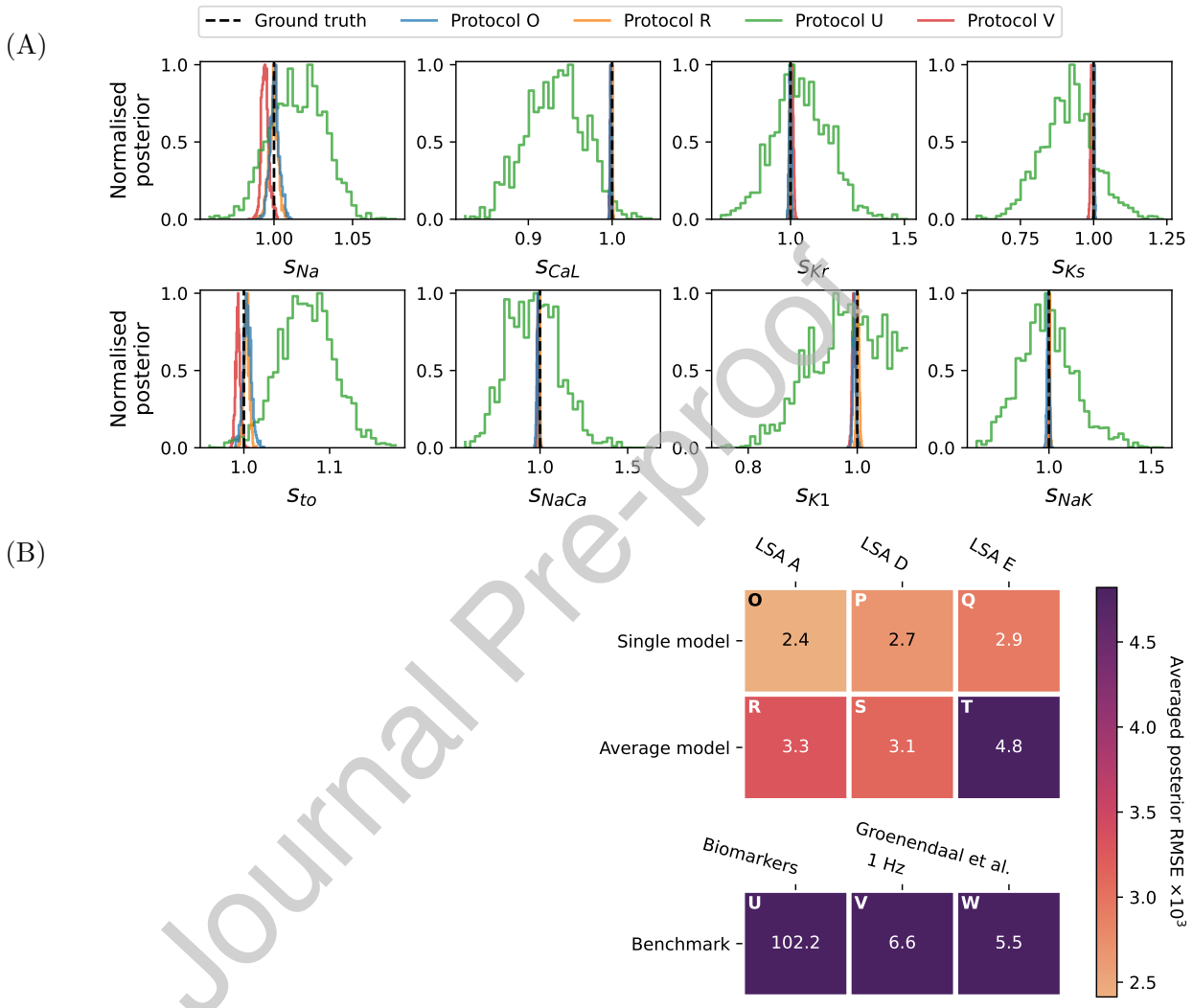


Figure S4: The posterior distribution of the parameters for the synthetic study with current clamp, generated with the ten Tusscher et al. (2004) model and fitted with the same model. **(A)** The marginal posterior for the LSA A-design (Protocols O and R), and for the two benchmark protocols. Note that y-axes are rescaled to unity for visualisation purpose. **(B)** The mean RMSE of the marginal posterior for all of the optimised protocols, as well as the benchmark protocols.

S2 Supplementary Tables

Protocols	s_{Na}	s_{CaL}	s_{Kr}	s_{Ks}	s_{to}	s_{NaCa}	s_{K1}	s_{NaK}	Mean
A	0.06	2.7	1.8	3.3	9.1	511.1	0.17	0.8	66.1
B	**0.03**	**0.9**	**2.0**	**2.4**	**7.4**	**349.5**	**0.06**	**0.9**	**45.4**
C	6.01	3.6	9.0	9.7	6.6	356.6	0.97	2.1	49.3
D	**0.09**	**5.6**	**2.1**	**3.8**	**9.6**	**339.3**	**0.10**	**1.4**	**45.2**
E	0.05	1.5	1.5	3.1	5.8	719.8	0.13	1.5	91.7
F	0.07	2.2	6.3	43.1	11.5	568.5	0.21	3.5	79.4
G	0.07	2.6	1.6	2.9	11.9	357.8	0.07	0.7	47.2
H	0.04	1.1	3.8	4.3	8.8	536.1	0.07	2.3	69.6
I	1.17	3.0	9.6	13.2	7.7	431.7	0.99	1.8	58.7
J	0.10	7.7	2.0	2.2	9.2	356.9	0.08	0.8	47.4
K	0.21	3.1	1.7	2.3	7.1	377.9	0.09	0.9	49.2
L	0.11	2.0	7.8	24.6	9.8	529.5	0.18	2.6	72.1
M	0.28	3.1	10.7	16.5	13.7	459.4	4.46	4.9	64.1
N	3.27	7.4	2.8	12.1	29.9	634.0	0.77	2.6	86.6

Table S1: Table of model parameter posterior RMSEs to the true values for different voltage-clamp designs, corresponding to the values shown in Figure 4. The last column shows the mean of the posterior RMSEs which is equivalent to the values shown in Figure 4B. Highlighted are the best on average optimised protocols (bold) and the benchmark protocols (grey). All values are $\times 10^3$ for ease of reading.

Protocols	s_{Na}	s_{CaL}	s_{Kr}	s_{Ks}	s_{to}	s_{NaCa}	s_{K1}	s_{NaK}	Mean
O	**1.57**	**0.8**	**0.8**	**2.4**	**3.6**	**2.7**	**1.87**	**3.2**	**2.1**
P	1.62	0.9	1.3	3.4	3.9	5.6	2.80	3.3	2.9
Q	1.70	0.9	1.3	2.9	4.8	4.5	3.28	4.4	3.0
R	**2.34**	**0.8**	**0.8**	**4.1**	**5.4**	**3.2**	**3.10**	**3.5**	**2.9**
S	2.32	1.0	0.8	3.0	7.7	4.0	1.98	3.6	3.0
T	1.64	1.3	1.6	8.7	7.6	7.7	1.57	5.6	4.5
U	83.42	49.8	41.9	153.8	397.2	182.9	70.59	141.8	140.2
V	1.74	7.7	1.5	53.9	9.6	33.0	2.43	14.6	15.6
W	2.4	2.92	1.80	13.0	8.3	6.3	6.2	22.2	7.9

Table S2: Table of model parameter posterior RMSEs to the true values for different current-clamp designs, corresponding to the values shown in Figure 5. The last column shows the mean of the posterior RMSEs which is equivalent to the values shown in Figure 5B. Highlighted are the best on average optimised protocols (bold) and the benchmark protocols (grey). All values are $\times 10^3$ for ease of reading.

Protocols	s_{Na}	s_{CaL}	s_{Kr}	s_{Ks}	s_{to}	s_{NaCa}	s_{K1}	s_{NaK}	Mean
A	0.15	2.8	13.9	0.2	4.9	23.1	0.04	19.4	8.1
B	0.13	1.6	14.4	0.2	5.8	23.6	0.02	20.0	8.2
C	36.66	9.2	25.8	1.2	6.7	31.8	0.28	24.0	16.9
D	0.15	2.4	19.7	0.2	4.1	20.5	0.03	26.7	9.2
E	0.21	1.1	12.3	0.1	5.4	33.3	0.04	34.9	10.9
F	0.20	2.6	23.3	6.4	11.9	59.9	0.05	50.6	19.4
G	0.17	1.6	13.1	0.2	6.1	18.6	0.02	13.5	6.7
H	0.15	1.6	19.2	0.2	4.9	25.3	0.02	26.9	9.8
I	3.68	4.8	51.8	0.7	10.0	24.7	0.50	24.6	15.1
J	0.13	5.8	18.1	0.2	7.1	21.5	0.02	13.4	8.3
K	0.40	1.7	14.1	0.3	6.3	44.3	0.03	22.8	11.2
L	0.21	2.4	25.1	7.3	7.3	52.6	0.06	43.6	17.3
M	0.52	1.9	53.2	2.4	8.9	71.2	10.26	78.5	28.4
N	2.25	6.1	18.5	1.7	17.0	72.2	0.21	41.4	19.9

Table S3: Table of model parameter posterior RMSEs to the true values for different voltage-clamp designs, corresponding to the values shown in Figure S3. The last column shows the mean of the posterior RMSEs which is equivalent to the values shown in Figure S3B. Highlighted are the benchmark protocols. All values are $\times 10^3$ for ease of reading.

Protocols	s_{Na}	s_{CaL}	s_{Kr}	s_{Ks}	s_{to}	s_{NaCa}	s_{K1}	s_{NaK}	Mean
O	1.8	0.74	3.46	1.5	2.5	4.4	3.2	1.8	2.4
P	2.1	0.71	3.70	1.8	2.5	6.0	2.6	2.0	2.7
Q	2.9	0.59	3.98	0.9	4.9	2.9	2.9	4.3	2.9
R	3.5	0.72	2.91	1.4	3.9	8.4	3.3	1.9	3.3
S	2.1	0.60	3.09	1.5	2.6	6.3	3.4	5.5	3.1
T	2.5	1.24	10.38	5.3	3.3	9.8	3.3	2.7	4.8
U	24.9	71.25	139.97	115.6	85.3	149.3	93.7	137.6	102.2
V	5.7	1.80	11.02	7.6	6.8	10.1	6.5	2.9	6.6
W	3.1	1.80	4.50	2.7	6.8	14.2	5.0	6.0	5.5

Table S4: Table of model parameter posterior RMSEs to the true values for different current-clamp designs, corresponding to the values shown in Figure S4. The last column shows the mean of the posterior RMSEs which is equivalent to the values shown in Figure S4B. Highlighted are the benchmark protocols. All values are $\times 10^3$ for ease of reading.