### Title

Quantifying the impact of electric fields on single-cell motility




### Journal

Biophysical Journal, 120(16)

### ISSN

0006-3495



### Authors

Prescott, Thomas P
Zhu, Kan
Zhao, Min
et al.



### Publication Date

2021-08-01

### DOI

10.1016/j.bpj.2021.06.034

Peer reviewed

Manuscript submitted to **Biophysical** *Journal*

Article

# Quantifying the impact of electric fields on single-cell motility

Thomas P Prescott[1], Kan Zhu[2], Min Zhao[2], and Ruth E Baker[1]

[1]Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

[2]Department of Ophthalmology and Vision Science, Department of Dermatology, Institute for Regenerative Cures, University of California, Sacramento, CA 95817, USA

ABSTRACT Cell motility in response to environmental cues forms the basis of many developmental processes in multicellular organisms. One such environmental cue is an electric field (EF), which induces a form of motility known as electrotaxis. Electrotaxis has evolved in a number of cell types to guide wound healing, and has been associated with different cellular processes, suggesting that observed electrotactic behaviour is likely a combination of multiple distinct effects arising from the presence of an EF. In order to determine the different mechanisms by which observed electrotactic behaviour emerges, and thus to design EFs that can be applied to direct and control electrotaxis, researchers require accurate quantitative predictions of cellular responses to externally-applied fields. Here, we use mathematical modelling to formulate and parametrise a variety of hypothetical descriptions of how cell motility may change in response to an EF. We calibrate our model to observed data using synthetic likelihoods and Bayesian sequential learning techniques, and demonstrate that EFs impact cellular motility in three distinct ways. We also demonstrate how the model allows us to make predictions about cellular motility under different EFs. The resulting model and calibration methodology will thus form the basis for future data-driven and model-based feedback control strategies based on electric actuation.

SIGNIFICANCE Electrotaxis is attracting much interest and development as a technique to control cell migration due to the precision of electric fields as actuation signals. However, precise control of electrotactic migration relies on an accurate model of how cell motility changes in response to applied electric fields. We present and calibrate a parametrised stochastic model that accurately replicates experimental single-cell data and enables the prediction of input–output behaviour while quantifying uncertainty and stochasticity. The model allows us to quantify three distinct ways in which electric fields perturb the motile behaviour of the cell. This model and the associated simulation-based calibration methodology will be central to future developments in the control of electrotaxis.

## INTRODUCTION

Cell migration underpins key physiological processes central to developmental biology, as well as wound healing and tissue regeneration, and it plays a crucial role in invasive, metastatic cancers. There are ongoing efforts to intervene in and influence these phenomena to, for example, inhibit metastasis (1) or accelerate wound healing (2). However, the cellular processes driving collective migration are complex and multifaceted, deriving from diverse physical mechanisms and various external stimuli (3), making it challenging for researchers to accurately and robustly direct cell motility. Due to the ease with which electric fields can be controlled and applied to cells, research into the control of cell motility has recently focused on exploiting *electrotaxis*

(also known as galvanotaxis) (3–5). However, the precise effects of electric fields on intracellular processes and thus on cell motility are not fully understood, making quantitative predictions and control policy design impractical.

Electrotactic cells have been observed to change their motile behaviour in response to the presence of a direct current (DC) electric field (EF) (3–7). Researchers seeking to control cell motility exploit this phenomenon by applying external electrical cues to cell populations (2, 4–9). The key advantages of using electrical cues to guide cell migration include the ability to exploit endogenous, evolved biological functionality to respond to precisely controllable DC EFs. This compares favourably to using chemoattractants to guide motility, since chemical signals experienced by the cell cannot be so precisely or flexibly controlled, especially dynamically, and chemoattractants are usually highly cell-specific. In contrast, light-directed motility allows for precise actuation signals. However, it requires sophisticated optogenetic manipulations of the cell population under control (10). As such, EFs provide a relatively precise and simply implemented actuation signal to achieve specified motile behaviours.

While an important strength of electrotactic cell control is that applying an EF for actuation is flexible enough to apply to any electrotactic cell type, the precise signal to be applied in order to achieve any specified goal needs to be carefully calibrated. At the most basic level, even the direction of migration within the same DC field has been shown to vary across different cell types, and within one cell type under different experimental conditions (5, 11). More broadly, a large number of biochemical and biophysical mechanisms have been implicated in the electrotactic response across different cell types (3). Each electrotactic mechanism, which may co-exist in combination at unknown relative strengths, may induce distinct observable effects on the dynamics of cellular motility. Overcoming this uncertainty in the observable electrotactic response is a fundamental challenge for designing EFs to control cell motility.

Mathematical models are a vital tool for quantifiably specifying the different ways in which cells can change their motility in response to EFs (12–14). In this paper we describe a parametrised stochastic model of the motile behaviour of a single human corneal epithelial cell, in which the cell's motility is driven by an internal polarity, in combination with the external influence of a DC EF. We assume that the cell can undergo both spontaneous and electrotactic polarisation. The model allows us to describe mathematically four distinct ways an EF may influence motility. We use experimentally observed trajectories of single cells, both with and without applied EFs, to calibrate the parameters of this model, thereby quantifying the extent to which different aspects of cell motility are impacted by the EF. The resulting calibrated model provides a vital first step towards being able to design feedback control policies and provide robustness guarantees, which are necessary if electrotaxis is to be used to control cell motility in practical applications such as wound healing or tissue engineering.

## Single-cell modelling

The agent-based modelling framework used in this work follows standard modelling assumptions outlined in (13). Specifically, we model the evolution of the velocity of a single cell in the overdamped regime, so that cell velocity is proportional to the sum of non-frictional forces on the cell. We provide full details on the mathematical model in *Materials and Methods* and in the *Supplementary Material*.

In the absence of any EF, the only non-frictional force acting on the cell is assumed to be an active force arising from the internal *polarity* of the cell. Thus, the cell velocity, $\mathbf{v} = \mathbf{v}_{\text{cell}}$, is comprised of a single component. Based on qualitative observations of spontaneous cellular motility, we assume that the cell randomly switches between two states, depolarised or polarised. In the depolarised state, the velocity stochastically fluctuates around a zero modal value, such that $\|\mathbf{v}_{\text{cell}}\| \approx 0$. In the polarised state, the modal values for the stochastic fluctuations of the velocity component are parametrised by $\|\mathbf{v}_{\text{cell}}\| \approx v$, where the scalar-valued parameter $v > 0$ has dimensions $\mu\text{m}\,\text{min}^{-1}$. In addition to the random changes in cell speed between depolarised and polarised states, the direction of cell motion (in the absence of an EF) is assumed to vary according to an unbiased random walk with positive diffusion constant $D > 0$, with dimensions $\text{min}^{-1}$. Eq. (2) in *Materials and Methods* provides the mathematical formulation of this model.

We hypothesise that a vector-valued DC EF, $\mathbf{u}$, can affect cell motility in a variety of ways. We use a number of extensions

of the model in order to determine how motility is impacted by the EF, specifying, in particular, four distinct ways in which it may affect the dynamics of a motile cell. We parametrise the magnitude of each hypothesised electrotactic effect, observed at a reference EF strength of 200 mV mm$^{-1}$, by the parameters $\gamma_1$, $\gamma_2$, $\gamma_3$ and $\gamma_4$, such that if $\gamma_i = 0$ then the corresponding hypothesised effect is not included in the model. Eq. (3) in *Materials and Methods* provides the mathematical formulation of this model.

The four means by which we model cell motility to be perturbed by the EF are:

**Velocity bias** ($\gamma_1$) The EF imparts an additional component of force on the cell. The resulting velocity, $\mathbf{v} = \mathbf{v}_{\text{cell}} + \mathbf{v}_{\text{EF}}$, is thus the sum of two components: the original polarity component, $\mathbf{v}_{\text{cell}}$, and an EF component, $\mathbf{v}_{\text{EF}}$. The EF velocity component acts in the direction of the field with magnitude $\gamma_1 v$.

**Speed increase** ($\gamma_2$) Polarised cells travel more quickly under the influence of an EF in the direction in which they are polarised. The modal magnitude of $\mathbf{v}_{\text{cell}}$ for polarised cells is increased by $\gamma_2 v$.

**Speed alignment** ($\gamma_3$) Polarised cells travel more quickly when the direction of their polarisation aligns with the EF, but slower if opposed to the EF. The modal magnitude of $\mathbf{v}_{\text{cell}}$ for polarised cells is increased by $\gamma_3 v \cos(\theta)$, where $\theta$ is the angle between $\mathbf{v}_{\text{cell}}$ (i.e. the polarity direction) and the EF direction.

**Polarity bias** ($\gamma_4$) The random walk determining cell polarity is biased so that cells preferentially polarise in the direction of the EF. The strength of this bias is parametrised by $\gamma_4$.

Two models can be distinguished: the *autonomous* model, where no EF is applied, and the *electrotactic* model, where a reference strength EF is applied. In each of these models, the cell velocity at time $t$, denoted $\mathbf{v}(t)$, undergoes a random walk. Figure 1 characterises each of these models by depicting the stationary probability distribution of this random walk. The top plot shows that, without an EF, the cell speed is most likely to be near zero or near $v$, with direction chosen uniformly at random. The bottom plot of this figure demonstrates how each electrotactic effect, quantified by the value of $\gamma_i$ for $i = 1, 2, 3, 4$, can be interpreted in terms of the probability distribution of the cell velocity: $\gamma_1$ translates the velocity distribution uniformly in the direction of the field; $\gamma_2$ rescales the domain of the distribution; $\gamma_3$ parametrises asymmetry in the shape of the velocity distribution; and $\gamma_4$ parametrises asymmetry in the density of the velocity distribution.

[Figure 1 about here; moved to end of manuscript by `endfloat`.]

## Outline

The primary goal of this work is to use single-cell experimental data to calibrate the parametrised mathematical model of spontaneous polarisation and electrotaxis. The model calibration process enables the identification of which of the four hypothesised electrotactic effects of EFs on cell motility can be observed in the experimental data. Importantly, the calibrated model also quantifies the relative contribution of each of these identified effects. In addition to the quantitative understanding of how electrotactic motility arises in cells, the calibrated model also allows us to simulate and predict the single-cell response to dynamic EFs.

The data used for model calibration is gathered from two experiments in which the trajectories of motile human corneal epithelial cells are recorded (a) without any EF applied, and (b) with a DC EF at a reference strength of 200 mV mm$^{-1}$, directed from left to right. These experiments are termed the *autonomous* and *electrotactic* experiments, respectively. We use these experiments in turn to first calibrate the parameters of the autonomous model and then the electrotactic model. To calibrate the electrotactic model, we first identify which of the four hypothesised electrotactic effects are supported by the data. Out of these identified electrotactic effects, we then proceed to quantify the relative contribution of each of them to the observed electrotaxis induced by the EF.

After formulating and calibrating the extended model of electrotaxis, we use simulations of the calibrated model to predict how the position and polarity of single cells evolve under the influence of different, dynamic EF inputs. The ability to make predictions using a calibrated, stochastic, uncertain model is a first step towards the future goal of model-based policy design for the electrotactic control of single-cell and population-level motility.

# MATERIALS AND METHODS

## Data collection

Two experiments were carried out, which we call the *autonomous* and *electrotactic* experiments. In both experiments, time-lapse images of human corneal epithelial cells, seeded at a low density, were acquired at 5 min intervals over 3 h. In the electrotactic experiment, the cells were subjected to a DC EF at a reference strength, 200 mV mm$^{-1}$, applied across the medium. In the autonomous experiment, no EF was applied. From the two resulting image sets, the positions of fifty cell centroids for each experiment were recorded over the entire time horizon. Visual confirmation from the raw experimental output confirms that cell collisions were rare, due to the low density (100 cells cm$^{-2}$) at which cells are initially seeded. We thus assume that cell–cell interactions can be neglected in the current model. We denote the resulting cell trajectory data $\mathbf{x}_{\mathrm{NoEF},i}(t_j)$ and $\mathbf{x}_{\mathrm{EF},i}(t_j)$ for the autonomous and electrotactic experiments, respectively, where each trajectory is translated to begin at the origin, such that $\mathbf{x}_{\mathrm{NoEF},i}(0) = \mathbf{x}_{\mathrm{EF},i}(0) = 0$ for all $i$. For each experiment, the index $i = 1, \ldots, 50$ refers to the cell being traced, and $t_j = 5j$ min for $j = 0, 1, \ldots, 36$ refers to the snapshot time points. We denote the entire set of data from each experiment as $\mathbf{x}_{\mathrm{NoEF}}$ and $\mathbf{x}_{\mathrm{EF}}$, respectively.

### Materials

EpiLife culture medium with $Ca^{2+}$ (60 µM), EpiLife defined growth supplement, and penicillin/streptomycin were purchased from ThermoFisher Scientific (Waltham, MA, USA). FNC Coating Mix was purchased from Athena Enzyme Systems (Baltimore, MD, USA). Dow Corning high-vacuum grease was purchased from ThermoFisher. Agar was purchased from MilliporeSigma (Burlington, MA, USA). Silver wires with 99.999% purity were purchased from Advent Research Materials Ltd. (Oxford, United Kingdom).

### Cell culture

Telomerase-immortalized human corneal epithelial cells (hTCEpi) were routinely cultured in EpiLife medium supplemented with EpiLife defined growth supplement and 1% (v/v) penicillin/streptomycin. Cells were incubated at 37 °C with 5% $CO_2$ until they reached ~70% confluence and were used between passages 55 and 65 for all cell migration assays.

### Electrotaxis assay

Electrotaxis experiments were performed as previously described (15, 16) with minor changes. Briefly, electrotaxis chambers (2 cm x 1 cm) were constructed in 100 mm petri dishes with glass strips and high-vacuum grease. Chambers were coated with FNC Coating Mix, following the manufacturer's instructions to facilitate cell attachment. Cells were seeded at a low density (100 cells cm$^{-2}$) and cultured overnight (12 h to 18 h) in the chambers to allow sufficient attachment. Chambers were covered with glass coverslips and sealed with high-vacuum grease. Electric currents were applied to the chamber through agar-salt bridges connecting with silver–silver chloride electrodes in Steinberg's solution (58 mM NaCl, 0.67 mM KCl and 0.44 mM $Ca(NO_3)_2$, 1.3 mM $MgSO_4$ and 4.6 mM Tris base, pH 7.4). Fresh cell culture medium (Epilife) was added into reservoirs to ensure good salt bridge contact and to support cell viability during electric stimulation. An EF strength of 200 mV mm$^{-1}$ was used, and field strengths were monitored at the beginning of the experiment and every 30 min afterwards to

ensure consistent EF application.

### Time-lapse imaging and quantification of cell migration

Cell migration was monitored and recorded by phase-contrast microscopy using an inverted microscope (Carl Zeiss, Oberkochen, Germany) equipped with a motorized stage and a regular 10× objective lens. Time-lapse images were acquired at 5 min intervals using Metamorph NX imaging software (Molecular Device, Sunnyvale, CA, USA). To maintain standard cell culture conditions (37 °C, 5% $CO_2$), a Carl Zeiss incubation system was used. Time-lapse images of cell migration were analyzed by using ImageJ software from the National Institutes of Health (http://rsbweb.nih.gov/ij/). Adherent cells in the images were manually tracked, and cells that divided, moved in and out of the field, or merged with other cells during the experiment were excluded from analysis. The position of a cell was defined by its centroid.

## Model construction

We constructed a mathematical model of single-cell dynamics. The model tracks the position of the cell centre in the plane, $\mathbf{x}(t) \in \mathbb{R}^2$, as a function of time, $t \geq 0$ min, with initial condition $\mathbf{x}(0) = 0$ at the origin. The position is a deterministic integral of cell velocity, $\mathbf{v}$, such that

$$\mathrm{d}\mathbf{x}(t) = \mathbf{v}(t)\,\mathrm{d}t, \tag{1}$$

and the stochastic dynamics of $\mathbf{v}$ are modelled. The key to this modelling task is the non-dimensional internal variable representing the cell polarity, $\mathbf{p}(t) \in \mathbb{R}^2$. We assume that the polarity imparts a force on the cell that corresponds to its active motility, resulting in a velocity component $\mathbf{v}_{\mathrm{cell}}(t)$.

### Modelling spontaneous polarisation and motility

We first describe the model of cellular motility with no biasing EF, which we will term the *autonomous* model. The only velocity component is that due to polarisation, so that we write the cell velocity as a single component,

$$\mathbf{v}(t) = \mathbf{v}_{\mathrm{cell}}(t) = v\mathbf{p}(t), \tag{2a}$$

where the parameter $v \geq 0$, with dimensions $\mu\mathrm{m\,min}^{-1}$, represents the modal magnitude of $\mathbf{v}_{\mathrm{cell}}$ for a polarised cell. Note that Eq. (2a) implies that the polarity variable, $\mathbf{p}$, is a non-dimensionalisation of the velocity component $\mathbf{v}_{\mathrm{cell}}$. We further assume that the polarity, $\mathbf{p}$, undergoes a random walk according to a Langevin diffusion, such that

$$\mathrm{d}\mathbf{p}(t) = -D\nabla W(\mathbf{p}(t))\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}\mathbf{B}, \tag{2b}$$

where $\mathbf{B}(t) \in \mathbb{R}^2$ is a two-dimensional Wiener process, and the parameter $D$ (in $\mathrm{min}^{-1}$) quantifies the speed at which the random walk approaches stationarity. The initial polarity, denoted $\mathbf{p}_0 = \mathbf{p}(0)$, also needs to be specified.

The *potential function* $W(\mathbf{p})$ in Eq. (2b) is defined to capture the intended features of the autonomous model, namely that cells stochastically spontaneously switch between polarised and depolarised states, and that the argument of the polarity is uniformly distributed, at stationarity. It can be shown (17, 18) that the transition rates between the polarised and depolarised states are determined by two non-dimensional *energy barriers*, denoted $\Delta W_{\mathrm{on}}$ and $\Delta W_{\mathrm{off}}$. For further details on the definition of $W$, see the *Supplementary Material*. We will calibrate the autonomous model in Eq. (2) by identifying the parameters $v$ and $D$, and also the two energy barriers, $\Delta W_{\mathrm{on/off}}$, that are sufficient to determine the potential function, $W$.

## Modelling motility bias due to an EF

We use the vector $\mathbf{u}(t)$ with non-dimensional magnitude $\|\mathbf{u}(t)\| = u(t)$ to describe a (time-varying) DC EF of strength $200u(t)$ mV mm$^{-1}$, directed parallel to $\mathbf{u}(t)$. In particular, the EF used in the electrotactic experiment, with magnitude 200 mV mm$^{-1}$ in the positive $x$ direction (left to right), is represented by the constant canonical unit vector, $\mathbf{u}(t) \equiv \mathbf{i}$, with constant (non-dimensional) magnitude $u(t) \equiv 1$. The autonomous model in Eq. (2) can be extended to include the four hypothesised effects of the EF. The *velocity bias* effect is accounted for by modelling the velocity using two components,

$$\mathbf{v}(t) = \mathbf{v}_{\text{cell}}(t) + \mathbf{v}_{\text{EF}}(t), \tag{3a}$$

where the EF induces a deterministic velocity component in the direction of the field,

$$\mathbf{v}_{\text{EF}}(t) = \gamma_1 v \mathbf{u}(t). \tag{3b}$$

The two hypothesised electrotactic effects of *speed increase* and *speed alignment* are both modelled through adapting the velocity component induced by the cell polarity, originally defined in Eq. (2a), into

$$\mathbf{v}_{\text{cell}}(t) = (1 + \gamma_2 u(t) + \gamma_3 \mathbf{u}(t) \cdot \hat{\mathbf{p}}(t))\, v\mathbf{p}(t), \tag{3c}$$

where $\hat{\mathbf{p}}$ is the unit vector in the direction of the polarity, $\mathbf{p}$. Finally, the hypothetical *polarity bias* effect is modelled in the stochastic evolution of the polarity variable $\mathbf{p}$. We add a drift term proportional to the EF to the Langevin diffusion equation, such that

$$\mathrm{d}\mathbf{p}(t) = -D\,[\nabla W(\mathbf{p}(t)) - \gamma_4 \mathbf{u}(t)]\;\mathrm{d}t + \sqrt{2D}\;\mathrm{d}\mathbf{B}, \tag{3d}$$

where $W(\mathbf{p})$ is the same potential function as used in Eq. (2b). As for the autonomous model, the initial value for the polarity, denoted $\mathbf{p}_0 = \mathbf{p}(0)$, is also required.

Note that substituting $\mathbf{u} = 0$ or setting $\gamma_i = 0$ for all $i = 1, 2, 3, 4$ into Eq. (3) recovers the dynamics of the autonomous model, in Eq. (2). We will term the extended model in Eq. (3) the *electrotactic* model. It is parametrised by the four parameters $v$, $\Delta W_{\text{on}}$, $\Delta W_{\text{off}}$ and $D$, with the same meaning and dimensions as in the autonomous model, and also by $\gamma_i$ for $i = 1, 2, 3, 4$, which, since $\mathbf{u}$ is non-dimensional, are all non-dimensional.

The models in Eq. (2) and Eq. (3) result in a stochastic path for the velocity, $\mathbf{v}(t)$, with a parametrically determined stationary distribution. Following Eq. (1), each path can be integrated to produce a stochastic trajectory of the cell position over time. The stationary distributions of $\mathbf{v}$ under the autonomous and electrotactic models are depicted in Figure 1, where the polarised and depolarised regimes in both models can be identified. Furthermore, the effect of each of the parameters $\gamma_i$, and hence each of the hypothesised electrotactic effects, can be identified by comparing the position, scale, and asymmetries of the two stationary distributions.

## Summarising simulations

For any given set of parameter values, $\theta = (v, \Delta W_{\text{on}}, \Delta W_{\text{off}}, D, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$, together with initial polarity, $\mathbf{p}_0$, and non-zero EF input, $\mathbf{u}(t)$, the stochastic model in Eq. (3) can be simulated. Note that, if the EF is zero, we simulate the autonomous model in Eq. (2). Each simulation produces a random trajectory, denoted $\omega = (\mathbf{p}(t), \mathbf{x}(t))_{t\geq 0}$. We will use *summary statistics* to analyse the model outputs by mapping each simulated trajectory, $\omega$, to a number (or small set of numbers) that summarise the trajectory. More details of the summary statistics can be found in the *Supplementary Material*.

We define one set of summary statistics based on simulated cell positions. We consider: (a) the total cell displacement over the time horizon of the experiments, $\|\mathbf{x}(180)\|$, denoted $Y_1(\omega)$; (b, c) the mean and standard deviation of the displacements,

$\|\mathbf{x}(t_j) - \mathbf{x}(t_{j-1})\|$, between the five-minute sample points, denoted $Y_2(\omega)$ and $Y_3(\omega)$; and (d) the angle, $\arg(\mathbf{x}(180))$, between the final cell position and the positive $x$ axis (i.e. the direction of the EF under the electrotactic experiment), denoted $Y_4(\omega)$. Note that the four summary statistics $Y_1$, $Y_2$, $Y_3$ and $Y_4$ can also be applied to the observed data, $\mathbf{x}_{\text{NoEF},i}$ and $\mathbf{x}_{\text{EF},i}$, in addition to any simulated trajectory, $\omega$.

In the models in Eq. (2) and Eq. (3), the polarity, $\mathbf{p}(t)$, evolves randomly from initial value $\mathbf{p}_0$. We define a further three summary statistics based on the simulated polarity, using a threshold polarity magnitude, $\bar{p}$. First, the *time to polarise*, $T_1(\omega)$, is defined as the first time $t \geq 0$ for which $\|\mathbf{p}(t)\| \geq \bar{p}$, for $\omega$ generated using parameter value $\theta$ and initial condition $\mathbf{p}_0 = 0$. The *time to depolarise*, $T_0(\omega)$, is the first time $t \geq 0$ for which $\|\mathbf{p}(t)\| \leq \bar{p}$, for $\omega$ generated using parameter value $\theta$ and initial condition $\mathbf{p}_0 = \mathbf{i}$. Finally, the binary value $\Pi_T(\omega)$ for any given $T \geq 0$ is the indicator function of $\|\mathbf{p}(T)\| \geq \bar{p}$. Given a parameter value, $\theta$, the conditional expectations of these statistics are denoted by a bar, so that $\bar{T}_1 = \mathbb{E}(T_1 \mid \theta)$ and so on. Thus, Bayesian uncertainty in parameter values will propagate to the estimated values of $\bar{T}_1$, $\bar{T}_0$, and $\bar{\Pi}_T$. Note that these summary statistics *cannot* be applied to the observed data, and can only be used to summarise simulated trajectories.

## Model calibration and selection

Given the experimental data sets, $\mathbf{x}_{\text{NoEF}}$ and $\mathbf{x}_{\text{EF}}$, the autonomous and electrotactic models can be *calibrated* by identifying the values of the parameters,

$$\theta = (v, \Delta W_{\text{on}}, \Delta W_{\text{off}}, D, \gamma_1, \gamma_2, \gamma_3, \gamma_4),$$

that are consistent with the observed behaviour. We employ a Bayesian approach to parameter inference, whereby prior beliefs about $\theta$, encoded in a *prior distribution*, $\pi(\theta)$, are updated in the context of the experimental data according to Bayes's rule,

$$\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) = \frac{\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)\pi(\theta)}{p(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})},$$

where $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)$ is the *likelihood* of observing the data under the models in Eq. (2) and Eq. (3) with the parameter value $\theta$. The resulting *posterior distribution*, $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, represents the remaining uncertainty in the parameter values, given the experimental data (19).

The simulation and inference algorithms used in this work have been developed in Julia 1.5.1 (20). The code is publicly available at `github.com/tpprescott/electro`.

### Bayesian synthetic likelihoods and sequential Monte Carlo

In practice, the likelihood cannot be calculated directly, and so we require a likelihood-free approach. We replace the true likelihood with a *synthetic likelihood*, where for each value of $\theta$ the likelihood is approximated by the likelihood of summarised data under an empirical Gaussian distribution, which is fit to the sample mean and covariance of a set of $n = 500$ summarised simulations (21–23). The summary statistics we use are $Y_1$, $Y_2$, $Y_3$ and $Y_4$, as described in *Summarising simulations*, above. To mitigate the computational burden of the large number of model simulations required for parameter inference, we combine a sequential Monte Carlo (SMC) algorithm with synthetic likelihoods (21, 24, 25). This approach is a popular strategy for efficiently sampling from a target distribution, and also allows the exploitation of parallelisation to speed inference (21, 25–27). We provide full details of the SMC inference approach using summary statistics and synthetic likelihoods in the *Supplementary Material*.

### Prior specification and model selection

The space of possible parameter values is defined as the product of intervals,

$$\Theta = (0,5]^3 \times (0,0.5] \times [0,2]^4,$$

where the interval bounds were chosen based on a preliminary qualitative, visual analysis of the simulation outputs in comparison to observed data. In order to identify which of the sixteen possible combinations of the four hypothesised electrotactic effects are best supported by the experimental data, we will define sixteen possible priors on $\Theta$. For each of the sixteen subsets, $X \subseteq \{1,2,3,4\}$, we define a uniform prior distribution $\pi_X(\theta)$ on $\Theta$ that takes a constant, positive value for parameter vectors $\theta$ if and only if $\gamma_i > 0$ for all $i \in X$, and is zero otherwise. Thus, by performing Bayesian inference using the prior distribution, $\pi_X$, we constrain the electrotactic model in Eq. (3) to model only electrotactic effects included in the subset $X \subseteq \{1,2,3,4\}$.

We define an optimisation problem that aims to prevent over-fitting, by balancing the closeness of the model fit to data while prioritising smaller parameter dimensions. The optimal subset, $X$, of electrotactic effects is defined as the maximiser of the objective function,

$$J_\mu(X) = \log p_X(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) - \mu(4 + |X|), \tag{4}$$

where the regularisation parameter $\mu > 0$ controls the cost of over-fitting by penalising the total number of non-zero parameters. This number is four, corresponding to $v$, $\Delta W_{\text{on/off}}$, and $D$, plus $|X|$, corresponding to the positive $\gamma_i$ for $i \in X$. We use $\mu = 0$ and $\mu = 2$ in our analysis, although the choice of $\mu$ is somewhat arbitrary. One interpretation of the value of $\mu$ is that it effectively imposes a 'prior' on the subsets, $X \subseteq \{1,2,3,4\}$, with probability mass proportional to $\exp(-\mu|X|)$.

The first term in $J_\mu(X)$ measures the closeness of fit between the data and the model, when constrained to only include the electrotactic effects in $X$. This fit is defined for each $X \subseteq \{1,2,3,4\}$ by the value of the partition function,

$$p_X(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) = \int \mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)\pi_X(\theta)\, \mathrm{d}\theta.$$

As the likelihoods $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)$ cannot be calculated directly, the partition functions $p_X(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$ are estimated for each $X$ by Monte Carlo sampling, where again the simulation-based synthetic likelihood is used in place of the true likelihood. More details of the specific sequential Monte Carlo sampling methodology used for this estimate are given in the *Supplementary Material*.

## RESULTS

We initially calibrate the autonomous model, based on the data set $\mathbf{x}_{\text{NoEF}}$ from the autonomous experiment alone, in order to confirm the principle of the modelling framework and its ability to replicate observed behaviours, and to check that the parameters are identifiable from the data. Then, we calibrate the full electrotactic model using the full data set, $\mathbf{x}_{\text{NoEF}}$ and $\mathbf{x}_{\text{EF}}$, in two stages. We first assess which subset of the four hypothesised electrotactic effects are best supported by the data. After choosing the optimal combination of electrotactic effects, we then calibrate the parameters of the selected electrotactic model.

### Parameters of the autonomous model are identifiable

We begin by confirming that the chosen modelling and inference approaches appropriately capture the autonomous experimental behaviour, $\mathbf{x}_{\text{NoEF}}$, where no external EF is applied. This scenario is modelled by the autonomous model in Eq. (2), which depends on four parameters, $\theta_{\text{NoEF}} = (v, \Delta W_{\text{on}}, \Delta W_{\text{off}}, D)$. The Bayesian synthetic likelihood approach was used to generate posterior samples for: the characteristic speed of a polarised cell, $v$ μm min$^{-1}$; the diffusion constant, $D$ min$^{-1}$, which determines the characteristic timescale of the spontaneous polarisation dynamics; and the dimensionless parameters, $\Delta W_{\text{on}}$ and $\Delta W_{\text{off}}$,

which respectively determine the relative rates at which cells switch from depolarised to polarised, and vice versa.

[Figure 2 about here; moved to end of manuscript by `endfloat`.]

Figure 2(a–d) depicts the marginals of the posterior distribution, $\pi(\theta_{\text{NoEF}} \mid \mathbf{x}_{\text{NoEF}})$, for each of the four calibrated parameters. The prior distribution used for Bayesian inference assumed that the parameters were independently uniformly distributed on the intervals $0 < v \leq 5\ \mu\text{m min}^{-1}$, $0 < \Delta W_{\text{on/off}} \leq 5$, and $0 < D \leq 0.5\ \text{min}^{-1}$. Each plot in Figure 2(a–d) demonstrates that the posteriors are concentrated within a small interval of the prior support, implying that the parameters of the autonomous model are identifiable from the experimental data, with quantifiable uncertainty.

The sample mean parameter value, calculated from the sample in Figure 2, can be used as a point estimate for the parameter values, $v = 1.82\ \mu\text{m min}^{-1}$, $\Delta W_{\text{on}} = 1.59$, $\Delta W_{\text{off}} = 0.25$, and $D = 0.031\ \text{min}^{-1}$. In Figure 2(e–f), we compare the observed trajectories from the autonomous experiment, $\mathbf{x}_{\text{NoEF}}$, to trajectories simulated from the autonomous model with parameter values given by this point estimate. A comparison between these plots shows that parameter inference based only on the selected four-dimensional summary statistics produces a close match (for this point estimate) between the visual characteristics of simulations and experimental observations.

Figure 2(a–d) quantifies the uncertainty in each parameter value resulting from the Bayesian approach to parameter inference. In order to make sense of this uncertainty in terms of the model outputs, simulations can be used to interpret how the uncertainty propagates to observable behaviour. Figure 2(g–i) depicts an estimate of the uncertainty in (g) the average time a simulated cell takes to polarise, $\bar{T}_1$, (h) the average time a simulated cell takes to depolarise, $\bar{T}_0$, and (i) the proportion of simulated cells that are polarised at any time, $\bar{\Pi}_\infty$. Each of these distributions are conditioned on the posterior parameter distribution in Figure 2(a–d). This procedure allows us to map quantified uncertainty in the parameter values to uncertainty in cell behaviour. The calibrated model suggests that the expected time for a cell to spontaneously polarise (i.e. without an EF applied) ranges from 22 min to 75 min (5% to 95% quantiles), with median value of 36 min. Similarly, the expected time for a cell to depolarise is 6 min to 21 min, with median value 10 min. Finally, the probability that a simulated cell is polarised (in any direction) at any one time is 0.29 to 0.40, with median value 0.34.

## Three of the four proposed electrotactic effects are supported by the data

Given that the autonomous model can be calibrated to the data set from the autonomous experiment, we now seek to calibrate the full electrotactic model to the entire data set from both experiments. However, some or all of the hypothesised electrotactic effects used to define the model in Eq. (3) may not be supported by the experimental data. Thus, we first use the data to select which of these proposed effects can be detected in the observed cell behaviours. Recall that the parameters $\gamma_1$, $\gamma_2$, $\gamma_3$ and $\gamma_4$ correspond to four distinct hypothesised electrotactic effects: velocity bias, speed increase, speed alignment, and polarity bias. Positive values of the parameters $\gamma_i$, for $i = 1, 2, 3, 4$, mean that the corresponding effect is included in the model. Conversely, setting any of these parameters to zero excludes the corresponding effect(s) from the model. There are a total of $2^4 = 16$ possible combinations of the four proposed electrotactic effects that the model in Eq. (3) can implement, through combinations of positive and zero parameter values.

[Figure 3 about here; moved to end of manuscript by `endfloat`.]

Each of the 16 possible combinations of the four electrotactic effects corresponds to a subset $X \subseteq \{1, 2, 3, 4\}$. We evaluate each combination of electrotactic effects, given by $X$, with respect to the objective function, $J_\mu(X)$, given in Eq. (4). This objective quantifies the trade-off between the model fit and the number of non-zero parameters in order to select a suitably accurate model while avoiding overfitting. Figure 3 ranks each of the 16 possible combinations of electrotactic effects, $X \subseteq \{1, 2, 3, 4\}$, using two different objective functions. The top plot considers $\mu = 0$, such that the maximiser of $J_0$ is the combination that gives the best fit to data, with no consideration given to the dimension of parameter space. The bottom plot

uses $\mu = 2$, which imposes a marginal cost on increasing the dimension of parameter space. Both objective functions are maximised by the subset $X = \{1, 2, 4\}$. This provides strong support for including the velocity bias, speed increase, and polarity bias effects of the EF in our model, and neglecting the hypothesised effect of speed alignment.

## The electrotactic effects of the EF on motility can be quantified

Recall that cell motility is modelled as the sum of an active force component, deriving from cell polarisation, and a component comprised of other external forces acting on the cell. In the preceding section, we found that a polarised cell applies a greater active force in the presence of an EF (positive *speed increase*, $\gamma_2 > 0$). However, there is insufficient evidence that the increase in this active force varies with the direction of the EF (zero *speed alignment*, $\gamma_3 = 0$). Instead, the observed directional bias in motility in the electrotactic experimental data, $\mathbf{x}_{\text{EF}}$, is found to arise from a combination of the cell tending to polarise in alignment with the EF (positive *polarity bias*, $\gamma_4 > 0$) and the EF imparting an additional external force on the cell (positive *velocity bias*, $\gamma_1 > 0$). In this section, we quantify the relative contributions of each of these three non-zero effects to the observed electrotactic behaviour.

Bayesian synthetic likelihoods were used to calibrate the electrotactic model by inferring the posterior distribution, $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, for

$$\theta = (v, \Delta W_{\text{on}}, \Delta W_{\text{off}}, D, \gamma_1, \gamma_2, \gamma_3, \gamma_4).$$

The chosen prior distribution, $\pi_{1,2,4}(\theta)$, is the product of independent uniform distributions on the intervals $0 < v \leq 5\ \mu\text{m min}^{-1}$, $0 < \Delta W_{\text{on/off}} \leq 5$, and $0 < D \leq 0.5\ \text{min}^{-1}$, multiplied by independent and uniformly distributed priors for the electrotactic parameters on the intervals $0 < \gamma_1, \gamma_2, \gamma_4 \leq 2$. The remaining parameter in $\theta$ is fixed at $\gamma_3 = 0$.

[Figure 4 about here; moved to end of manuscript by `endfloat`.]

Figure 4(a–c) shows three empirical marginals from the posterior sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, constructed using Bayesian synthetic likelihoods and SMC sampling. The three marginals shown correspond to the parameters of the three non-zero electrotactic effects, $\gamma_1$, $\gamma_2$, and $\gamma_4$. The posterior marginal distributions of the other non-zero parameters ($v$, $\Delta W_{\text{on/off}}$ and $D$) closely match those in Figure 2, as depicted in the *Supplementary Material*. Similarly to Figure 2, the posterior distribution is concentrated in a small region of the prior domain, providing evidence that each of the new parameters is identifiable using the chosen summary statistics.

The non-dimensional parameters $\gamma_1$, $\gamma_2$, and $\gamma_4$ can be interpreted as follows. For parameter $\gamma_1$, which quantifies *velocity bias*, the external force on the cell induced by the EF has magnitude approximately 39% to 65% that of the active force applied by a polarised cell (5% to 95% quantiles), with median value 51%. Similarly, for parameter $\gamma_2$, quantifying the *speed change*, the EF increases the speed of a polarised cell by approximately 9% to 24%, with median value of 16%. Finally, the field is found to induce a bias in cell polarisation towards alignment with the direction of the EF. Let $\phi$ denote the angle between the cell polarity direction and the EF. At stationarity, $\phi$ is distributed proportionally to $\exp(\gamma_4 \cos(\phi))$, with maximum at $\phi = 0$ and minimum at $\phi = \pi$. The parameter $\gamma_4$ thus determines the *polarity bias* effect towards $\phi = 0$, and is inferred to lie from 0.18 to 0.72, with median value of 0.49.

The mean of the SMC sample depicted in Figure 4(a–c) can be used as a point estimate for the parameter values: $v = 1.84\ \mu\text{m min}^{-1}$, $\Delta W_{\text{on}} = 1.51$, $\Delta W_{\text{off}} = 0.29$, $D = 0.029\ \text{min}^{-1}$, $\gamma_1 = 0.52$, $\gamma_2 = 0.16$, $\gamma_3 = 0$, and $\gamma_4 = 0.47$. Figure 4(d–e) compares the data from the electrotactic experiment, $\mathbf{x}_{\text{EF}}$, against 50 simulations from the electrotactic model, Eq. (3), using this point estimate. The observed bias in motility towards the direction of the EF is reflected in the stochastically simulated outputs. This provides visual confirmation that parameters inferred by Bayesian synthetic likelihood, based on the chosen summary statistics, produce simulated outputs that share observable characteristics with the experimental data.

## DISCUSSION & CONCLUSION

The primary goal of this work has been to use mathematical modelling to quantitatively tease apart the contributions of multiple hypothesised means by which EFs induce electrotaxis in single cells. We have presented an empirical, parametrised, agent-based model of electrotactic cell motility, and shown that it can be calibrated to single-cell trajectory data using likelihood-free Bayesian inference. To our knowledge, although many models of single-cell and collective motility under environmental cues have been developed (13), there have been few mathematical models of electrotaxis (28, 29), and this work is the first use of detailed mathematical modelling at a single-cell level to quantify motility under electrotaxis. Moreover, the inferred parameter values of the calibrated model provide quantitative, mechanistic insights into experimentally-observed electrotaxis.

Specifically, by calibrating the model to experimental observations of electrotaxis in human corneal epithelial cells, we have concluded that the observed bias in motility is the product of three distinct effects of the EF. First, the presence of the EF directly biases the motility of all cells, independently of their polarisation state, in the direction of the EF. Second, polarised cells are more motile in the presence of the EF. Third, cells preferentially break their symmetry to polarise in alignment with the direction of the EF. By carefully calibrating the parametrised mathematical model to experimental data, we have quantified the relative contributions of each of these distinct effects to electrotaxis.

A key strength of the model presented in Eq. (3) is its flexibility. The parametric design means that the Bayesian calibration methodology used in this work can be recapitulated to calibrate the same model to electrotaxis assays using other cell types or with different experimental conditions. Thus, observed differences in spontaneous and electrotactic motility between different cells and experimental conditions (3, 5) can be modelled and predicted within a common parametric framework. It is also important to acknowledge that we have chosen from only four hypothetical observable effects of electrotaxis. Other electrotactic effects may be reasonably included in the modelling process: for example, the EF may induce changes to the rate of polarisation and depolarisation (3). The electrotaxis model can straightforwardly be extended and recalibrated to account for any alternative hypothetical effects.

We have also considered EFs at a single reference strength, requiring a single parameter to quantify each hypothesised electrotactic effect. However, the characteristics of electrotaxis have been observed to vary nonlinearly with EF strength (5). The model is sufficiently flexible to account for this phenomenon through the replacement of the parameters $\gamma_i$ with functions $\Gamma_i(u)$ that vary with the EF strength, $u$ mV mm$^{-1}$. The challenge will then be to use experimental data gathered from assays using EFs of different strengths to infer each of the functions $\Gamma_i$ in place of each of the parameters $\gamma_i$.

The model we have presented predicts single-cell electrotactic behaviour. However, there is a wealth of data and analysis on electrotaxis in the context of cell populations (3, 4, 6–9, 13). The electrotaxis model in this paper is a starting point for a comprehensive agent-based model that also incorporates phenomena such as volume exclusion, adhesion, elastic collisions, contact inhibition, and so on (13, 30, 31). Multifidelity approaches (27, 32) that can link experiments and information at the single-cell and multicellular level will be vital to identify and quantify the biasing effects of EFs on the collective motility of cell populations (12, 14, 33).

The seven positive parameter values for the electrotactic model were inferred using data generated using two experimental inputs: a zero EF, and a constant EF of magnitude 200 mV mm$^{-1}$. A fundamental motivation for developing this model is to allow predictions of observed behaviour under different EF inputs, with the ultimate goal of controlling cell motility by designing and applying dynamic EFs. In the *Supplementary Material* we show, using two examples of dynamic EFs, that the calibrated model can be simulated to produce predictions of single-cell electrotaxis under dynamic inputs. The model considered in this paper, and the Bayesian uncertainty quantification of its parameters, are important tools for enabling stochastic model predictive control designs of such policies based on output feedback and filtering (34). We have therefore provided a significant step towards the real-time model predictive control of populations of electrotactic cells.

## AUTHOR CONTRIBUTIONS

The experiments and image analysis were performed by KZ and MZ. The modelling, simulation and inference were carried out by TPP and REB. All authors contributed to writing the manuscript.

## ACKNOWLEDGMENTS

The experiments in the Zhao lab were supported by the AFOSR-MURI grant FA9550-16-1-0052 and NIH grant 1R01EY019101. TPP and REB are supported by BBSRC through grant BB/R000816/1, and REB is supported by a Royal Society Wolfson Research Merit Award.

## REFERENCES

1. Stuelten, C. H., C. A. Parent, and D. J. Montell, 2018. Cell motility in cancer invasion and metastasis: insights from simple model organisms. *Nature Reviews Cancer* 18:296–312.

2. Kai, H., T. Yamauchi, Y. Ogawa, A. Tsubota, T. Magome, T. Miyake, K. Yamasaki, and M. Nishizawa, 2017. Accelerated wound healing on skin by electrical stimulation with a bioelectric plaster. *Advanced Healthcare Materials* 6:1700465.

3. Cortese, B., I. E. Palamà, S. D'Amone, and G. Gigli, 2014. Influence of electrotaxis on cell behaviour. *Integrative Biology* 6:817–830.

4. Cho, Y., M. Son, H. Jeong, and J. H. Shin, 2018. Electric field–induced migration and intercellular stress alignment in a collective epithelial monolayer. *Molecular Biology of the Cell* 29:2292–2302.

5. Saltukoglu, D., J. Grünewald, N. Strohmeyer, R. Bensch, M. H. Ulbrich, O. Ronneberger, and M. Simons, 2015. Spontaneous and electric field–controlled front–rear polarization of human keratinocytes. *Molecular Biology of the Cell* 26:4373–4386.

6. Zhao, M., A. Agius-Fernandez, J. V. Forrester, and C. D. McCaig, 1996. Directed migration of corneal epithelial sheets in physiological electric fields. *Investigative Ophthalmology & Visual Science* 37:2548–2558.

7. Zhao, M., B. Song, J. Pu, J. V. Forrester, and C. D. McCaig, 2003. Direct visualization of a stratified epithelium reveals that wounds heal by unified sliding of cell sheets. *The FASEB Journal* 17:397–406.

8. Cohen, D. J., W. J. Nelson, and M. M. Maharbiz, 2014. Galvanotactic control of collective cell migration in epithelial monolayers. *Nature Materials* 13:409–417.

9. Sun, Y.-H., Y. Sun, K. Zhu, B. Reid, X. Gao, B. W. Draper, M. Zhao, and A. Mogilner, 2017. Electric fields accelerate cell polarization and bypass myosin action in motility initiation. *Journal of Cellular Physiology* 233:2378–2385.

10. Hughes, R. M., and D. S. Lawrence, 2014. Optogenetic engineering: light-directed cell motility. *Angewandte Chemie International Edition* 53:10904–10907.

11. Wang, E., M. Zhao, J. V. Forrester, and C. D. McCaig, 2003. Bi-directional migration of lens epithelial cells in a physiological electrical field. *Experimental Eye Research* 76:29–37.

12. Lee, R. M., H. Yue, W.-J. Rappel, and W. Losert, 2017. Inferring single-cell behaviour from large-scale epithelial sheet migration patterns. *Journal of The Royal Society Interface* 14:20170147.

13. Camley, B. A., and W.-J. Rappel, 2017. Physical models of collective cell motility: from cell to tissue. *Journal of Physics D: Applied Physics* 50:113002.

14. Matsiaka, O. M., C. J. Penington, R. E. Baker, and M. J. Simpson, 2018. Discrete and continuum approximations for collective cell migration in a scratch assay with cell size dynamics. *Bulletin of Mathematical Biology* 80:738–757.

15. Nakajima, K., K. Zhu, Y.-H. Sun, B. Hegyi, Q. Zeng, C. J. Murphy, J. V. Small, Y. Chen-Izu, Y. Izumiya, J. M. Penninger, and M. Zhao, 2015. KCNJ15/Kir4.2 couples with polyamines to sense weak extracellular electric fields in galvanotaxis. *Nature Communications* 6.

16. Zhu, K., Y. Takada, K. Nakajima, Y. Sun, J. Jiang, Y. Zhang, Q. Zeng, Y. Takada, and M. Zhao, 2019. Expression of integrins to control migration direction of electrotaxis. *The FASEB Journal* 33:9131–9141.

17. Erban, R., and S. J. Chapman, 2019. Stochastic Modelling of Reaction–Diffusion Processes. Cambridge University Press.

18. Pavliotis, G., 2014. Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations. Springer, New York, NY.

19. Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013. Bayesian Data Analysis. Chapman and Hall/CRC, 3rd edition.

20. Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah, 2017. Julia: A fresh approach to numerical computing. *SIAM Review* 59:65–98.

21. Price, L. F., C. C. Drovandi, A. Lee, and D. J. Nott, 2017. Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics* 27:1–11.

22. Priddle, J. W., S. A. Sisson, D. T. Frazier, and C. Drovandi, 2019. Efficient Bayesian synthetic likelihood with whitening transformations. arXiv:1909.04857.

23. Picchini, U., U. Simola, and J. Corander, 2020. Adaptive MCMC for synthetic likelihoods and correlated synthetic likelihoods. arXiv:2004.04558.

24. Hainy, M., C. C. Drovandi, and J. M. McGree, 2016. Likelihood-free extensions for Bayesian sequentially designed experiments. *In* mODa 11 - Advances in Model-Oriented Design and Analysis, Springer International Publishing, 153–161.

25. Del Moral, P., A. Doucet, and A. Jasra, 2012. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing* 22:1009–1020.

26. Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface* 6:187–202.

27. Prescott, T. P., and R. E. Baker, 2020. Multifidelity approximate Bayesian computation with sequential Monte Carlo parameter sampling. arXiv:2001.06256v1.

28. Akiyama, M., T. Sushida, S. Ishida, and H. Haga, 2017. Mathematical model of collective cell migrations based on cell polarity. *Development, Growth & Differentiation* 59:471–490.

29. Vanegas-Acosta, J., D. Garzón-Alvarado, and A. Zwamborn, 2012. Mathematical model of electrotaxis in osteoblastic cells. *Bioelectrochemistry* 88:134–143.

30. Binny, R. N., A. James, and M. J. Plank, 2016. Collective cell behaviour with neighbour-dependent proliferation, death and directional bias. *Bulletin of Mathematical Biology* 78:2277–2301.

31. Warne, D. J., R. E. Baker, and M. J. Simpson, 2017. Optimal quantification of contact inhibition in cell populations. *Biophysical Journal* 113:1920–1924.

462 32. Prescott, T. P., and R. E. Baker, 2020. Multifidelity approximate Bayesian computation. *SIAM/ASA Journal on Uncertainty*
463 *Quantification* 8:114–138.

464 33. Zhang, Y., G. Xu, R. M. Lee, Z. Zhu, J. Wu, S. Liao, G. Zhang, Y. Sun, A. Mogilner, W. Losert, T. Pan, F. Lin, Z. Xu, and
465 M. Zhao, 2017. Collective cell migration has distinct directionality and speed dynamics. *Cellular and Molecular Life*
466 *Sciences* 74:3841–3850.

467 34. Mesbah, A., 2016. Stochastic model predictive control: an overview and perspectives for future research. *IEEE Control*
468 *Systems* 36:30–44.

# LIST OF FIGURES

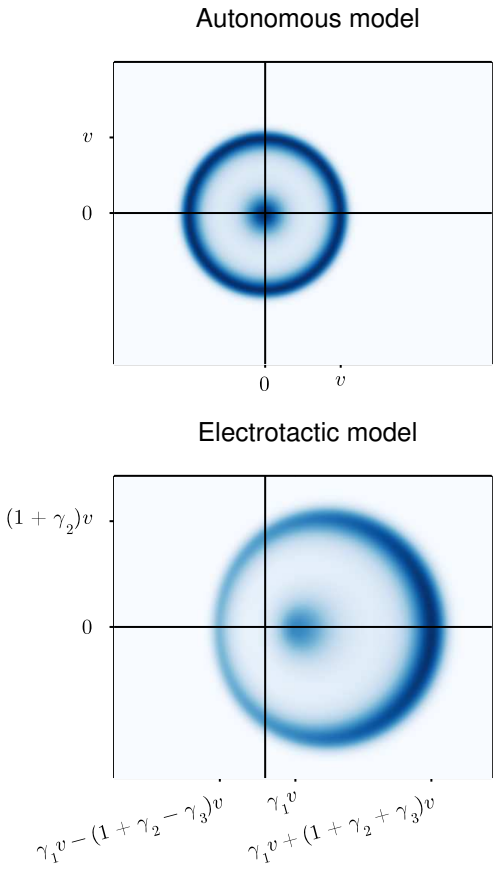


Figure 1: Comparison of the stationary distributions for the random velocity, **v**, under the autonomous and electrotactic models, where darker regions correspond to greater probability. The bottom plot shows the hypothesised electrotactic effects of an EF, applied in the positive $x$ direction, parametrised by $\gamma_1, \ldots, \gamma_4$. The effects of $\gamma_1$, $\gamma_2$ and $\gamma_3$ are visible in the shape of the distribution. *Polarity bias* ($\gamma_4$) produces asymmetry in the distribution density, shown as a darker region to the right of the figure.

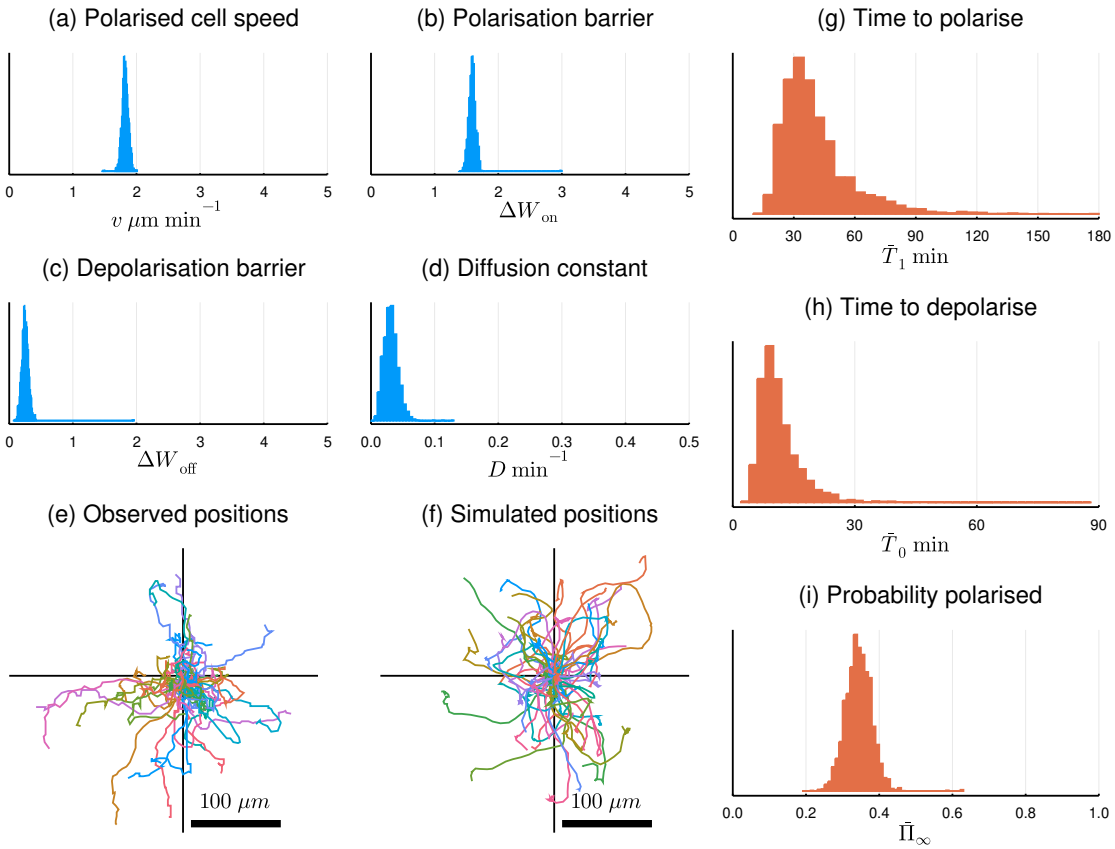


Figure 2: Parameter inference and simulation of autonomous model. (a–d) All one-dimensional projections of the posterior sample from $\pi(\theta_{\text{NoEF}} \mid \mathbf{x}_{\text{NoEF}})$, with axes scaled to the support of the prior. The covariance structure of the posterior is given in the *Supplementary Material*. (e–f) Comparison of autonomous data and simulations of autonomous model in Eq. (2). (f) Simulations use parameters $v = 1.82$ µm min$^{-1}$, $\Delta W_{\text{on}} = 1.59$, $\Delta W_{\text{off}} = 0.25$, and $D = 0.031$ min$^{-1}$. (g–i) Posterior predictive samples for $\bar{T}_1$ (expected time to polarisation), $\bar{T}_0$ (expected time to depolarisation), and $\bar{\Pi}_\infty$ (probability of a simulated cell being polarised) under the posterior, $\pi(\theta_{\text{NoEF}} \mid \mathbf{x}_{\text{NoEF}})$.

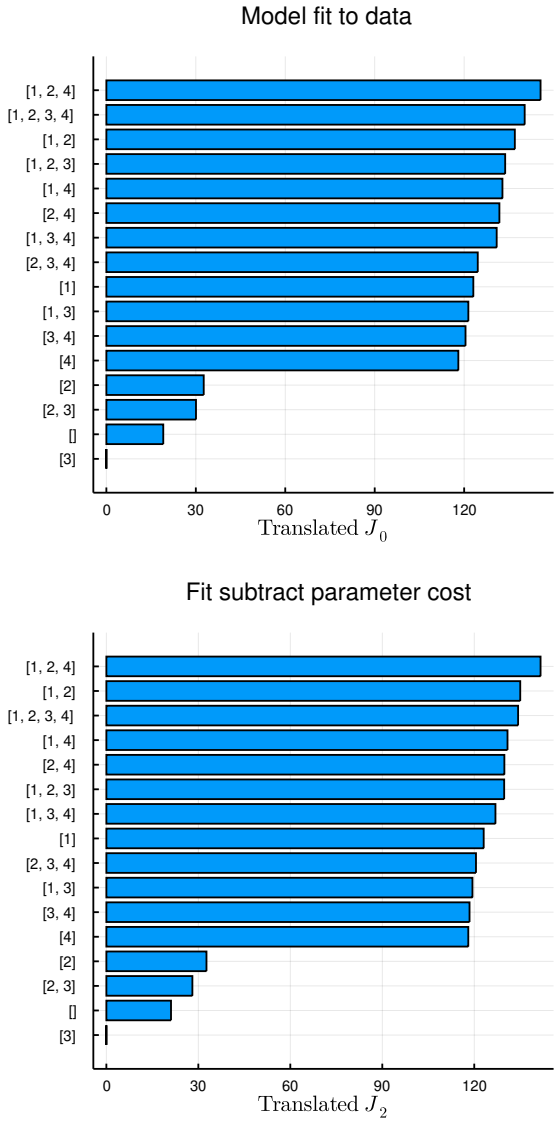


Figure 3: Objective functions $J_0(X)$ and $J_2(X)$ from Eq. (4), for combinations of electrotactic effects indexed by $X \subseteq \{1, 2, 3, 4\}$. Greater values are preferred. Each objective function is translated to have zero minimum value.

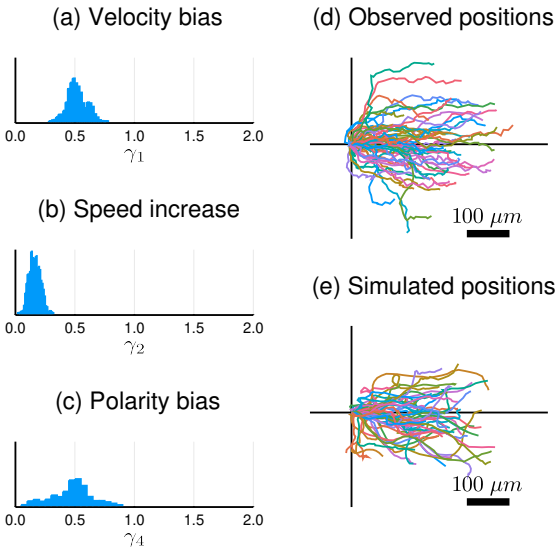


Figure 4: Parameter inference and simulation of electrotactic model. (a–c) One-dimensional projections for $\gamma_1$, $\gamma_2$, and $\gamma_4$ of the posterior sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, with axes scaled to the support of the prior. The covariance structure of the full posterior is given in the *Supplementary Material*. (d–e) Comparison of the electrotactic experimental data, $\mathbf{x}_{\text{EF}}$, and simulations of the electrotactic model in Eq. (3). (d) Experimentally observed trajectories $\mathbf{x}_{\text{EF},1}, \ldots, \mathbf{x}_{\text{EF},50}$. (e) Fifty simulations from model in Eq. (3) with parameter values $v = 1.84$ µm min$^{-1}$, $\Delta W_{\text{on}} = 1.51$, $\Delta W_{\text{off}} = 0.29$, $D = 0.029$ min$^{-1}$, $\gamma_1 = 0.52$, $\gamma_2 = 0.16$, $\gamma_3 = 0$, and $\gamma_4 = 0.47$.

# Supplementary Material

## Quantifying the impact of electric fields on single-cell motility

Thomas P Prescott  Kan Zhu  Min Zhao  Ruth E Baker

## 1 Mathematical model of electrotaxis

The autonomous model of cellular velocity in Eq. (2) is given by

$$\mathbf{v}(t) = v\mathbf{p}(t), \tag{2a}$$

$$\mathrm{d}\mathbf{p}(t) = -D\nabla W(\mathbf{p}(t))\ \mathrm{d}t + \sqrt{2D}\ \mathrm{d}\mathbf{B}, \tag{2b}$$

and the electrotactic model in Eq. (3) is given by

$$\mathbf{v}(t) = \mathbf{v}_{\text{cell}} + \mathbf{v}_{\text{EF}}, \tag{3a}$$

$$\mathbf{v}_{\text{EF}}(t) = \gamma_1 v\mathbf{u}(t), \tag{3b}$$

$$\mathbf{v}_{\text{cell}}(t) = (1 + \gamma_2 u(t) + \gamma_3 \mathbf{u}(t)\cdot\hat{\mathbf{p}}(t))\, v\mathbf{p}(t), \tag{3c}$$

$$\mathrm{d}\mathbf{p}(t) = -D\left(\nabla W(\mathbf{p}(t)) - \gamma_4\mathbf{u}(t)\right)\ \mathrm{d}t + \sqrt{2D}\ \mathrm{d}\mathbf{B}. \tag{3d}$$

Here, velocity is denoted by $\mathbf{v}$ and cell polarity is denoted by $\mathbf{p}$. The vector $\mathbf{u}$ is the non-dimensionalised EF with magnitude $\|\mathbf{u}\| = u$, scaled such that $u = \alpha$ represents a field of strength $200\alpha$ mV mm$^{-1}$. The two-dimensional standard Wiener process is denoted by $\mathbf{B}$, and $\hat{\mathbf{p}}$ is the unit vector in the direction of polarity. Both models depend on the parameters $v$, with units µm min$^{-1}$, and $D$, with units min$^{-1}$. Thus $\mathbf{p}$ is a non-dimensional quantity. The additional parameters, $\gamma_1, \ldots, \gamma_4$, in the electrotactic model parametrise the four hypothesised electrotactic effects, as described in the main text.

Also common to both models is the potential function $W(\mathbf{p})$. This function is defined to capture the intended features of the autonomous model, namely that cells stochastically, spontaneously switch between polarised and depolarised states, and that the direction of the polarity is uniformly distributed at stationarity. Denoting $p = |\mathbf{p}|$, it follows from the latter requirement that the potential function $W(\mathbf{p}) = W(p)$ must be radially symmetric. The interpretation of the parameter $v$ as the modal speed of a polarised cell also implies that the polarised and depolarised states are characterised by $\mathbf{p}$ stochastically switching between the regimes $p \approx 0$ and $p \approx 1$. We therefore require a potential function with local minima at $p = 0$ and $p = 1$. We also require both a local maximum at $p = \bar{p}$ for $\bar{p} \in (0, 1)$, and $W \to \infty$ as $p \to \infty$, for these two minima to define two 'wells' where $\mathbf{p}(t)$ is found with the highest probability. Following [11], this function is implemented as

$$W(p) = \beta\left(\frac{1}{6}p^6 - \frac{1}{4}\left(\bar{p}^2 + 1\right)p^4 + \frac{1}{2}\bar{p}^2p^2\right), \tag{S1}$$

where the parameter values $\bar{p} \in (0, 1)$ and $\beta > 0$ define the local minimum values of the wells at $p = 0$ and $p = 1$.

It can be shown [17, 18] that the rates at which the polarity stochastically switches between the two wells are determined solely by the timescale parameter, $D$ min$^{-1}$, and the non-dimensional values of the two *energy barriers*

of the potential function,

$$\Delta W_{\text{on}} = W(\bar{p}) - W(0) = \beta(3\bar{p}^4 - \bar{p}^6)/12, \tag{S2a}$$

$$\Delta W_{\text{off}} = W(\bar{p}) - W(1) = \beta(1 - \bar{p}^2)^3/12, \tag{S2b}$$

which represent the changes in potential, $W$, between the minima at $p = 0$ and $p = 1$ and the local maximum at $p = \bar{p}$. These parameters are sufficient to specify $\beta$ and $\bar{p}$ in Eq. (S1), and thus uniquely determine the potential function, $W$. Hence, we calibrate the models in Eq. (2) and Eq. (3) by inferring the common parameters, $v$, $D$, and $\Delta W_{\text{on/off}}$, together with the parameters $\gamma_1, \ldots, \gamma_4$ specific to the electrotactic model.

**Note on polarity definition** Our description of the model interprets the variable $\mathbf{p}$ as the cell polarity, and treats velocity as the combination of a polarity component and a component due to the EF. Another interpretation of $\mathbf{p}$ is available if we specifically define single-cell polarity as the non-dimensionalisation of the velocity by $v$. In the electrotactic model, this alternative definition identifies cell polarity as the variable

$$\mathbf{v}/v = (1 + \gamma_2 u + \gamma_3 \mathbf{u} \cdot \hat{\mathbf{p}})\mathbf{p} + \gamma_1 \mathbf{u}.$$

The variable $\mathbf{p}$, with dynamics (3d), is then interpreted as a slowly-responding component of the cell polarity (in the alternative definition) to the EF input, while $\gamma_1 \mathbf{u}$, identifiable with *velocity bias*, is an instantly-responding component of the cell polarity. However, these definitions are internal to the model, in the sense that they have no effect on the observable position or velocity of simulated cells. Thus, in the current work, we choose to identify 'cell polarity' as the modelled variable $\mathbf{p}$, while noting that alternative interpretations are possible.

## 2 Likelihood-free Bayesian inference

The Bayesian inference framework uses the experimental data, $\mathbf{x}_{\text{NoEF}}$ and $\mathbf{x}_{\text{EF}}$, to update a prior distribution, $\pi(\theta)$, into a posterior distribution, $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, by multiplying by the *likelihood*, $\mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)$, according to Bayes' rule,

$$\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}}) \propto \mathcal{L}(\mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}} \mid \theta)\pi(\theta).$$

To define the likelihood, we first consider simulations of the models in Eq. (2) and Eq. (3).

For a given parameter vector, $\theta$, initial polarity, $\mathbf{p}_0$, and non-zero EF input, $\mathbf{u}(t)$, the model in Eq. (3) is simulated and a trajectory, $\omega = (\mathbf{p}(t), \mathbf{x}(t))_{t \geq 0}$, is produced. This stochastic trajectory has conditional density $p(\omega \mid \theta, \mathbf{p}_0, \mathbf{u}(t))$. We assume that there is a known distribution, $\varphi(\mathbf{p}_0)$, for the initial polarity: for the inference procedure carried out in the main text, we assume that $\varphi$ is a Gaussian distribution with zero mean and diagonal covariance matrix, with component-wise variances of 0.1. For the two specific experimental inputs, $\mathbf{u}(t) \equiv 0$ and $\mathbf{u}(t) \equiv \mathbf{i}$, we integrate the density $p$ with respect to $\varphi(\mathbf{p}_0)$ and thus define two densities,

$$p_{\text{NoEF}}(\omega \mid \theta) = \int p(\omega \mid \theta, \mathbf{p}_0, \mathbf{u}(t) \equiv 0)\varphi(\mathbf{p}_0)\, \mathrm{d}\mathbf{p}_0, \tag{S3}$$

$$p_{\text{EF}}(\omega \mid \theta) = \int p(\omega \mid \theta, \mathbf{p}_0, \mathbf{u}(t) \equiv \mathbf{i})\varphi(\mathbf{p}_0)\, \mathrm{d}\mathbf{p}_0, \tag{S4}$$

for trajectories simulated by the autonomous and electrotactic models, respectively. Each observed trajectory in

the experimental data set, $\mathbf{x}_{\mathrm{NoEF},i}$ and $\mathbf{x}_{\mathrm{EF},i}$, thus defines a set in the simulation space,

$$\Omega(\mathbf{x}_{\mathrm{NoEF},i}) = \left\{\omega = (\mathbf{x}(t), \mathbf{p}(t))_{t\geq 0} \;:\; \mathbf{x}(t_j) = \mathbf{x}_{\mathrm{NoEF},i}(t_j)\right\},$$
$$\Omega(\mathbf{x}_{\mathrm{EF},i}) = \left\{\omega = (\mathbf{x}(t), \mathbf{p}(t))_{t\geq 0} \;:\; \mathbf{x}(t_j) = \mathbf{x}_{\mathrm{EF},i}(t_j)\right\},$$

of all simulated trajectories that are indistinguishable from the observed data. We thus define the likelihoods of each simulation as

$$\mathcal{L}_{\mathrm{NoEF}}(\mathbf{x}_{\mathrm{NoEF},i} \mid \theta) = \int_{\Omega(\mathbf{x}_{\mathrm{NoEF},i})} p_{\mathrm{NoEF}}(\omega \mid \theta)\, \mathrm{d}\omega,$$
$$\mathcal{L}_{\mathrm{EF}}(\mathbf{x}_{\mathrm{EF},i} \mid \theta) = \int_{\Omega(\mathbf{x}_{\mathrm{EF},i})} p_{\mathrm{EF}}(\omega \mid \theta)\, \mathrm{d}\omega,$$

for the cell indices $i = 1, \ldots, 50$. The likelihoods of each trajectory thus combine to give the posterior,

$$\begin{aligned}\pi(\theta \mid \mathbf{x}_{\mathrm{NoEF}}, \mathbf{x}_{\mathrm{EF}}) &\propto \mathcal{L}(\mathbf{x}_{\mathrm{NoEF}}, \mathbf{x}_{\mathrm{EF}} \mid \theta)\pi(\theta) \\ &= \prod_{i=1}^{50} \mathcal{L}_{\mathrm{NoEF}}(\mathbf{x}_{\mathrm{NoEF},i} \mid \theta)\mathcal{L}_{\mathrm{EF}}(\mathbf{x}_{\mathrm{EF},i} \mid \theta)\pi(\theta). \end{aligned} \tag{S5}$$

However, it is clear that the likelihood of each of the experimentally observed trajectories cannot easily be calculated. We therefore identified the posterior parameter distribution using a likelihood-free (i.e. simulation-based) Bayesian inference approach, harnessing the concept of *synthetic likelihoods*.

## 2.1 Synthetic likelihoods

We focus on the autonomous case first; the electrotactic case follows in exactly the same way, albeit with an obvious change of notation. The synthetic likelihood approach approximates the likelihoods, $\mathcal{L}_{\mathrm{NoEF}}(\mathbf{x}_{\mathrm{NoEF},i} \mid \theta)$ for $i = 1, \ldots, 50$, in two stages. The first stage is to reduce the dimension of the data space by defining a function of the simulated and observed trajectories that maps the data to a low-dimensional summary statistic. The second stage is to (a) use repeated simulation of the summarised model at the parameter value $\theta$ to fit an empirical multivariate Gaussian distribution for the summary statistic, and then (b) to approximate the likelihood with the *synthetic likelihood* of the experimental data, defined as the likelihood of the summarised data under the fitted empirical Gaussian distribution.

We define the function $Y : \omega \mapsto \mathbb{R}^4$ as:

$$Y_1(\omega) = \|\mathbf{x}(180)\|, \tag{S6a}$$
$$Y_2(\omega) = \frac{1}{36}\sum_{j=1}^{36} \|\mathbf{x}(t_j) - \mathbf{x}(t_{j-1})\|, \tag{S6b}$$
$$Y_3(\omega) = \left(\frac{1}{36}\sum_{j=1}^{36} \|\mathbf{x}(t_j) - \mathbf{x}(t_{j-1}) - Y_2(\omega)\|^2\right)^{1/2}, \tag{S6c}$$
$$Y_4(\omega) = \arccos\left(\frac{\mathbf{i}\cdot\mathbf{x}(180)}{Y_1(\omega)}\right), \tag{S6d}$$

for sample time points $t_j = 5j$ min. Thus, the entries of $Y(\omega)$ denote the random values of

- the net displacement,
- the mean and standard deviation of the displacements over five-minute intervals,

- and the angle between the final position and the positive $x$ axis,

for stochastic simulations $\omega$ of the electrotactic model in Eq. (3), given $\theta$, $\mathbf{p}_0$, and $\mathbf{u}(t)$. Note that we can also calculate the values of the function $Y$ in Eq. (S6) for the experimentally observed data, $\mathbf{x}_{\mathrm{NoEF},i}$ and $\mathbf{x}_{\mathrm{EF},i}$, for $i = 1, \ldots, 50$. With a slight abuse of notation, we denote the resulting summarised experimental data by $y_{\mathrm{NoEF},i} = Y(\mathbf{x}_{\mathrm{NoEF},i})$ and $y_{\mathrm{EF},i} = Y(\mathbf{x}_{\mathrm{EF},i})$, respectively.

For a given value of $\theta$, the synthetic likelihood approach [21–23] assumes that the random value of $Y(\omega)$ under the density $p_{\mathrm{NoEF},i}(\omega \mid \theta)$ is a Gaussian random variable with parameter-dependent mean $\mu_{\mathrm{NoEF}}(\theta)$ and covariance $\Sigma_{\mathrm{NoEF}}(\theta)$. We estimate this mean and covariance with the sample mean and covariance of simulated summary statistics $Y(\omega_k)$, for $k = 1, \ldots, n$, produced by simulating the autonomous model $n$ times using the parameter value $\theta$. The resulting approximation of each trajectory's likelihood, $\tilde{\mathcal{L}}_{\mathrm{NoEF},n} \approx \mathcal{L}_{\mathrm{NoEF}}$, is summarised as

$$\tilde{\mathcal{L}}_{\mathrm{NoEF},n}(\mathbf{x}_{\mathrm{NoEF},i} \mid \theta) = \mathcal{N}\left(y_{\mathrm{NoEF},i} \mid \hat{\mu}_{\mathrm{NoEF}}(\theta), \hat{\Sigma}_{\mathrm{NoEF}}(\theta)\right) \quad i = 1, \ldots, 50, \tag{S7a}$$

$$\hat{\mu}_{\mathrm{NoEF}}(\theta) = \frac{1}{n}\sum_{k=1}^{n} Y(\omega_k), \tag{S7b}$$

$$\hat{\Sigma}_{\mathrm{NoEF}}(\theta) = \frac{1}{n}\sum_{k=1}^{n} (Y(\omega_k) - \hat{\mu}_{\mathrm{NoEF}}(\theta))(Y(\omega_k) - \hat{\mu}_{\mathrm{NoEF}}(\theta))^T, \tag{S7c}$$

$$\omega_k \sim p_{\mathrm{NoEF}}(\cdot \mid \theta) \quad k = 1, \ldots, n, \tag{S7d}$$

where $\mathcal{N}$ denotes the Gaussian density and where the chosen number of simulations, $n$, needs to be appropriately large [22]. In our case, we choose $n = 500$.

Finally, with the appropriate adaptation of the procedure above to also apply to the electrotactic model, we can multiply each these trajectory synthetic likelihoods into an overall synthetic likelihood for the experimental data,

$$L_{\mathrm{NoEF},n}(\theta) = \prod_{i=1}^{50} \tilde{\mathcal{L}}_{\mathrm{NoEF},n}(\mathbf{x}_{\mathrm{NoEF},i} \mid \theta), \tag{S8a}$$

$$L_{\mathrm{NoEF},n}(\theta) = \prod_{i=1}^{50} \tilde{\mathcal{L}}_{\mathrm{EF},n}(\mathbf{x}_{\mathrm{EF},i} \mid \theta), \tag{S8b}$$

$$\mathcal{L}(\mathbf{x}_{\mathrm{NoEF}}, \mathbf{x}_{\mathrm{EF}} \mid \theta) \approx L_n(\theta) = L_{\mathrm{NoEF},n}(\theta) L_{\mathrm{EF},n}(\theta), \tag{S8c}$$

each calculation of which requires $n$ simulations of the autonomous model and $n$ of the the electrotactic model.

## 2.2 SMC and sequential inference

In order to produce a sample from the posterior distribution, we use sequential Monte Carlo (SMC) with synthetic likelihoods [21–27], as outlined in Algorithm 1. This method is chosen in order to exploit parallelisation, mitigating the computational burden of MCMC-based approaches that is incurred due to the large numbers of model simulations required for accurate likelihood-free inference. SMC defines a sequence of intermediate importance distributions that evolve towards the target posterior. This approach is particularly useful in comparison to naive rejection sampling: since we will use non-informative priors, rejection sampling is too inefficient, as it proposes parameters in extremely low-likelihood regions of parameter space too frequently.

In Algorithm 1 we produce a weighted sample from the Bayesian synthetic likelihood approximation to the posterior, $\pi(\theta \mid \mathbf{x}_{\mathrm{NoEF}}, \mathbf{x}_{\mathrm{EF}})$. The intermediate distributions at each iteration are proportional to the tempered distributions

$$\pi_{T_1,T_2}(\theta) \propto L_{\mathrm{NoEF},n}(\theta)^{T_1} L_{\mathrm{EF},n}(\theta)^{T_2} \pi(\theta),$$

**Algorithm 1** Sequential Inference: Synthetic Likelihood SMC

**Input:** Observed summary statistics $y_{\text{NoEF}}$ and $y_{\text{EF}}$; prior $\pi$; perturbation kernel $K(\cdot \mid \theta)$; .

**Output:** Weighted sample set of parameters $\theta_i$ with weights $W_i$, from the synthetic likelihood approximation to the posterior $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$.

Sample $N$ independent $\theta_i$ from $\pi$.
Set weights $W_i^0 = 1/N$ for $i = 1, \ldots, N$.
Initialise $T_1 = 0$, $T_2 = 0$ and $r = 0$.
**repeat**
  Update $r \leftarrow r + 1$.
  Find $\Delta T_1 \in [\Delta T_{\min}, 1 - T_1]$ to solve $ESS(\{W_i^r\}) = \alpha ESS(\{W_i^{r-1}\})$, for weights $W_i^r$ such that

$$\log W_i^r = \log W_i^{r-1} + \Delta T_1 \log L_{\text{NoEF},n}(\theta_i),$$

  for the synthetic likelihoods, $L_{\text{NoEF},n}(\theta_i)$. Use $\Delta T_1 = \Delta T_{\min}$ or $\Delta T_1 = 1 - T_1$ if $ESS(\{W_i^r\})$ is, respectively, uniformly less than or uniformly greater than $\alpha ESS(\{W_i^{r-1}\})$ on the interval $[\Delta T_{\min}, 1 - T_1]$.
  Update $T_1 \leftarrow T_1 + \Delta T_1$.
  **if** $ESS(\{W_i^r\}) < N_{\min}$ **then**
    Resample from $\{\theta_i\}$ according to weights $W_i^r$.
    Reset weights $W_i^r = 1/N$.
  **end if**
  Update parameters $\theta_i \leftarrow \theta_i^\star$ to perturbed values, sampled as $\theta_i^\star \sim K(\ \mid \theta_i)$.
**until** $T_1 = 1$.
**repeat**
  Update $r \leftarrow r + 1$.
  Find $\Delta T_2 \in [\Delta T_{\min}, 1 - T_2]$ to solve $ESS(\{W_i^r\}) = \alpha ESS(\{W_i^{r-1}\})$, for weights $W_i^r$ such that

$$\log W_i^r = \log W_i^{r-1} + \Delta T_2 \log L_{\text{EF},n}(\theta_i),$$

  for the synthetic likelihoods, $L_{\text{EF},n}(\theta_i)$. Use $\Delta T_2 = \Delta T_{\min}$ or $\Delta T_2 = 1 - T_2$ if $ESS(\{W_i^r\})$ is, respectively, uniformly less than or uniformly greater than $\alpha ESS(\{W_i^{r-1}\})$ on the interval $[\Delta T_{\min}, 1 - T_2]$.
  Update $T_2 \leftarrow T_2 + \Delta T_2$.
  **if** $ESS(\{W_i^r\}) < N_{\min}$ **then**
    Resample from $\{\theta_i\}$ according to weights $W_i^r$.
    Reset weights $W_i^r = 1/N$.
  **end if**
  Update parameters $\theta_i \leftarrow \theta_i^\star$ to perturbed values, sampled as $\theta_i^\star \sim K(\ \mid \theta_i)$.
**until** $T_2 = 1$.

where the sequence of temperatures $(T_1, T_2)$ evolves from $(0, 0)$ to $(1, 1)$ such that first $T_1$ increases to 1 with $T_2$ held at 0, and then $T_2$ increases to 1 with $T_1$ held at 1. In Algorithm 1, we define $K(\cdot \mid \theta)$ to be a multivariate Gaussian density with mean $\theta$ and diagonal covariance matrix, with component-wise variances of 0.001. The effective sample size, $ESS$, used to adaptively choose the increment in temperature, is defined as

$$ESS(\{W_i\}) = \left(\sum_i W_i\right)^2 / \sum_j W_j^2$$

for any finite set of sample weights, $W_i$.

Note that, due to the sequence in which the intermediate distributions are produced, we can terminate Algorithm 1 early, where $T_1 = 1$ and where $T_2 = 0$ (i.e. at line 13). The sample resulting from this early termination is a weighted sample from the intermediate distribution,

$$\pi_{1,0}(\theta) \propto L_{\mathrm{NoEF},n}(\theta)\pi(\theta),$$

which approximates the posterior $\pi(\theta \mid \mathbf{x}_{\mathrm{NoEF}})$ conditioned on the autonomous experiment alone. The results depicted in Figure 2(a–d), in the main text, are the result of such an early termination. The covariance structure of this sample is depicted in Figure S1.

The intermediate sample was used as a common starting point for sixteen subsequent runs of Algorithm 1 for each of the priors $\pi_X$, all starting at line 14 of Algorithm 1 with $T_1 = 1$ and $T_2 = 0$ and evolving to $T_1 = T_2 = 1$. The output of the completed algorithm, for the prior $\pi_X$ using the subset $X = \{1, 2, 4\}$, is depicted in Figure S2. The comparison of the marginals of the common parameter set, $v$, $\Delta W_{\mathrm{on/off}}$, and $D$, between the early termination and the completed algorithms is given in Figure S3.

# 3 Simulated behaviour from dynamic inputs

## 3.1 Summary statistics

Recall that the polarity, $\mathbf{p}(t)$, evolves randomly from initial value $\mathbf{p}_0$. We will use the polarity, $\mathbf{p}(t)$, to define four possible cell states: depolarised, denoted $\Omega_0$; polarised in the positive $x$ direction, denoted $\Omega_+$; polarised in the negative $x$ direction, denoted $\Omega_-$; and polarised perpendicular to the $x$ axis, denoted $\Omega_\perp$. We further coarse-grain the polarity by defining a cell as polarised, denoted $\Omega_1$, if it is not depolarised. Here, the sets $\Omega_\star$ are defined by the following partition of the state space of polarity vectors, $\mathbf{p} = [p_x, p_y]$, such that

$$\begin{aligned}
\Omega_0 &= \{\mathbf{p} \in \mathbb{R}^2 \mid \|\mathbf{p}\| < \bar{p}\},\\
\Omega_+ &= \{\mathbf{p} \in \Omega_1 \mid p_x > |p_y|)\},\\
\Omega_- &= \{\mathbf{p} \in \Omega_1 \mid p_x < -|p_y|)\},\\
\Omega_\perp &= \{\mathbf{p} \in \Omega_1 \mid |p_x| < |p_y|\},
\end{aligned}$$

where $\Omega_1 = \Omega_0^C = \Omega_+ \cup \Omega_- \cup \Omega_\perp$ and $\bar{p}$ is the polarity threshold defined by the energy barriers, $\Delta W_{\mathrm{on/off}}$, according to Eq. (S2). These four coarse-grained states are depicted in Figure S4.

We use the following summary statistics, defined by these coarse-grained polarisation states, to analyse model simulations, $\omega = (\mathbf{x}(t), \mathbf{p}(t))_{t \geq 0}$. Each summary statistic condenses $\omega$ into a time or a binary value. We first define the *hitting times*, $T_\star(\omega)$, as the earliest time $t \geq 0$ at which $\mathbf{p}(t) \in \Omega_\star$, which may be zero if $\mathbf{p}(0) \in \Omega_\star$ for the given $\omega$. Thus, $T_0$ is the first time at which the simulated cell becomes depolarised. Similarly, $T_+$ is the first time at which the simulated cell becomes polarised in the positive $x$ direction, and so on. For the fixed time point, $T$, we

also define the binary indicator $\Pi_T(\omega)$ to take the value 1 if the cell is polarised at time $T$ (i.e. if $\mathbf{p}(T) \in \Omega_1$), and 0 otherwise. Finally, we define the binary indicator $P_{\perp\to-}(\omega)$ to take the value 1 if the cell is polarised perpendicular to the $x$ axis *before* it is first polarised in the negative $x$ direction (i.e. if $T_\perp(\omega) < T_-(\omega)$), and 0 otherwise. The resulting summary statistics are given by

$$\begin{aligned}
T_0(\omega) &= \inf\{t \geq 0 \ : \ \mathbf{p}(t) \in \Omega_0\},\\
T_1(\omega) &= \inf\{t \geq 0 \ : \ \mathbf{p}(t) \in \Omega_1\},\\
T_-(\omega) &= \inf\{t \geq 0 \ : \ \mathbf{p}(t) \in \Omega_-\},\\
T_\perp(\omega) &= \inf\{t \geq 0 \ : \ \mathbf{p}(t) \in \Omega_\perp\},\\
\Pi_T(\omega) &= \mathbb{I}(\mathbf{p}(T) \in \Omega_1),\\
P_{\perp\to-}(\omega) &= \mathbb{I}(T_\perp(\omega) < T_-(\omega)).
\end{aligned}$$

Note that, unlike the summary statistics, $Y$, defined in Eq. (S6), we cannot apply these functions to the experimental data since cell polarity cannot be directly measured.

Finally, each of the summary statistics can be averaged across many simulations: the averaged value is notated with a bar, to give the conditional expectation of the summary statistic for a given parameter value and for a given initial condition (or distribution of initial conditions). Thus $\bar{T}_0 = \mathbb{E}(T_0 \mid \theta, \mathbf{p}_0)$ is the average time a simulated cell takes to depolarise for parameter values $\theta$ and initial polarity value $\mathbf{p}_0$, and so on. Note that $\bar{\Pi}_T$ is then the probability that a simulated cell is polarised at time $T$. Similarly, $\bar{P}_{\perp\to\infty}$ is the probability that a simulated cell is polarised perpendicular to the $x$ axis before first being polarised in the negative $x$ direction. Thus, posterior uncertainty in $\theta$ propagates to uncertainty in the conditional expectations, allowing us to interpret parametric uncertainty in terms of simulated model outputs.

## 3.2 Simulation results

We simulated the model under two new dynamic inputs: a *switching* field, $\mathbf{u}_{\text{switch}}(t)$, and a *stopping* field, $\mathbf{u}_{\text{stop}}(t)$. For $0 \leq t \leq 90$ min, both inputs apply a DC EF of magnitude $200\,\text{mV}\,\text{mm}^{-1}$ in the positive $x$ direction. The switching field, $\mathbf{u}_{\text{switch}}(t)$, reverses this field over $90 < t \leq 180$ min, by applying a $200\,\text{mV}\,\text{mm}^{-1}$ DC EF in the negative $x$ direction. In contrast, the stopping field, $\mathbf{u}_{\text{stop}}$, removes the field over $90 < t \leq 180$ min.

### 3.2.1 Point estimate

Five hundred simulations of the electrotactic model in Eq. (3) were run under each input, $\mathbf{u}_{\text{switch}}$ and $\mathbf{u}_{\text{stop}}$. The parameters used in these simulations were the point estimates given by the empirical mean of the posterior sample in Figure S2, $v = 1.84\,\mu\text{m}\,\text{min}^{-1}$, $\Delta W_{\text{on}} = 1.51$, $\Delta W_{\text{off}} = 0.29$, $D = 0.029\,\text{min}^{-1}$, $\gamma_1 = 0.52$, $\gamma_2 = 0.16$, $\gamma_3 = 0$, and $\gamma_4 = 0.47$. The results of these simulations are depicted in Figures S5 and S6. For the switching input, $\mathbf{u}_{\text{switch}}$, the ensemble average position in Figure S5 shows clear migration from left to right in the first 90 minutes, which then reverses direction for the second 90 minutes. In contrast, the ensemble average position for the stopping input, $\mathbf{u}_{\text{stop}}$, moves much less in the second 90 minutes. Note that the small number of individual trajectories depicted in Figure S5 illustrate that the underlying behaviour of each agent is highly stochastic, and that the ensemble average behaviour is not reflective of each individual cell's trajectory.

Beyond predicting directly observable behaviour (i.e. cell position), the model also enables us to analyse the internal polarisation state of the cells, which cannot be as easily observed. For the 500 simulations under each dynamic EF, the fraction of simulated cells with polarity in each of the four coarse-grained states shown in Figure S4 was recorded. The trajectories of these proportions are shown in Figure S6 for each of $\mathbf{u}_{\text{switch}}$ and $\mathbf{u}_{\text{stop}}$. Under the given point estimate of the parameter values, by the switching/stopping time at $t = 90$, over 50% of cells are

depolarised. Approximately 20–25% of cells are polarised in alignment with the EF, a similar number are polarised perpendicular to the field, and the remainder are polarised counter to the field.

After the field switches direction for $t > 90$, as shown in the top plot of Figure S6, there is little change in the proportion of cells that are depolarised, or polarised perpendicular to the $x$ axis. However, over the subsequent 90 min, the distribution of cells in each state evolves to a new stationary distribution. The proportion of cells polarised in alignment with the new EF (i.e. in the negative $x$ direction) increases, and the proportion of cells polarised counter to the new direction decreases. In contrast, for the EF $\mathbf{u}_{\text{stop}}$, the distribution evolves over the subsequent 90 min horizon to a new symmetric stationary distribution, with roughly equal proportions polarised in the positive and negative $x$ directions, with a slightly higher proportion of depolarised cells.

### 3.2.2 Bayesian parameter uncertainty

The simulated dynamics in Figure S5 and Figure S6 were generated using a point estimate of each of the seven positive parameter values, and setting $\gamma_3 = 0$. However, the output of the Bayesian parameter inference procedure is an empirical sample, depicted in Figure S2, which reflects the Bayesian posterior uncertainty in these seven positive parameter values. Similarly to the analysis of the autonomous model, simulations can be used to interpret how this uncertainty propagates to uncertainty in the cell's simulated behaviour. Figure S7 depicts an estimate of the uncertainty in (a) the average time taken for a cell to depolarise, $\bar{T}_0$, (b) the average time taken for a cell to be polarised in the negative $x$ direction, $\bar{T}_-$, and (c) the probability that a cell is polarised perpendicular to the $x$ axis before becoming polarised in the negative $x$ direction, $\bar{P}_{\perp\to-}$. These posterior predictive distributions are constructed assuming a cell is initially polarised in the positive $x$ direction, and for the value of the EF input over $t > 90$; that is, for $\mathbf{u} = -\mathbf{i}$ in the case of $\mathbf{u}_{\text{switch}}$ and $\mathbf{u} = 0$ for $\mathbf{u}_{\text{stop}}$.

We observe that the uncertainty in the expected time taken for cells to depolarise, $\bar{T}_0$, is similar under both EFs: from 5.8 min to 25.2 min for the switched input (median value 10.8 min), and from 6.9 min to 28.5 min for the stopped input (median value 12.5 min). However, the distribution of values for $\bar{T}_-$ shows that cells under the switched field are significantly quicker than under the stopped field to reverse their polarisation to align with the negative $x$ axis. The expected time for a cell to reverse its polarisation under the switched field is 51 min to 227 min, with median value 93 min, while under the stopped field the expected time a cell takes to reverse its polarisation is 84 min to 346 min, with median value 147 min.

We observe in Figure S6 that the distribution of cells in each of the coarse-grained polarisation states reaches its new steady state within approximately 30 min. This is notably faster than the timescales, shown in Figure S7, of individual polarised cells switching their direction of polarity. This is likely due to the fact that the majority of cells at $t = 90$ are *not* actually polarised in the positive $x$ direction, and these are therefore more able to quickly respond to the switch in the EF. In contrast, those cells that are polarised in the positive $x$ direction are relatively slower to reverse the direction of their polarity.

Finally, an important experimental observation is that, in reversing the direction of motility under a switched input, many cells transiently travel perpendicular to the field direction [4, 8]. The value of $\bar{P}_{\perp\to-}$, depicted in Figure S7, gives the probability that a simulated cell is polarised perpendicular to the $x$ axis before reversing polarity from the positive to negative $x$ direction. Under the switched input, 77% to 88% (median value 82%) of simulated cells demonstrate this behaviour. Thus, simulations from this model replicate a characteristic feature of the cellular response to dynamic EFs, without this feature being explicitly specified in the model's design or used to calibrate the model's parameters.

# List of Figures

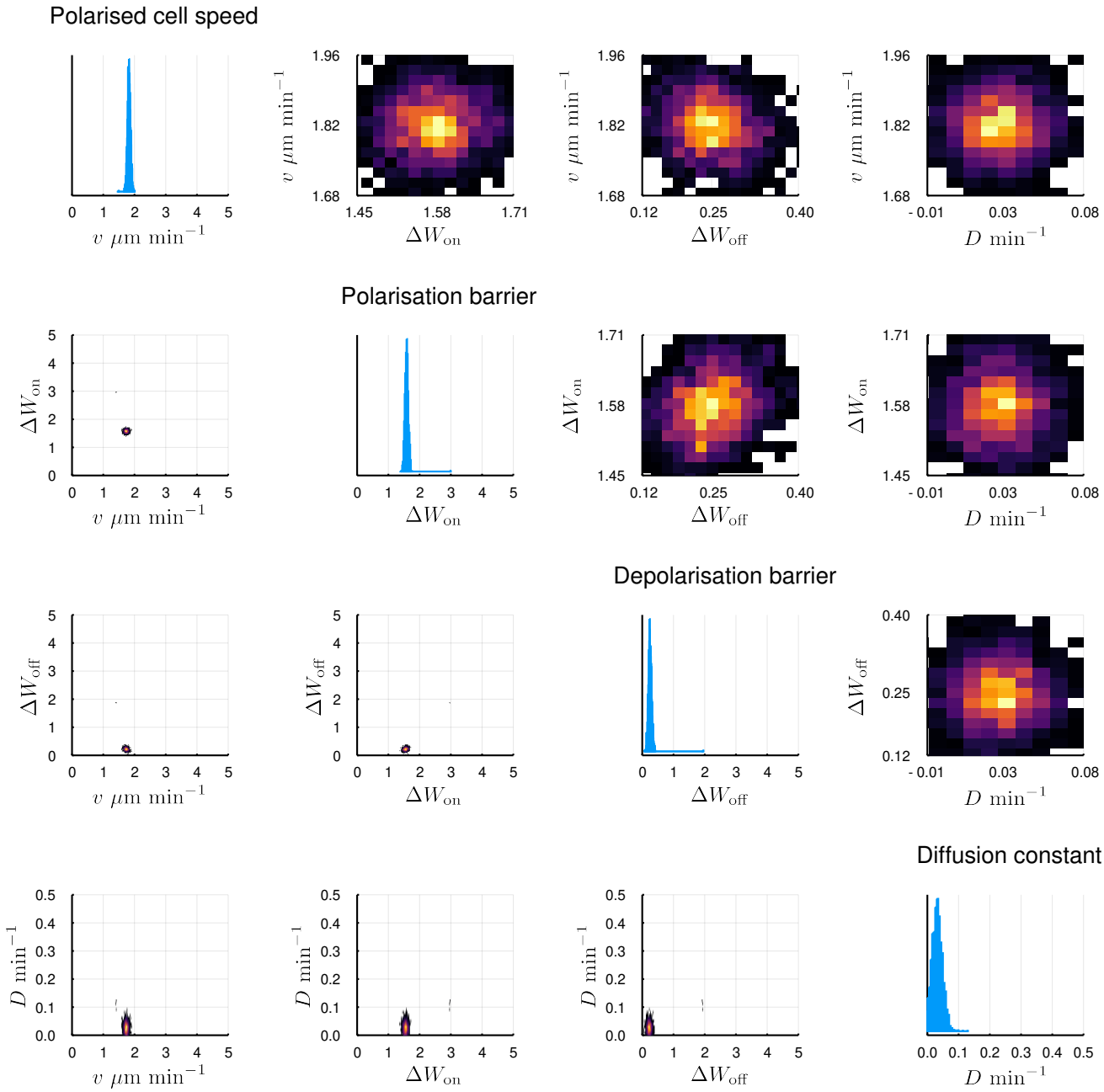


Figure S1: Weighted sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}})$ generated by the early termination of Algorithm 1 for identified parameters, $v$, $\Delta W_{\text{on}}$, $\Delta W_{\text{off}}$, and $D$. Diagonal plots are empirical histograms (as in Figure 3). Off-diagonal heatmaps represent empirical pairwise distributions, where brighter colours correspond to greater density. Axes in the bottom-left are scaled to the prior support; the top-right scales are zoomed in to the region of greatest positive posterior likelihood.

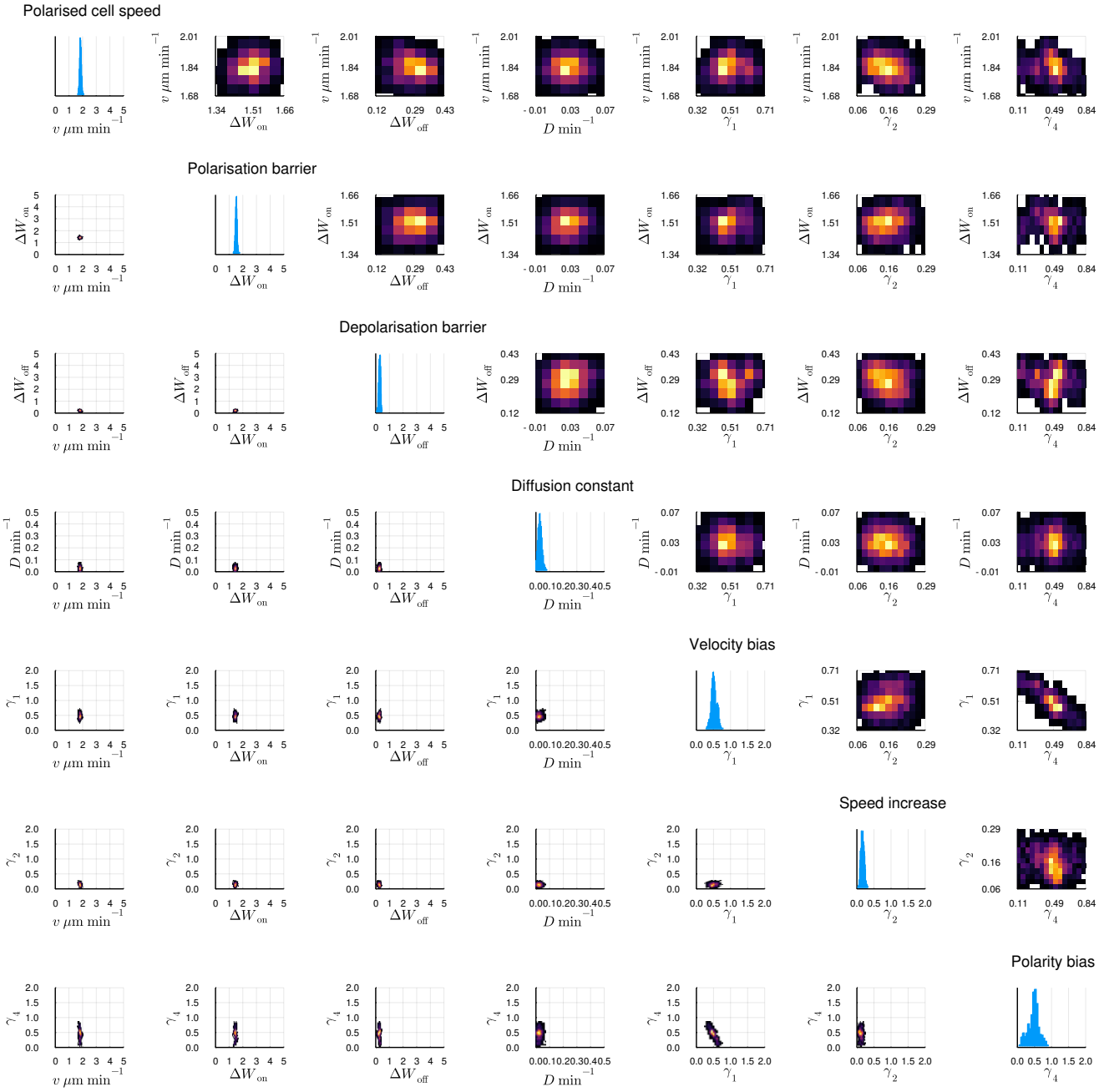


Figure S2: Weighted sample from $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$, generated by the completion of Algorithm 1 for identified parameters, $v$, $\Delta W_{\text{on}}$, $\Delta W_{\text{off}}$, $D$, $\gamma_1$, $\gamma_2$, and $\gamma_4$. Diagonal plots are empirical histograms (as in Figure S3 and Figure 7). Off-diagonal heatmaps represent empirical pairwise distributions, where brighter colours correspond to greater density. Axes in the bottom-left are scaled to the prior support; the top-right scales are zoomed in to the region of greatest positive posterior likelihood.

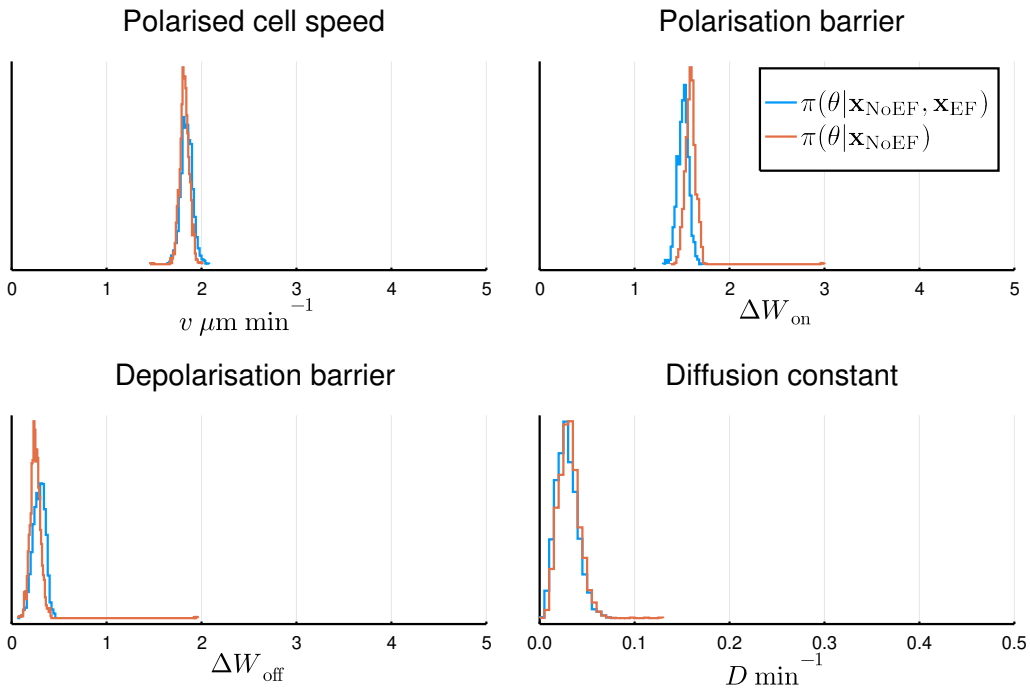


Figure S3: One-dimensional projections for $v$, $\Delta W_{\mathrm{on}}$, $\Delta W_{\mathrm{off}}$, and $D$ of the sample from the posterior, $\pi(\theta \mid \mathbf{x}_{\mathrm{NoEF}}, \mathbf{x}_{\mathrm{EF}})$, generated by the completion of Algorithm 1, with axes scaled to the support of the prior. Dotted curves replot the one-dimensional projections of the estimated intermediate posterior distribution, $\pi(\theta \mid \mathbf{x}_{\mathrm{NoEF}})$, generated by early termination of Algorithm 1, and also shown in Figure S1.

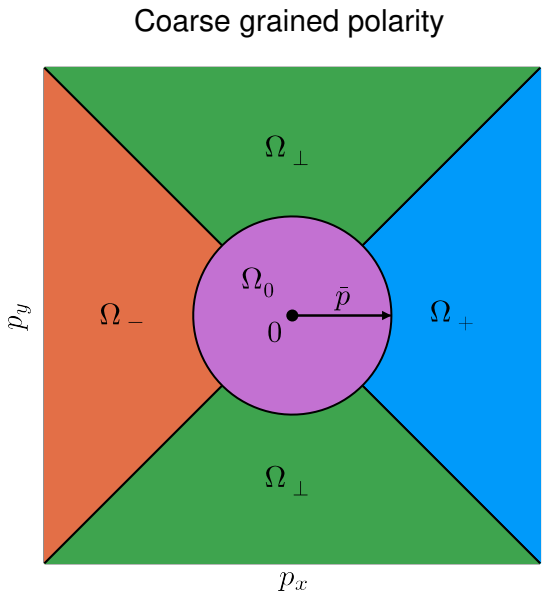


Figure S4: Coarse-grained classification of polarity, $\mathbf{p}(t)$, into one of four states: depolarised ($\Omega_0$); polarised in the positive $x$ direction ($\Omega_+$); polarised in the negative $x$ direction ($\Omega_-$); polarised perpendicular to the $x$ axis ($\Omega_\perp$).

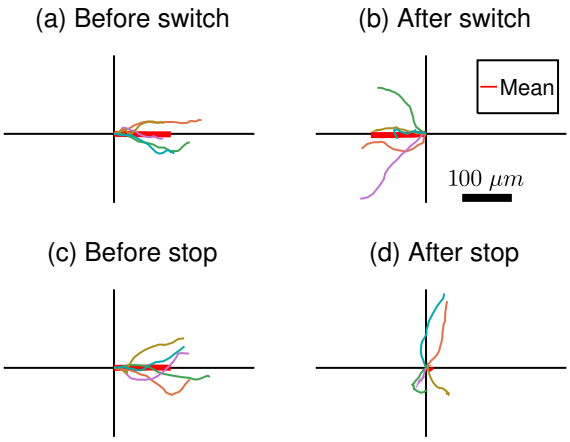


Figure S5: Model predictions of displacement, $\mathbf{x}(t)$, under (a, b) the switching EF, $\mathbf{u}_{\text{switch}}$, and (c, d) the stopping EF $\mathbf{u}_{\text{stop}}$, shown (a, c) before the step change at $t = 90\,\text{min}$, and (b, d) after the step change at $t = 90\,\text{min}$, where $\mathbf{x}(90)$ is reset to zero. Simulations are generated from the electrotactic model in Eq. (3), using the sample mean of the posterior sample depicted in Figure S2. Thicker, red curves represent the ensemble average of 500 simulations, and the other colours are five randomly selected trajectories.

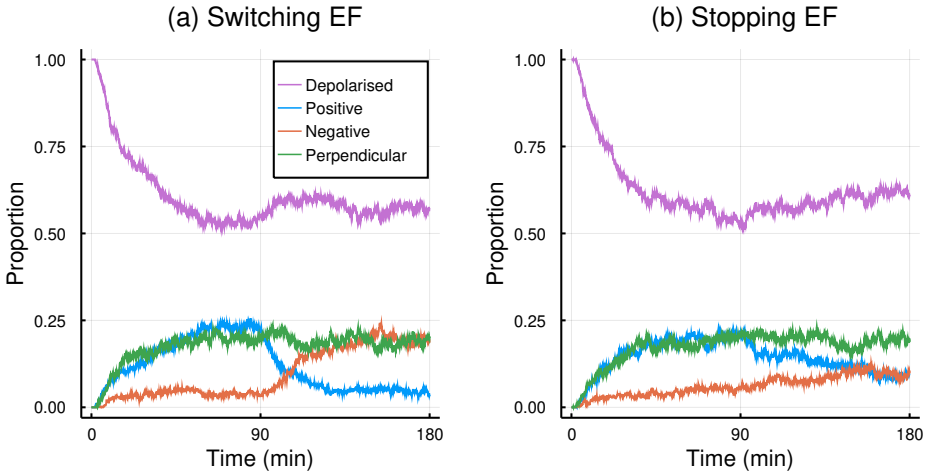


Figure S6: Fraction of 500 simulations with polarity $\mathbf{p}(t)$ in each of the four coarse-grained states depicted in Figure S4, for (a) the switching input, $\mathbf{u}_{\text{switch}}$, and (b) the stopping input, $\mathbf{u}_{\text{stop}}$. Simulations are generated from the electrotactic model in Eq. (3), using the sample mean of the posterior sample depicted in Figure S2.

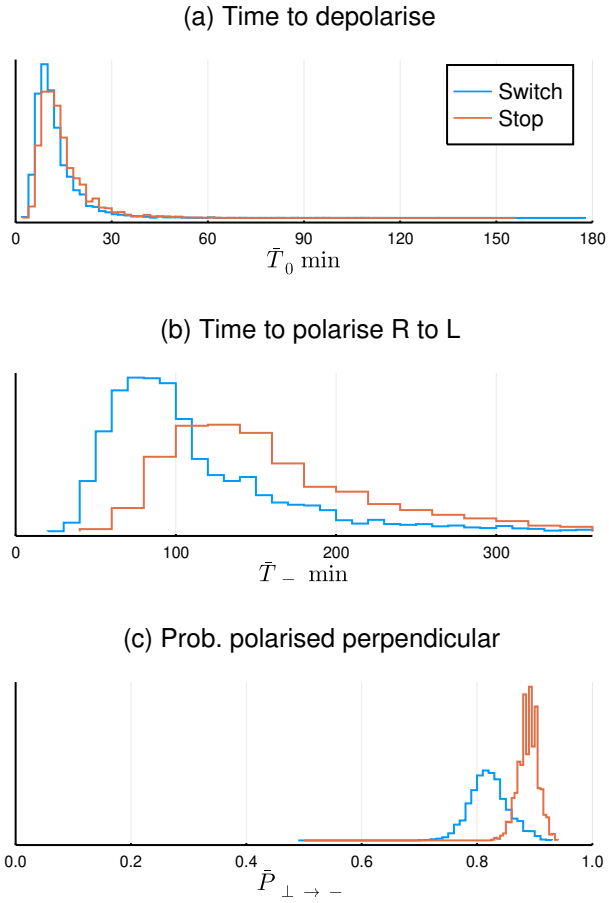


Figure S7: Posterior predictive samples for (a) expected time to depolarisation, $\bar{T}_0$ min, (b) expected time to polarisation in the negative $x$ direction, $\bar{T}_-$ min, and (c) probability of perpendicular polarisation before negative polarisation, $\bar{P}_{\perp \to -}$, under the posterior $\pi(\theta \mid \mathbf{x}_{\text{NoEF}}, \mathbf{x}_{\text{EF}})$ depicted in Figure S2. Simulations assume initial polarity $\mathbf{p}_0 = \mathbf{i}$ in the positive $x$ direction, with constant EFs $\mathbf{u} = -\mathbf{i}$ (Switch) and $\mathbf{u} = 0$ (Stop).