

Systematic Analysis of Simulated Atrial Electrograms to Distinguish Scar from Slow Conduction

Robbe Daelman
Student number: 02106237

Promotor: Prof. Dr. Nele Vandersickel
Supervisor: Sebastiaan Lootens

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Science in Physics and Astronomy

Academic year: 2025–2026



**GHENT
UNIVERSITY**

Preamble

This thesis begins with a background chapter that provides the necessary context for readers who may not be familiar with the subject matter. Its purpose is to introduce the concepts required to understand the subsequent chapters. Readers with prior knowledge of the field may therefore choose to skip parts of this chapter and proceed directly to the introduction. In particular, section 1.5 gives a detailed explanation of electrogram signals in simulations, including the extraction methods, the frequently used features within the field, the modelling of fibrotic tissue and its effect on the signals, and the analysis of these signals.

The appendix contains additional tables and figures that support the main text but are not essential for following its line of reasoning. These materials provide supplementary information that may be useful for interpreting the results presented in the main chapters.

Finally, this master's dissertation is part of an examination. Any comments made by the assessment committee during the oral presentation of the dissertation are therefore not included in this text.

Preface

I present this thesis as the final step in completing my studies in Physics and Astronomy at Ghent University. Over the past year, I have learned a great deal, both about cardiac arrhythmias and about conducting scientific research. The completion of this thesis would not have been possible without the support of many people, to whom I would like to express my sincere gratitude.

First and foremost, I would like to thank my promoter, Prof. Dr. Nele Vandersickel, for her guidance throughout the year. Your insights were invaluable in bringing this thesis to completion, and your continuous encouragement gave me the motivation to keep refining my work. You helped me improve my research skills and critical thinking, and made me a better, more well-rounded scientist.

I am especially grateful to my supervisor, Sebastiaan Lootens, who consistently made time to support me throughout this journey. Your feedback on the code, figures, and writing undoubtedly helped elevate this work to a higher level. You were always willing to answer my many questions and provide clarification whenever needed. Your meticulous proofreading and engagement with my thesis greatly improved the quality of this work and made the process much more enjoyable.

I would also like to thank the entire Biophysics research group at Ghent University for attending the biweekly meetings, providing feedback on my progress and results, reviewing my code, and suggesting ideas to improve this work. On a professional level, every one of you was always willing to help with any issue I encountered. On a personal level, the countless lunches we shared and the activities we did together were a welcome break from the demanding process of writing this thesis.

I would further like to thank Dr. Mattias Duytschaever and Dr. Alexander Panfilov for taking the time to listen to my progress and for providing valuable feedback.

As my time studying Physics and Astronomy comes to an end, I would like to thank everyone who shared this journey with me: the friends who have been with me throughout my life, as well as the friends I made during my years at university, and my roommates. Thank you for always putting up with my physics talk.

Finally, I would like to thank my family for their love, support, and endless patience. Much to the dismay of some of you, I brought up thesis-related topics at the most inconvenient times, but you nonetheless put up with it. I could not have done this without your support.

Robbe Daelman
June 2026

AI Statement

Several AI tools were used during the writing of this thesis. All content, including the ideas, analysis, interpretation, and conclusions, is my own and I take full responsibility for the contents of this thesis. All AI-generated suggestions were critically assessed, verified where necessary, and edited before being incorporated.

Since the simulation package used in this thesis is not widely available and is highly specific to this work, AI models were not used to directly write the simulation code itself. AI tools were only used to support auxiliary tasks, as described below. AI tools were also not implemented to generate data, perform simulations, produce numerical results or draw scientific conclusions.

ChatGPT: ChatGPT, a large language model developed by OpenAI, was used for simple coding tasks, such as writing plotting functions and other small auxiliary code snippets. It was also used to improve the language, clarity, and readability of the thesis, and to suggest possible improvements to figures. It was furthermore used to restructure paragraphs and improve the academic tone of the text, and to help with LaTeX-related formatting.

Claude: Claude, a large language model developed by Anthropic, was used for simple coding tasks, such as writing plotting functions and other small auxiliary code snippets and for debugging. It was also used to improve the language, clarity, and readability of the thesis, and to suggest possible improvements to figures. It was furthermore used to restructure paragraphs and improve the academic tone of the text.

GitHub Copilot: GitHub Copilot, an AI-assisted code completion tool integrated in Visual Studio Code, was used to support more advanced programming tasks such as code completion, restructuring auxiliary code, creating helper classes, handling input/output routines, and debugging. It was not used to directly write the simulation code, but only to assist with surrounding and supporting code.

Elicit: Elicit, an AI tool developed for scientific literature search, was used to help identify relevant papers and obtain preliminary summaries of scientific literature. All papers and suggestions identified using Elicit were manually checked and critically evaluated before being used in the thesis, as would be done for scientific sources found by any other method.

Abstract

Introduction: Cardiac ablation planning requires accurate identification of non-conductive tissue. Current methods, such as phase mapping and voltage maps, often overestimate the extent of non-conductive regions and provide only a simplified representation of the underlying electrogram (EGM) signals. Moreover, current markers cannot reliably distinguish structural remodelling from functional slow conduction. This thesis investigates whether additional EGM-derived features can be used to distinguish differences in electrical conduction velocity through cardiac tissue.

Methods: Simulations were performed for five different tissue configurations of increasing complexity and reduced symmetry. All simulations were conducted in 2D using a single stimulus and included a central obstacle surrounded by a grid of EGM electrodes. From each EGM signal, 11 features were extracted based on quantities previously described in the literature or commonly used in ECG analysis. PCA was then applied to identify the linear combinations of features that account for the largest variance in the data. The major principal components were then used in supervised machine learning approaches to predict the conduction velocity and distance to affected tissue.

Results: The first principal component accounted for 65–70% of the variance across all simulation types and was dominated by features related to signal strength, including maximum amplitude and total signal area. The second through fourth principal components together explained approximately 25% of the variance and primarily contained features related to peak structure, timing, and signal asymmetry. The values of these principal components, particularly the first component, differed systematically between simulations with different conduction velocities. A logistic regression classified between zero and non-zero distance with 89.3% balanced accuracy, and a random forest regressor predicted the averaged conduction with an absolute mean error of 1.2%.

Conclusion: EGM-derived features beyond phase and maximum voltage carry discriminative information about cardiac tissue conduction properties. The low-dimensional structure identified by PCA generalises across tissue configurations of increasing complexity, and the resulting feature representations support both classification of electrode distance to affected tissue and prediction of local conduction velocity. These findings suggest that multidimensional EGM feature analysis provides a more appropriate framework for distinguishing scar tissue from functional slow conduction regions than isolated signal markers alone.

Contents

Preamble	i
Preface	ii
Abstract	iv
Contents	v
List of Figures	vii
List of Tables	ix
List of abbreviations	x
1 Background	1
1.1 The heart	1
1.1.1 Anatomy of the heart	1
1.1.2 Coronary circulation	2
1.1.3 Types of cardiac tissue	3
1.2 Cardiac conduction	4
1.2.1 Electrical pathways in the heart	4
1.2.2 The action potential	4
1.3 Cardiac arrhythmias	7
1.3.1 Types of cardiac arrhythmias	7
1.3.2 Mechanisms of arrhythmias	8
1.3.3 Detection of cardiac arrhythmias	8
1.3.4 Treatment of cardiac arrhythmias	10
1.4 Computational modelling of the heart	11
1.4.1 Mathematical formulation of cardiac propagation	11
1.4.2 Numerical methods	12
1.4.3 The Courtemanche-Ramirez-Nattel model	12
1.4.4 Initial conditions and pacing	13
1.4.5 The Finitewave simulation package	14
1.5 EGM signals in simulation	14
1.5.1 Extracting EGM signals from simulated cardiac tissue	14
1.5.2 Features extracted from EGM signals	15
1.5.3 Modelling of fibrotic tissue and the effect on uEGMs	19
1.5.4 Analysis of EGM signals	21
2 Introduction	24
3 Methods	26
3.1 Setup of the simulation in Finitewave	26

3.1.1	Integration scheme	26
3.1.2	The cardiac model	27
3.1.3	The cardiac cell mesh	27
3.1.4	Prepacing	27
3.1.5	Applying stimuli	28
3.1.6	Setting up the EGM electrodes	28
3.2	5 Datasets of EGM signals in increasingly complex S1 simulations	28
3.2.1	Dataset 1: Modified conduction simulations	29
3.2.2	Dataset 2: Conduction gradient simulations	29
3.2.3	Dataset 3: Modified stimulus height simulations	30
3.2.4	Dataset 4: Probabilistic boundary simulations	31
3.2.5	Dataset 5: Random infarcted tissue	31
3.2.6	Defining properties of the setup	31
3.3	Analysis of EGM signals	34
3.3.1	11 features extracted from the EGM signals	34
3.3.2	Principal component analysis	39
3.3.3	Supervised machine learning	40
4	Results	43
4.1	Qualitative analysis of EGM signals	43
4.1.1	Dataset 1: Modified conduction simulations	43
4.1.2	Dataset 2: Conduction gradient simulations	44
4.1.3	Dataset 3: Applying stimuli at an angle	44
4.1.4	Dataset 4: Adding randomness to the boundaries of affected tissue	45
4.1.5	Dataset 5: Generating random affected tissue	45
4.1.6	Effect of changes in the parameters on the recorded EGM signals	47
4.2	Quantitative analysis of EGM signals	53
4.2.1	EGM feature extraction	53
4.2.2	Principal component analysis (PCA)	55
4.2.3	Supervised machine learning	60
5	Discussion	62
5.1	Overview of current literature on EGM signal analysis	62
5.2	Main findings	62
5.2.1	Qualitative analysis of EGM signals	62
5.2.2	Quantitative analysis of EGM signals	64
5.3	Limitations	66
5.4	Future prospects	67
6	Conclusion	69
	Bibliography	70
A	Supplementary material	82
A.1	Supplementary material for methods	82
A.2	Supplementary materials for results	84
A.2.1	Correlation between different features	84
A.2.2	PCA loadings and explained variance	87
A.2.3	Correlations between the principal components	90

List of Figures

1.1	Anatomy of the heart	2
1.2	Cardiac venous system	2
1.3	Composition of the cardiac wall	3
1.4	Electrical conduction pathways in the heart	5
1.5	Structural organization of a gap junction	6
1.6	Action potential in myocytes and pacemakers	7
1.7	Types of rotation in AT	9
1.8	ECG signal	9
1.9	Electrical propagation in cardiac tissue	11
3.1	Overview setups	32
3.2	Averaged conduction definition	33
3.3	Overview feature definitions	35
3.4	Example of a Poincaré plot	38
3.5	Example of PCA analysis	40
4.1	Overview simulation setups with indicated electrode	46
4.2	EGM signals for every simulation type	46
4.3	Overview simulation frames	47
4.4	EGM signals for variable width	49
4.5	EGM signals for variable conduction	50
4.6	EGM signals for changing simulation parameters	52
4.7	Overview of results of all features across all simulation types	54
4.8	Overview of the results of the PCA analysis across all simulation types	57
4.9	PC1 vs PC2 for all simulation types	58
4.10	Correlations between the principal components for the full dataset	59
4.11	Confusion matrix logistic regression	61
A.1	Effect of averaged conduction on simulation setups	83
A.2	Feature correlations substrate simulation	84
A.3	Feature correlations conduction gradient simulation	84
A.4	Feature correlations variable stimulus simulation	85
A.5	Feature correlations probabilistic boundary simulation	85
A.6	Feature correlations diffuse infarct simulation	86
A.7	PCA loadings for the substrate simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.	87
A.8	PCA loadings for the gradient simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.	87
A.9	PCA loadings for the variable stimulus simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.	88

A.10 PCA loadings for the probabilistic boundary simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.	88
A.11 PCA loadings for the diffuse infarct simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.	89
A.12 PCA loadings for all simulations in one dataset. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.	89
A.13 Correlations between the principal components for the substrate simulations . . .	90
A.14 Correlations between the principal components for the conduction gradient simulations	91
A.15 Correlations between the principal components for the variable stimulus simulations	92
A.16 Correlations between the principal components for the probabilistic boundary simulations	93
A.17 Correlations between the principal components for the diffuse infarct simulations	94

List of Tables

3.1	Simulation parameters	32
4.1	Random forest regression performance	60
4.2	Logistic regression performance	61
A.1	Model constants for the implementation of the Courtemanche model and tissue-level parameters in Finitewave.	82

List of abbreviations

AF atrial fibrillation

AP action potential

APD action potential duration

AT atrial tachycardia

AV atrioventricular

biEGM bipolar electrogram

CV conduction velocity

ECG electrocardiogram

EGM electrogram

LAT local activation time

PCA principal component analysis

SA sinoatrial

uEGM unipolar electrogram

VF ventricular fibrillation

VT ventricular tachycardia

Chapter 1

Background

1.1 The heart

The cardiovascular system is responsible for circulating blood throughout the body. It delivers oxygenated blood and nutrients to peripheral organs and, in turn, removes waste products such as carbon dioxide. The heart functions as the pump of this system and receives deoxygenated blood from the whole body, which it then pumps to the lungs for re-oxygenation. This oxygenated blood then flows back to the heart, from where it is pumped to the rest of the body. The heart is a highly intricate, multiscale biological system that contains several types of cells, each with its own function. [1]

1.1.1 Anatomy of the heart

As shown by the red and blue colours in figure 1.1, the heart can be divided into a right part, which receives the deoxygenated blood from the rest of the body and pumps it to the lungs, and the left part, which receives oxygenated blood from the lungs and pumps this to the rest of the body. Furthermore, the heart can be split up into the upper part, which consists of the thin-walled atria where the blood enters the heart, and the lower part, which consists of the thick-walled ventricles which pump blood to either the lungs or all other parts of the body [1].

The heart hence consists of four main components: the left atrium (LA), the right atrium (RA), the left ventricle (LV) and the right ventricle (RV). In order to maintain separation between oxygenated and deoxygenated blood, the atria are separated by the interatrial septum and the ventricles by the interventricular septum. Furthermore, it is imperative that blood can only flow from the atria towards the ventricles, and not the other way around. This unidirectional circulation is ensured by valves, which prevent backflow by well-timed opening and closing. [2]

The full blood circulation within the heart is shown in figure 1.1. Blood coming from the body first passes through either the superior or inferior vena cava before entering the right atrium. From there, it moves through the tricuspid valve, which ensures unidirectional flow, into the right ventricle, which then pumps the blood through the pulmonary valve into the pulmonary artery towards the lungs. Blood returning from the lungs then flows through the pulmonary veins into the left atrium, which pumps it through the mitral valve, which again ensures a unidirectional flow, into the left ventricle. This then finally pumps the blood through the aortic valve into the aorta and out of the heart. The aorta then distributes the oxygen-rich blood throughout the rest of the body. [3].

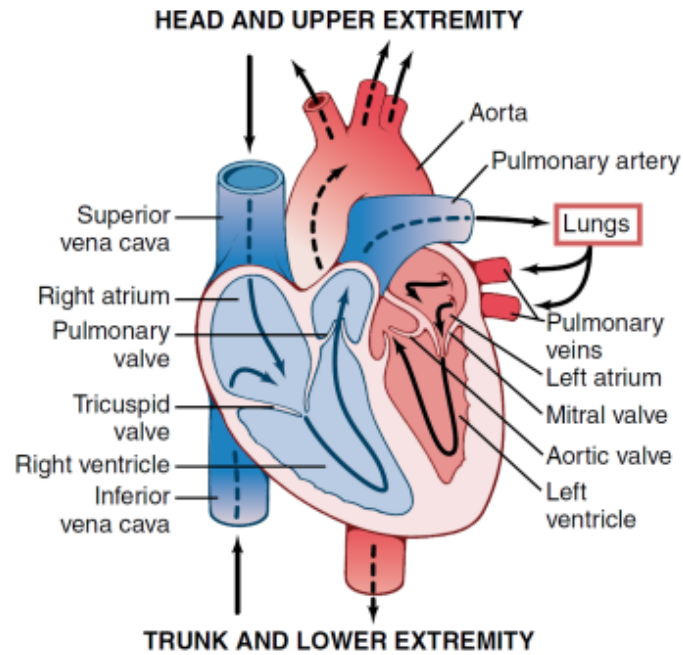


Figure 1.1: Anatomy of the heart. Blue areas indicate deoxygenated blood, while red parts indicate oxygenated blood. [3]

1.1.2 Coronary circulation

The heart itself also needs a blood supply, which is implemented through the coronary system and is shown in figure 1.2. The two main compartments of this system are the arterial compartment, which delivers oxygenated blood and the venous compartment, which removes waste products. The coronary venous system, which is subdivided into the greater and lesser cardiac venous systems, complements the arterial supply by draining deoxygenated blood from the heart [4].

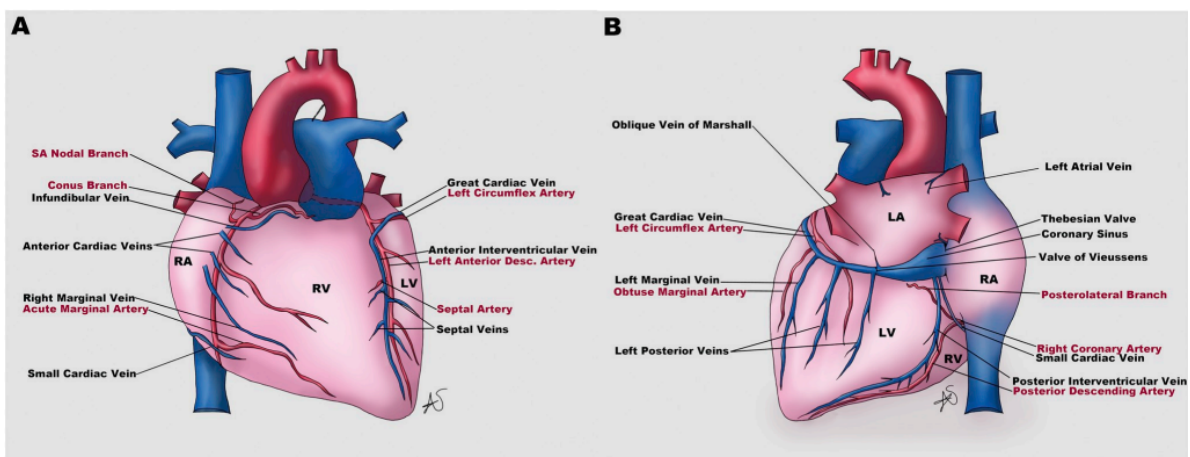


Figure 1.2: The greater cardiac venous system. Illustrations showing coronary venous anatomy of the greater cardiac venous system on coronal (A) and posterolateral (B) views of the heart. [4]

1.1.3 Types of cardiac tissue

The muscle tissue in the cardiac wall ensures that the heart can contract rhythmically and forcefully, allowing for continuous blood circulation throughout the body. It is a multilayered structure designed to withstand continuous mechanical stress, maintain electrical conductivity and support the heart's metabolic demands [5]. The heart wall contains three main layers, which are shown in figure 1.3.

The outermost layer is called the epicardium, which is also referred to as the visceral layer of the serous pericardium. It contains coronary blood vessels, lymphatic vessels, and autonomic nerves that supply the heart. Its primary function is to minimise friction between the beating heart and adjacent thoracic structures and to provide a protective covering to the heart. The epicardium is primarily made up of a single layer of mesothelial cells, which form a smooth protective outer lining, and connective tissue [6].

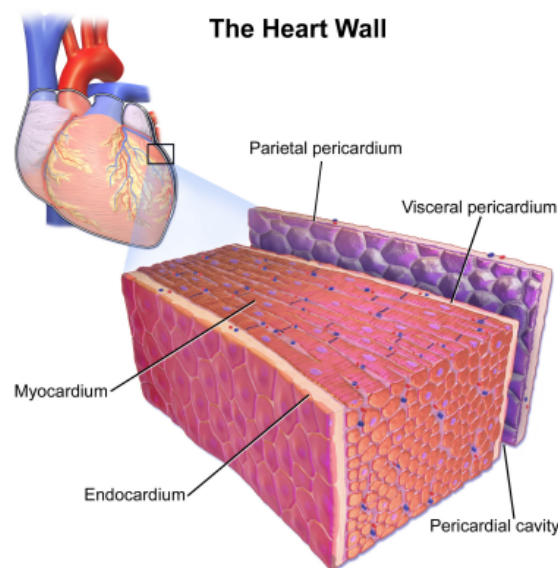


Figure 1.3: Composition of the cardiac wall [7].

The middle layer is referred to as the myocardium, which makes up the largest portion of the heart wall. It is thick compared to the other layers and mainly composed of cardiomyocytes, the contractile heart muscle cells, but also contains other cells such as fibroblasts and neural cells [8]. The myocardium is the muscular engine of the heart, responsible for generating the force needed to pump blood throughout the body. Since different chambers have different pressure demands, the myocardium differs in thickness, being relatively thin in the atria and significantly thicker in the ventricles [9].

Finally, the endocardium makes up the innermost layer of the cardiac wall, lining the heart chamber and covering the heart valves. Its main components are a thin layer of endothelial cells, which form a smooth inner lining in contact with the blood, and underlying subendothelial connective tissue. This smooth endothelial lining ensures minimal resistance to blood flow within the heart [9]. The endocardium also includes parts of the cardiac conduction system, which coordinate the timing and propagation of electrical impulses that control the contraction of the cardiac chambers [10].

The pericardium surrounds the heart and provides structural support and reduces friction during

cardiac motion. It is double-walled and consists of two main components: the serous pericardium and the fibrous pericardium. The serous pericardium is a thinner, double-layered membrane located within the fibrous pericardium. It is divided into the parietal layer, which lines the inner surface of the fibrous pericardium, and the visceral layer, which is continuous with the epicardium [11]. Fibrous pericardium is the tough, dense outer layer that anchors the heart within the thoracic cavity. It prevents overexpansion when blood volume increases, and it stabilises the heart's position by limiting excessive movement [12].

Between these two serous layers lies the pericardial cavity, which contains a small amount of lubricating serous fluid. This fluid reduces friction generated by cardiac movements, enabling smooth and efficient heartbeats [12].

1.2 Cardiac conduction

1.2.1 Electrical pathways in the heart

The electrical activity in the heart is responsible for the rhythmic contraction of the atria and the ventricles. Both working cells and pacemaker cells play a crucial role in proper cardiac conduction [13]. Working cells make up the majority of the myocardium and have a stable resting membrane potential, ranging from -90 mV to -80 mV. They require external stimulation to generate electrical signals. On the contrary, pacemaker cells have an unstable resting potential and exhibit automaticity, which is the intrinsic ability to generate a signal without the need for external stimuli [14]. The location of the pacemaker cells in the heart is shown in Figure 1.4. The sinoatrial node (SA) is the heart's primary pacemaker and hence contains cardiac pacemaker cells. It is located in the right atrium. The atrioventricular node is a secondary pacemaker, also located in the right atrium near the atrioventricular junction and it also contains pacemaker cells. The final locations in the heart containing pacemaker cells are the bundle of His and the Purkinje fibres in the ventricular septum and ventricular walls, which can act as backup pacemakers in case the others fail [3].

A normal heartbeat begins at the SA node, which sets the intrinsic heart rhythm due to its fast, spontaneous depolarisation rate. In healthy individuals, this rhythm lies between 60 and 100 beats per minute (bpm) [9]. After leaving the SA, the electrical wave travels over the atria, causing them to contract and eject blood into the ventricles. After this, the excitation wave converges at the AV node, which forms the electrical bridge between the atria and the ventricles. In a healthy heart, supraventricular pulses can only pass through the AV node, as the atria and ventricles are electrically isolated by the heart's fibrous skeleton [6]. The conduction velocity through the AV node is approximately 0.05 m/s, much slower than in atrial or ventricular muscle, where velocities can reach 1 m/s. This ensures that the atria have fully contracted and filled the ventricles before any ventricular contraction begins [13].

Subsequently, the excitation wave propagates down the bundle of His and separates into the right and left bundle branches, which further divide into specialised fibres. An example of these specialised fibres is the Purkinje fibre network, which rapidly transmits the excitation through the ventricular myocardium. This coordinated spread of excitation ensures simultaneous depolarisation and contraction of both ventricles, facilitating the efficient ejection of blood into the pulmonary artery and aorta.

1.2.2 The action potential

Although electrical activity in the heart spreads as waves across the tissue, it arises from the activity of individual cells. The cellular electrical signal is known as the action potential (AP), a

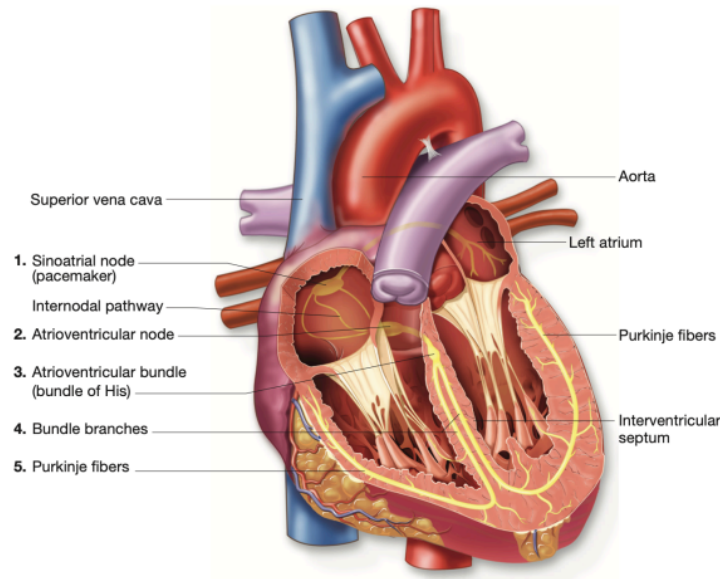


Figure 1.4: Electrical conduction pathways in the heart [15]

rapid transient change in the electrical potential across the cardiomyocyte membrane triggered by an external stimulus [16]. The action potential is mediated by ion currents through the cell membrane and depends on three key factors: the concentration gradients of ions across the cell membrane, the selective permeability of the membrane to specific ions, and the active transport of ions across the membrane by electrogenic pumps [17].

Intracellular and extracellular fluids contain different concentrations of ions, whose movement is governed by both chemical and electrical forces. Chemically, ions diffuse from areas of high concentration to low concentration and electrically, they respond to the electric field created by the membrane potential. When both of these forces are perfectly balanced, the myocardial cell is in electrochemical equilibrium. Each ion has a distinct equilibrium potential, determined by its concentration gradient and charge. The overall resting membrane potential results from the combined influence of all permeable ions across the cell membrane. The regulation of the ion movements is performed by ion channels located on the membrane of the cardiac myocytes. These protein complexes span the lipid bilayer and selectively permit the passage of specific ions [17].

Action potential in cardiomyocytes

Cardiac myocytes cannot generate APs spontaneously and hence require external stimulation, which is provided by electrical impulses originating from the SA node. These impulses are transmitted through gap junctions, which are clusters of intercellular channels that connect neighbouring cells. Each channel consists of a pair of connexons, each consisting of six subunits embedded in the cell membrane, as illustrated in Figure 1.5. When connexons of adjacent cells are aligned, they directly connect the intracellular cytoplasm and enable rapid ion flow between cells. Because cardiac myocytes are elongated and electrically coupled in a direction-dependent manner, conduction is anisotropic: excitation propagates more rapidly along the long axis of the myocytes than transverse to it. In some cardiac structures, longitudinal conduction can be approximately an order of magnitude faster than transverse conduction [18]. As a result, excitation spreads rapidly through the myocardium, while its direction and velocity are strongly influenced by cellular alignment and gap-junctional coupling [19].

As an electrical impulse reaches a cardiomyocyte, it triggers a defined sequence of ion channel activations that initiate the cardiac AP [13]. A typical AP of a myocyte is shown in Figure 1.6. Upon stimulation, the cardiomyocyte membrane rapidly depolarises due to the opening of Na^+ channels, which generate a large inward sodium current. After this initial upstroke, the membrane potential is maintained temporarily at depolarised values by a balance between inward Ca^{2+} currents and outward K^+ currents, before repolarisation restores the resting membrane potential. This sequence of ionic currents produces the characteristic cardiac action potential and allows electrical excitation to propagate through the tissue [9].

During and shortly after the action potential, the cardiomyocyte enters a refractory period, during which its responsiveness to new stimuli is temporarily suppressed. For most of this interval, the cell is in an absolute refractory state, where further excitation is impossible. As repolarisation nears completion, this is followed by a relative refractory period, during which a stronger-than-normal stimulus may trigger a new action potential, although with reduced amplitude and slower kinetics. The refractory period ensures that each action potential is followed by sufficient recovery, helping the heart maintain a regular and efficient rhythm [20].

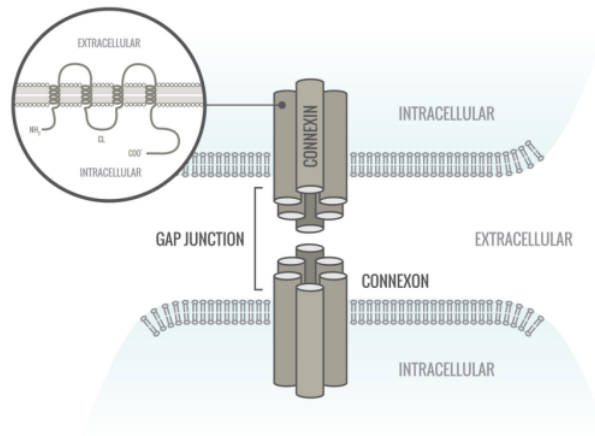


Figure 1.5: Structural organization of a gap junction [21].

Action potential in pacemaker cells

Unlike working cardiomyocytes, pacemaker cells in the SA and AV nodes do not maintain a stable resting membrane potential. Instead, they reach a maximum diastolic potential, followed by a slow spontaneous depolarisation that gradually brings the membrane to threshold. This automatic depolarisation is mainly driven by inward pacemaker currents, including the hyperpolarisation-activated (I_f), background (Na^+) currents, transient (Ca^{2+}) currents, and the sodium-calcium exchanger (I_{NCX}). The SA node has the least negative maximum diastolic potential and the steepest spontaneous depolarisation slope, explaining why it acts as the primary pacemaker of the heart. Once the threshold is reached, the pacemaker action potential upstroke is driven mainly by ($I_{Ca,L}$), in contrast to working cardiomyocytes, where rapid depolarisation is primarily mediated by (I_{Na}). Pacemaker action potentials also lack the pronounced early repolarisation and plateau observed in working cardiomyocytes; instead, gradual repolarisation restores the membrane potential to the maximum diastolic potential, allowing the cycle of spontaneous depolarisation and repolarisation to repeat [9, 13]. The pacemaker AP is shown in figure 1.6, where it can be compared with the myocyte AP.

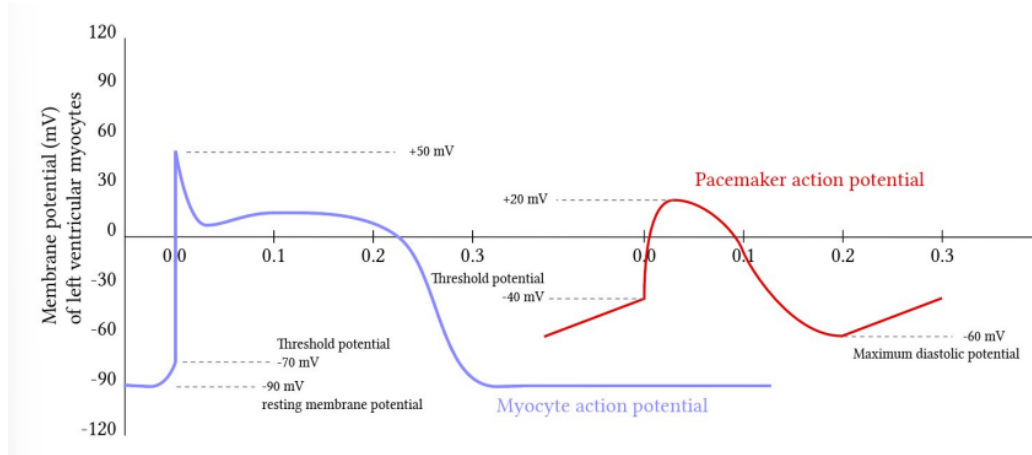


Figure 1.6: Comparison of action APs in pacemaker cells and cardiac myocytes [22].

1.3 Cardiac arrhythmias

1.3.1 Types of cardiac arrhythmias

A cardiac arrhythmia is defined as an aberrant heart rhythm, but this covers a broad range of conditions. These can be classified in different ways, the most basic being by their location. Arrhythmias can occur either in the atria or the ventricles and are hence referred to as atrial arrhythmias and ventricular arrhythmias, respectively [9, 14].

Another classification looks at the regularity of the arrhythmia, where a heartbeat is referred to as regular when the duration between beats stays constant over time. If heartbeats are not regularly spaced, the arrhythmia is classified as irregular. This can be caused by conditions such as fibrillation, which is characterised by completely disorganised electrical activity, causing asynchronous contractions of the tissue. Notable irregular arrhythmias are atrial fibrillation (AF) and ventricular fibrillation (VF). A special case occurs when a regularly irregular rhythm happens when extra (ectopic) beats are added into a normal rhythm. The spacing between beats is still regular, but extra beats happen between each main beat. In other words, an irregular pattern repeats regularly [9].

A final classification looks at the speed of the arrhythmia, measured in beats per minute (bpm). A typical resting heart rate lies between 60 and 100 bpm. Arrhythmias causing an abnormally slow heartbeat (less than 60 bpm) are referred to as bradycardias, while arrhythmias resulting in a resting heartbeat faster than 100 bpm are called tachycardias. The main types of tachycardia are atrial tachycardia (AT) and ventricular tachycardia (VT) [23].

Atrial arrhythmias are generally speaking less severe than ventricular arrhythmias. AT, for example, is an arrhythmia in the atria with a faster than normal heartbeat and is not life-threatening, however it causes discomfort for the patient and can develop into more severe arrhythmias [24]. More severe ventricular arrhythmias are VT and VF. VT results in a very high heart rate, which prevents normal circulation of blood through the body. If left untreated, VT can deteriorate into VF, where the heart's electrical activity becomes completely disordered. The heart will stop delivering blood, resulting in a cardiac arrest. This condition works very quickly and, if left untreated, will result in death within minutes [9].

1.3.2 Mechanisms of arrhythmias

Wave dynamics

As explained in section 1.2.2, cardiac tissue has a refractory period after every excitation. That is, there is a period of time during which the same cardiac tissue cannot be re-excited by nearby APs. Due to this refractory period, waves of electrical signals can collide with each other, resulting in the termination of the signal. Furthermore, a single wave can enter tissue that is in this refractory state and can hence die out [9].

Focal sources

A focal source refers to a localised region from which spontaneous electrical activation originates and spreads through the surrounding tissue, producing approximately concentric activation patterns. A single premature activation from such a region is often referred to as an ectopic beat, whereas repeated focal activation can contribute to the initiation or maintenance of arrhythmias. When a focal source activates tissue at a frequency higher than that of the SA node, it can disrupt normal sinus rhythm and impose an abnormal pacing pattern on the surrounding myocardium [25, 26].

Re-entry

Re-entry is a mechanism in which an electrical excitation wave persists by repeatedly circulating through cardiac tissue and reactivating regions that have recovered excitability. A re-entry can occur when an action potential is shorter than the duration of travel around the centre, hereby reactivating tissue that is past its refractory period [27]. Re-entry can either occur around anatomical obstacles like valves, veins, arteries, and scar tissue or it can occur when an excitable core re-excites the site of origin. The latter is referred to as functional re-entry or a rotor [14]. A rotor is said to be meandering when its core moves over the cardiac tissue. Re-entry circuits can rotate at a higher frequency than the normal sinus rhythm, allowing them to overdrive the physiological excitation sequence and impose an abnormal activation pattern on the surrounding tissue [27].

Anatomical re-entry can be further split up into two types, true- and suppressed re-entry, both shown in figure 1.7. A true re-entry makes a complete rotation around the anatomical hole, while a suppressed loop collides with the wavefront of a true loop and will therefore make an incomplete rotation. Until recently, these suppressed loops were often considered irrelevant. However, due to the recent discovery of the topology-derived index theorem, it was revealed that suppressed loops must also be taken into account [28].

1.3.3 Detection of cardiac arrhythmias

Electrocardiogram (ECG)

An electrocardiogram (ECG) is a widely used, non-invasive technique for assessing the electrical activity of the heart. It is recorded by placing electrodes at specific positions on the patient's body, where they detect voltage differences generated by the propagation of electrical activation through the myocardium. The resulting ECG waveform reflects the coordinated depolarisation and repolarisation of the atria and ventricles during the cardiac cycle, as illustrated in figure 1.8. A typical ECG consists of a P wave, QRS complex, and T wave. The P wave corresponds

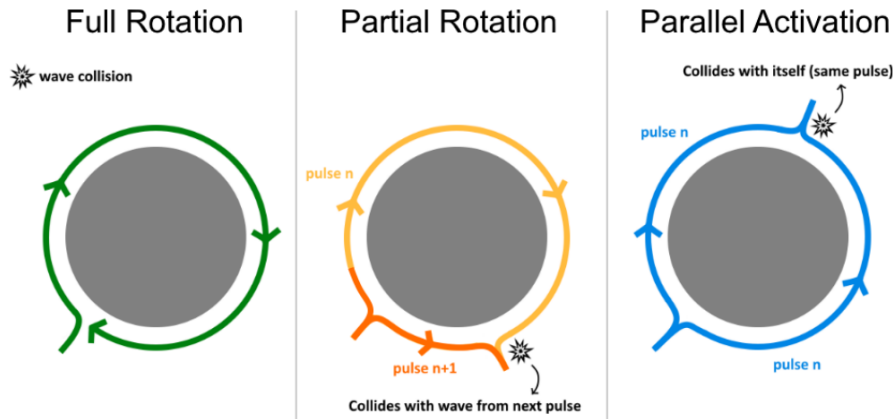


Figure 1.7: Types of rotation in AT. You can either have a full rotation or a partial rotation [29].

to atrial depolarisation, the QRS complex mainly reflects ventricular depolarisation, and the T wave represents ventricular repolarisation [9].

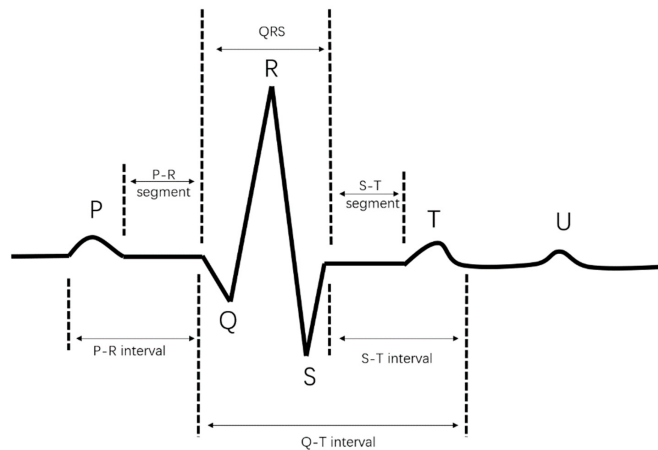


Figure 1.8: ECG signal with indicated signature peaks [30].

Intracardiac electrogram (EGM)

In addition to body-surface ECG recordings, the electrical activity of the heart can be measured locally from within the cardiac chambers using intracardiac electrograms. These signals are acquired during catheter-based cardiac mapping, in which a thin catheter carrying multiple electrodes is introduced through the blood vessels, commonly via the groin, and guided to the heart under fluoroscopic imaging. By recording electrograms at known spatial locations on the endocardial surface, cardiac mapping links local electrical activation patterns to the anatomical structure of the chamber. This makes it possible to identify abnormal activation sequences and arrhythmogenic regions that may not be resolved from body-surface recordings alone [31].

The origin of these intracardiac signals lies in the coordinated electrical activation of the myocardium. Electrical activity in cardiac tissue is generated by the opening and closing of ion channels in the cell membrane, which gives rise to the cardiac action potential, defined as the time-varying difference between the intra- and extracellular potential of a cell, as described in section 1.2.2.

Cardiac myocytes are electrically coupled through gap junctions, allowing excitation to propagate from cell to cell, preferentially along the local alignment direction of the myocytes. This coordinated propagation produces spatial and temporal variations in transmembrane voltage throughout the myocardium [32, 1].

The spatial distribution of the transmembrane voltage of the cells creates the extracellular field. As an advancing depolarisation wave will alter the spatial distribution of transmembrane voltage of the cells, it will also alter the extracellular potential, which can be measured using a unipolar EGM (uEGM). This principle is illustrated in figure 1.9, which also shows the unipolar electrogram morphology, which is symmetric and consists of a positive phase (R-peak) and a negative phase (S-peak). The R-peak indicates an approaching wavefront, while the downslope indicates the moment that the wave is underneath the electrode. Finally, the S-peak indicates that the wavefront is moving away from the electrode [32].

In clinical practice, uEGMs are affected by several sources of noise and interference, including hardware noise, respiratory artefacts, cardiac motion, and far-field signals originating from distant parts of the heart [33]. To mitigate these issues, bipolar EGMs, in which the signals of two closely spaced electrodes are subtracted [34], are frequently used. This subtraction suppresses common-mode noise that affects both electrodes [35], such as far-field cardiac activity and respiratory artefacts [32]. However, bipolar EGMs remain sensitive to several measurement uncertainties, including bipolar vector orientation, inter-electrode spacing, electrode size, and tissue contact [36, 37, 38, 39]. Nevertheless, unipolar EGMs retain information about activation direction and far-field propagation that may be lost in bipolar recordings [40, 41, 42]. Overall, bipolar EGMs are generally preferred in clinical practice because they provide improved suppression of far-field activity and common-mode noise [33].

Besides propagation-related effects, catheter and electrode design also influence recorded EGM morphology and the interpretation of EGM-derived features. Electrode size, shape, inter-electrode spacing, and tissue contact affect both unipolar and bipolar recordings and can substantially alter measured amplitudes, durations, and waveform morphology [34, 43, 33]. Larger electrodes spatially average electrical activity over a wider tissue region, generally increasing signal amplitude and duration while reducing spatial specificity [44, 43]. Similarly, bipolar recordings remain strongly dependent on electrode orientation and inter-electrode spacing, introducing additional uncertainty in substrate characterisation [38, 45]. Consequently, differences in catheter configuration and recording setup complicate direct comparison of EGM-derived features across studies and clinical mapping systems.

1.3.4 Treatment of cardiac arrhythmias

Cardiac arrhythmias can be treated using pharmacological or interventional approaches. Pharmacological treatment relies on anti-arrhythmic drugs, which modify cardiac excitability or conduction and can help restore normal rhythm. However, these drugs may also have pro-arrhythmic effects and can therefore introduce new arrhythmias [46]. A second treatment option is catheter ablation, in which a thin, flexible tube called a catheter is inserted into the heart. Electrodes at the catheter tip are used to locally heat the cardiac tissue, creating scar tissue that no longer conducts electrical signals. This can be an effective treatment for arrhythmias driven by re-entry circuits or focal sources, because the ablated region can interrupt the abnormal activation pattern [47, 48]. Therefore, accurately identifying the arrhythmogenic region is crucial for effective ablation therapy [28].

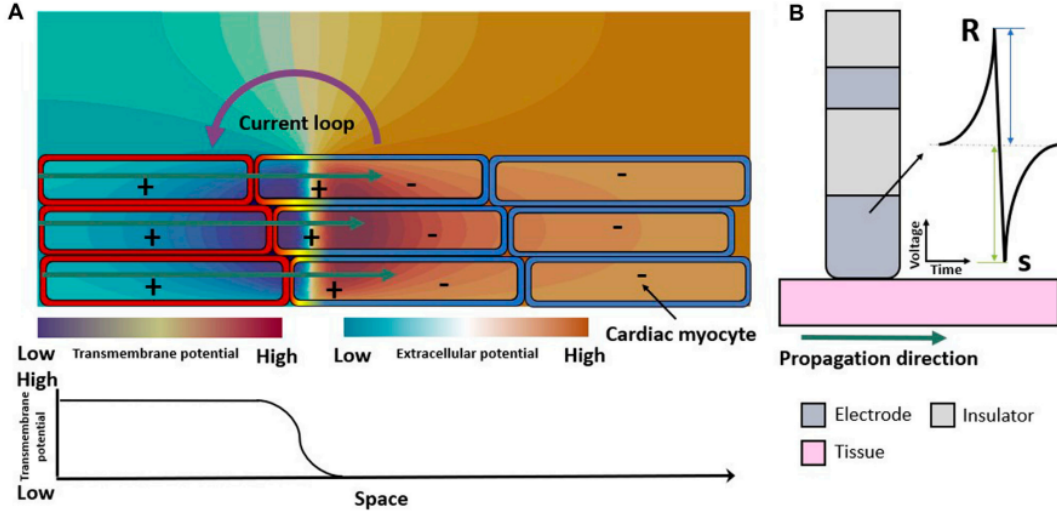


Figure 1.9: Electrical propagation in healthy cardiac tissue. **(A)** Extracellular field caused by the depolarisation of cardiomyocytes when an excitation propagates from left to right (green arrows). Spatial transmembrane voltage distribution is colour-coded on the membranes and shown in the bottom row. The leftmost cells are already depolarised and in the action potential plateau, while those on the right are still at resting membrane voltage. **(B)** Symmetric unipolar electrogram measured at the surface of the cardiac tissue (pink). The initial positive wave (R-peak) is caused by the wavefront approaching the electrode (dark grey), the polarity changes when the wavefront passes underneath the electrode, and the S-peak is caused by the wavefront travelling away from the measuring electrode. [32]

1.4 Computational modelling of the heart

The most accurate and complete model describing the electrical depolarisation of the myocardium and its cells consists of three parts: The Hodgkin-Huxley model of an excitable cell [49], the tissue homogenization approach [50], and the set of bidomain equations [51][32]. However, other models of excitable cardiac cells exist, some of which model all of the ionic currents explicitly and are hence considered more realistic [52]. On top of that, the bidomain equations are computationally expensive [53, 54], and other formulations attaining similar results exist [53]. The sections below will discuss these different approaches, with an emphasis on the methods used further on in the thesis.

1.4.1 Mathematical formulation of cardiac propagation

In the bidomain model, the cardiac tissue is described as a homogeneous medium that is composed of two domains: the intracellular and the extracellular domains. Both of these computational domains coexist, and they occupy the same geometrical space in the bidomain approach.

$$\nabla \cdot (\sigma_i \nabla \phi_i) = \beta I_m \quad (1.1)$$

$$\nabla \cdot (\sigma_e \nabla \phi_e) = -\beta I_m - I_{\text{extra}} \quad (1.2)$$

$$I_m = C_m \frac{\partial V_m}{\partial t} + I_{ion}(V_m, v) - I_{intra} \quad (1.3)$$

$$V_m = \phi_i - \phi_e \quad (1.4)$$

where ϕ represents the electrical potential, the indices i and e refer to the intracellular and extracellular spaces, respectively. σ is the conductivity tensor and β is the surface-to-volume ratio of the cardiac myocytes. I_{ion} is the total transmembrane ionic current density defined by the cellular model, which depends on the transmembrane voltage V_m and a vector v of further state variables. The external stimuli are described by a transmembrane current density I_{intra} and an extracellular current density I_{extra} . A bath surrounding the tissue is treated as an extension of the extracellular space [32].

The monodomain model assumes an equal anisotropy for the extracellular and intracellular conductivity, giving rise to only one parabolic differential equation [55]:

$$\nabla \cdot (\sigma_m \nabla V_m) = \beta I_m + \beta I_{tr} \quad (1.5)$$

with V_m defined in the same way as in equation 1.4 and β is again the surface-to-volume ratio of the cardiac myocytes. I_m is defined as in equation 1.3 and represents the transmembrane ionic current density, while I_{tr} is the transmitted stimulus current which is applied externally. σ_m is the bidomain equivalent monodomain conductivity and is defined as:

$$\sigma_m = \left(\frac{\sigma_i \sigma_e}{\sigma_i + \sigma_e} \right)^{-1} \quad (1.6)$$

Although the monodomain model is a stark simplification of the bidomain model, it has been shown that it is sufficiently accurate to study and understand the electrical propagation in the cardiac tissue, both under electrically remodelled tissue and under physiological conditions. In particular, electrograms computed with monodomain and bidomain models were visually indistinguishable [53].

1.4.2 Numerical methods

Numerical methods are required to simulate cardiac electrophysiology because the governing equations for action potential propagation generally cannot be solved analytically in realistic tissue geometries. Instead, the continuous cardiac domain is discretised into a finite number of spatial elements or grid points, after which the evolution of the transmembrane potential is computed over time. Different numerical approaches can be used for this purpose. Finite difference methods approximate spatial derivatives on a structured grid and are therefore computationally efficient for simple or regular geometries [53, 56]. In contrast, finite element methods divide the domain into elements that can more easily conform to complex anatomical shapes, making them particularly suitable for realistic cardiac chamber geometries [57, 58]. Both approaches are widely used to discretise cardiac tissue and simulate the propagation of electrophysiological signals.

1.4.3 The Courtemanche-Ramirez-Nattel model

The Courtemanche-Ramirez-Nattel model is a detailed ionic model of the human atrial action potential [59]. In this model, the time evolution of the transmembrane voltage is determined by

the balance between membrane capacitance, externally applied stimulation, and a set of ionic currents that represent the main ion channels, pumps, and exchangers present in atrial cardiomyocytes. These include the fast sodium current, L-type calcium current, several potassium currents, background currents, and calcium-handling mechanisms. By explicitly representing these currents and their associated gating variables, the model can reproduce the characteristic morphology and rate-dependent behaviour of human atrial action potentials.

Because the model describes the electrophysiology of a single atrial cell, it provides the local ionic dynamics required for tissue-scale simulations but does not by itself describe propagation between neighbouring cells. To simulate excitation spreading through cardiac tissue, the cellular model must therefore be coupled to a spatial propagation model, such as the monodomain or bidomain equation. In this coupling, the ionic model determines the local transmembrane current, while tissue-level parameters such as the diffusion coefficient and background current control how electrical activation spreads through the simulated domain. These propagation-related parameters are not part of the original single-cell Courtemanche-Ramirez-Nattel formulation and must therefore be chosen separately, typically based on values reported in the literature [52, 60].

The detailed ionic formulation of the Courtemanche-Ramirez-Nattel model makes it suitable for simulations in which realistic atrial action potential morphology and electrogram generation are important. However, its large number of state variables also means that the simulated state depends on the preceding activation history of the cell [61]. For this reason, the choice of initial conditions and the use of prepacing are important considerations before using the model in tissue-level simulations.

1.4.4 Initial conditions and prepacing

Ionic cardiac cell models contain a large number of state variables in addition to the transmembrane voltage, such as gating variables, ionic concentrations, and recovery variables. These variables introduce short-term memory into the model, since their values at a given time depend on the preceding activation history of the cell [62, 63]. As a result, starting a simulation from arbitrary initial conditions can lead to non-physiological behaviour. For example, gating variables may not initially be in equilibrium with the membrane voltage, ionic concentrations may not correspond to a realistic cellular state, and the resulting action potential morphology and action potential duration may change substantially over the first few beats.

To reduce the influence of these arbitrary initial conditions, cardiac tissue simulations are commonly prepaced before the actual simulation of interest. Prepacing consists of repeatedly stimulating the tissue with a fixed pacing protocol until the relevant state variables have approached a stable beat-to-beat response. In practice, this means that quantities such as the action potential morphology, action potential duration, gating variables, and ionic concentrations no longer change appreciably between consecutive beats [64]. The final state after prepacing is then used as the initial condition for the subsequent simulations.

It should be noted that not all ionic models converge to a unique steady state in a strict sense. The final state may depend on the pacing cycle length, stimulation protocol, and number of prepacing beats. For the Courtemanche atrial cell model, for instance, it has been shown that the model does not necessarily reach a true equilibrium state and that its state can remain dependent on the preceding pacing history [61]. Prepacing should therefore be interpreted as a way to obtain a reproducible and physiologically more realistic initial state, rather than as a guarantee of complete equilibrium.

1.4.5 The Finitewave simulation package

All simulations were performed with the Finitewave simulation package developed by the Biophysics Group at Ghent University. Finitewave is a Python package for simulating cardiac electrophysiology using finite-difference methods. It provides tools for defining cardiac tissue geometries, assigning spatially varying model parameters, applying stimulation protocols, and computing the propagation of electrical waves through the tissue. In addition, the package includes functionality for storing and visualising simulation outputs, making it suitable for systematically generating and analysing large sets of electrophysiological simulations [65].

1.5 EGM signals in simulation

Clinical EGM signals recorded with catheters are often complex and difficult to interpret visually, as their morphology is influenced by several factors beyond the underlying tissue substrate. These include the incident angle of the activation wavefront, inter-electrode spacing, electrode size and shape, superposition of multiple wavefronts, and variability in tissue contact. In addition, signal preprocessing choices and mapping density can further affect the recorded EGMs and the features derived from them [33]. Moreover, several diagnostic measures used by clinical practitioners do not have precise definitions, which also contributes to the difficulty of interpreting EGM signals [66]. In this context, simulations are useful as they can provide insight into the factors that alter the morphology and amplitude of the electrograms [32].

1.5.1 Extracting EGM signals from simulated cardiac tissue

Simulating realistic EGM signals requires both an accurate model of cardiac excitation and a method for extracting extracellular potentials from that simulation. The choice of computational approach involves trade-offs between physical fidelity and cost, spanning the selection of propagation model, the method used to approximate extracellular potentials, and the definition of a reference electrode. The following subsection outlines these considerations and introduces the analytical framework used to compute unipolar EGM signals.

Monodomain-based approaches to EGM extraction

Extracting EGM signals directly from a bidomain simulation is computationally expensive, as it requires explicit calculation of the extracellular potential throughout the tissue. Monodomain-based approaches avoid this by simulating excitation propagation without solving for extracellular potentials, which are instead approximated in a subsequent step [32]. This separation reduces computational cost substantially, making monodomain-derived methods the practical choice for large-scale simulations [54, 67, 53].

Hybrid and post-processing methods

All hybrid approaches share the same basic strategy: simulate excitation propagation with the monodomain model, then compute extracellular potentials separately. The methods differ in how the second step is performed and how much bidomain physics they recover.

The simplest assumption is an infinite homogeneous volume conductor, in which the cardiac tissue is immersed in an unbounded, uniform extracellular medium [68]. This allows direct study of the relationship between depolarization wavefront propagation and unipolar electrogram

morphology [45, 69], but ignores the influence of heterogeneous surrounding structures such as the cardiac chambers, lungs, and liver [32]. The reaction-eikonal model takes a different approach to reducing cost: rather than simplifying the extracellular step, it reproduces physiological propagation on a coarser computational mesh [67].

More physically faithful methods recover extracellular potentials through post-processing. In the pseudo-bidomain approach, transmembrane currents from a monodomain simulation drive a separate reconstruction of extracellular potentials, avoiding the cost of solving the full bidomain system [54, 70]. The augmented monodomain approach surrounds the tissue with a conducting bath, inducing wavefront curvature effects and recovering key characteristics of bidomain propagation and conduction velocity at substantially lower cost [54]. A related strategy solves the elliptic bidomain equation only at the time points where extracellular potentials are needed, using a precomputed transmembrane voltage distribution. This preserves important bidomain characteristics while remaining far cheaper than a full bidomain simulation [70], and produces EGM signals consistent with clinical recordings [71].

Reference potential selection

In electro-anatomical mapping, the reference potential is typically recorded at a remote site or computed as an average across multiple electrodes, such as Wilson’s central terminal [68, 32]. In bidomain simulations of unipolar EGMs, a common alternative is to define the reference as the average extracellular potential over the surface furthest from the tissue [72, 73]. Although this approximates rather than replicates a truly remote reference such as a surface patch on the patient’s back, it substantially reduces non-physical baseline drift [70].

Reference quality and computational cost trade off directly: a more distant reference improves stability but expands the computational domain [54, 67]. Numerically, averaging the potential over a remote surface or volume is preferable to imposing a fixed Dirichlet boundary condition, as the latter introduces additional artefacts [51, 54].

Analytical formulation of the unipolar EGM

For the monodomain model, the EGM signal can be derived analytically by assuming that the electrode and all source points are immersed in an infinite, homogeneous volume conductor. Here, homogeneity refers to the medium, not the sources. Under this assumption, the unipolar electrogram exhibits dipole-like behaviour: the measured potential depends on both the strength and orientation of the propagating electrical source relative to the electrode. The unipolar EGM at position $\mathbf{x}$ is then the extracellular potential contributed by distributed cardiac source dipoles throughout the tissue volume, referenced to an electrode at infinity:

$$\phi_e = \frac{1}{4\pi\sigma} \iiint_V \frac{I_{\text{src}}}{|\mathbf{x} - \mathbf{x}_{\text{src}}|} dV \quad (1.7)$$

where ϕ_e denotes the extracellular potential, σ represents the conductivity of the volume conductor, I_{src} is the source current density and $||x - x_{\text{src}}||$ is the distance from the source point to the measuring point [32].

1.5.2 Features extracted from EGM signals

Quantitative analysis of EGM signals requires defining waveform features that can be systematically extracted and linked to the electrophysiological properties of the underlying tissue.

These features enable substrate characterisation and may assist in guiding catheter ablation [34, 74, 33]. The following subsection describes the most relevant EGM features — amplitude, asymmetry, entropy, fractionation, timing, area, and temporal derivatives — along with their electrophysiological interpretation, known limitations, and key findings from clinical and computational studies.

Overview and interdependence of EGM features

Alterations in conduction velocity, tissue anisotropy, fibrotic remodelling, and wavefront interaction can modify multiple aspects of EGM morphology simultaneously, including signal amplitude, duration, asymmetry, fractionation, and temporal complexity. Several of these characteristics, like reduced amplitude, prolonged duration, and increased fractionation, have been associated with conduction slowing and structural remodelling [42, 75, 66]. Because overlapping mechanisms drive multiple features at once, EGM interpretation relies on the combined analysis of complementary waveform descriptors rather than any single feature in isolation.

Peak-to-peak amplitude

The peak-to-peak amplitude, defined as the difference between the maximum and minimum voltage of the EGM signal [32], is widely used in clinical practice to distinguish healthy from pathological tissue through voltage thresholds applied to both unipolar [75] and bipolar [40] recordings. It also serves as an indicator of the severity and spatial extent of conduction inhomogeneities [76], as conduction anisotropy produces direction-dependent amplitude variations, particularly for propagation parallel to myocardial fibres [77, 78]. These variations are partly linked to changes in sodium current dynamics and the resulting propagation properties [79]. One limitation is that peak-to-peak amplitude increases with tissue thickness, since a larger volume of electrically active tissue contributes to the measured potential and the influence of depolarising cells on the epicardial side diminishes accordingly. Tissue thickness does not substantially affect EGM morphology or duration, however, acting primarily on signal amplitude [80].

Deflection amplitude

Unlike peak-to-peak amplitude, deflection amplitude refers to the maximum positive and negative voltage excursions measured individually relative to the baseline. Several recording-related factors influence deflection amplitude in clinical practice. Electrode size inversely affects recorded amplitude [81], while bipolar amplitude increases with inter-electrode spacing due to the greater temporal activation delay between electrodes [78]. Bipolar amplitudes are also strongly influenced by wavefront direction and electrode orientation [38, 45], whereas unipolar amplitudes are more sensitive to far-field activity [82, 41]. Additional sources of amplitude variation include contact pressure, catheter motion from cardiac contraction and respiration [83], electrode-tissue interface polarisation [83], and electromagnetic interference [84]. At the tissue level, bipolar amplitude decreases with reduced CV, and larger non-conductive regions produce both lower amplitudes and increased bipolar widths, with the amplitude reduction proportional to scar size [45]. Uniform changes in tissue conductivity do not substantially alter EGM morphology, and the amplitude reduction associated with slow conduction appears primarily driven by reduced intracellular conductivity σ_i rather than conduction velocity itself, while EGM width is more directly related to conduction velocity [80].

Reduced deflection amplitude is strongly associated with slow conduction, conduction inhomogeneity, and structural remodelling. Slow conduction zones are generally characterised by

low-amplitude EGMs accompanied by fractionation and prolonged duration, whereas higher-amplitude EGMs are more indicative of wavefront collisions [85]. Bipolar voltage predicts regions of slow conduction with an area under the curve of 0.733 at a 95% confidence interval [86], and low-amplitude regions independently associate with local conduction slowing [87] and increased conduction inhomogeneity [42]. Both unipolar and bipolar amplitude reductions have been independently linked to scar tissue [88] and correlate inversely with collagen deposition [89]. Amplitude interpretation remains dependent on recording configuration and local propagation characteristics; however, limiting its specificity as a standalone substrate marker.

Waveform asymmetry

Waveform asymmetry quantifies the imbalance between positive and negative deflections in unipolar electrograms, typically measured as the signed waveform area or the difference between peak deflections [90, 42]. In healthy tissue with planar wavefront propagation, surface electrograms are approximately symmetric both in vivo and in silico [32]. In homogeneous and isotropic tissue, highly asymmetric electrograms are rare, though considerable variability in amplitude and symmetry persists. Introducing anisotropy increases symmetry variability, particularly for propagation across myocardial fibres, with highly curved wavefronts producing negatively asymmetric electrograms and colliding wavefronts producing positively asymmetric ones [79].

Electrogram asymmetry relates directly to the sine of the angle of the equivalent dipole relative to the tissue surface, with epicardial-endocardial activation delay and transmural coupling identified as key determinants [90]. At the tissue level, increasing the density and length of collagenous septa reduces conduction velocity by up to 75%, substantially increasing electrogram asymmetry [91], and collagen deposition more broadly alters asymmetry in a graded fashion [89]. Waveform asymmetry therefore reflects both propagation geometry and structural remodelling, making it a useful descriptor of the electrophysiological substrate.

Entropy-based complexity measures

Highly irregular activation dynamics during atrial fibrillation produce electrograms with substantial temporal and amplitude variability, motivating entropy-based descriptors as measures of signal irregularity and waveform organisation. Shannon entropy quantifies the amplitude distribution and overall complexity of bipolar EGMs and may assist in AF rotor mapping [92]. Dynamic approximate entropy correlates positively with electrogram fractionation in rotor simulations and can identify highly fractionated EGMs and localise rotor tips during simulated AF [69]. Both measures thus target the increased signal complexity associated with disorganised conduction, though their clinical translation remains an active area of investigation.

Electrogram fractionation

Electrogram fractionation describes electrograms containing multiple deflections and prolonged activation components, arising from discontinuous propagation and localised conduction heterogeneity. It is commonly quantified by the number of peaks relative to a healthy reference signal and is considered one of the most important markers of infarcted or diseased tissue, despite the lack of a universally accepted definition or recording methodology [66]. Fractionation corresponds to increased electrogram complexity and is often accompanied by prolonged signal duration. Reduced conduction velocity promotes increasingly fractionated morphologies in computational models [91], and clinical slow-conduction zones are commonly characterised by fractionated, low-amplitude, and prolonged EGMs [85]. Fractionated and double potentials are

frequently observed in regions of slowed conduction and conduction inhomogeneity [87, 42], and fractionation has been associated with anatomical scar, irrespective of its intramural location [93, 88]. Increasing collagen deposition is additionally associated with increased fractionation [89].

Simulations and clinical studies consistently link electrogram fractionation to tissue heterogeneity and fibrotic remodelling. Multiphasic and fractionated electrograms are rarely observed in homogeneous tissue models but occur frequently where abrupt conductivity changes or patchy heterogeneity are present [79]. Patchy fibrosis produces markedly more fractionated electrograms than diffuse fibrosis, with simulated electrograms exhibiting up to ten zero crossings, comparable to clinically observed levels [94]. Regions with increased fibrosis show prolonged fractionation duration, and increased electrical activity, and fractionated electrograms preferentially occur in regions with high phase singularity density and elevated late gadolinium enhancement intensity [95]. These findings support the hypothesis that atrial fibrosis contributes to the maintenance of persistent AF by stabilising rotors and disrupting normal wave propagation, making fractionated electrograms indirect markers of fibrotic AF-driving substrate rather than independent arrhythmogenic sources [95]. Nevertheless, fractionation interpretation remains challenging because different electrophysiological mechanisms and recording device properties can produce similar fractionated morphologies [66].

Electrogram duration and timing

Conduction slowing and delayed activation substantially prolong electrogram waveforms, making duration and timing descriptors important markers of abnormal substrate. Prolonged EGM duration is strongly associated with slow conduction, fibrosis, and scar formation [88, 89, 85]. Slow conduction zones are commonly characterised by long-duration EGMs accompanied by low amplitudes and fractionation, whereas shorter, higher-voltage EGMs are more indicative of wavefront collisions [85]. Prolonged bipolar EGM duration independently associates with scar tissue irrespective of its intramural location [88], and EGM duration increases with collagen deposition [89].

Quantitative timing thresholds further support duration as a scar marker. Late potentials and prolonged electrograms accurately identify areas of anatomical scar [93], and infarct age correlates significantly with bipolar endocardial EGM width [96]. In clinical mapping, EGMs exceeding 70 ms corresponded to 99% of dense scar regions (< 0.2 mV) and 95% of scar border zone regions (0.2-1.0 mV) [97]. These thresholds highlight the practical utility of EGM duration in identifying scar tissue, though sensitivity to recording configuration and propagation direction limits its use as a standalone marker.

Waveform area

Waveform area integrates both amplitude and temporal properties of the EGM signal, capturing conduction slowing and structural remodelling that may not be fully reflected by either measure alone. Clinically, EGM area quantifies local conduction delay [98], and increasing collagen deposition reduces EGM area in a graded fashion [89].

Temporal derivatives

The temporal derivatives of electrogram waveforms, including upstroke and downstroke characteristics, reflect local activation dynamics and abrupt voltage transitions caused by heterogeneous propagation. Early studies reported that extracellular electrogram derivatives become

increasingly complex and polyphasic in aged and structurally disrupted tissue, reflecting microscopic conduction disturbances and zigzag propagation from fibrosis-induced electrical uncoupling [77]. Simulations of heterogeneous conduction and anisotropic propagation reproduce similar derivative changes [79], supporting their use as markers of fibrotic or infarcted tissue. Derivative-based features are sensitive to local propagation geometry, however, and their morphology can vary substantially with wavefront direction and electrode orientation.

Combined interpretation of EGM features

Many EGM waveform features are strongly interdependent, as conduction slowing, anisotropic propagation, fibrotic remodelling, and wavefront collision simultaneously alter amplitude, duration, fractionation, asymmetry, and entropy-related measures. Multiple features therefore covary within regions of heterogeneous conduction or structural remodelling [42, 74, 33], and regions exhibiting prolonged and fractionated electrograms are commonly characterised by reduced amplitudes and increased waveform irregularity as well [85, 42]. These interdependencies mean that substrate characterisation based on any single feature risks both misclassification and loss of complementary information.

Different tissue configurations and propagation mechanisms can produce overlapping electrogram morphologies, making it difficult to infer local conduction properties or fibrosis topology from a single EGM characteristic. Distinguishing structural fibrosis from functional slow-conduction regions based solely on electrogram morphology remains challenging [34, 99, 45]. Multidimensional approaches integrating multiple EGM features have therefore been explored to improve characterisation of conduction heterogeneity and arrhythmogenic substrate [74, 100, 42].

1.5.3 Modelling of fibrotic tissue and the effect on uEGMs

Fibrotic remodelling alters both the structural organisation and electrophysiological properties of cardiac tissue, producing conduction slowing, conduction heterogeneity, and arrhythmogenic activity [101, 18, 77]. These alterations substantially modify excitation propagation and electrogram morphology, making fibrosis modelling central to the interpretation of unipolar electrograms and atrial substrate characterisation [79, 94, 32]. Because different fibrosis representations produce distinct activation dynamics and electrogram morphologies, model choice directly influences the inferred relationship between tissue structure and waveform characteristics [102, 95, 32].

Conductivity-based fibrosis representations

In continuum tissue models, fibrosis is commonly represented through local reductions in tissue conductivity or diffusion coefficients, reproducing impaired electrical coupling and conduction slowing without explicitly modelling fibroblasts [103, 104, 105]. These reductions reproduce clinically relevant propagation abnormalities, including slow-conduction pathways, wavefront fragmentation, conduction block, and stable reentrant activity associated with atrial fibrillation and atrial tachycardia [101, 106, 105]. Because cardiac conduction is intrinsically anisotropic, fibrosis can additionally alter preferential propagation pathways and increase directional conduction heterogeneity [18, 77].

Conductivity scaling factors generate regions with varying degrees of conduction impairment, ranging from mildly heterogeneous border zones to severely infarcted or electrically disconnected

tissue. Spatial conductivity gradients and probabilistic fibrosis distributions can further reproduce diffuse fibrosis patterns and heterogeneous infarct borders, avoiding unrealistically abrupt transitions between healthy and diseased myocardium [107, 108, 109]. The spatial organisation of these heterogeneities strongly influences activation dynamics, as small-scale or highly heterogeneous fibrotic structures promote conduction block, wave break formation, action potential dispersion, and stabilisation of reentrant activity [107, 101, 108]. Conductivity-based models offer computationally efficient frameworks for large-scale simulations but may not fully capture microscopic conduction discontinuities or the electrophysiological effects of explicit fibroblast-myocyte interactions [102, 95, 56].

Structural and cellular fibrosis models

Two approaches extend beyond continuum conductivity reductions by incorporating explicit cellular or structural heterogeneity. In fibroblast-myocyte coupling models, fibroblast proliferation and electrical coupling to atrial myocytes reduce excitability, shorten action potential duration, slow conduction, and produce increasingly fractionated atrial electrograms during AF [110, 56]. Structural percolation models represent fibrosis through randomly distributed low- or non-conductive regions that disrupt normal propagation pathways. As tissue connectivity approaches the percolation threshold, highly irregular activation patterns, conduction block, wave break formation, and sustained reentry emerge despite preserved global propagation [109, 108]. Both approaches produce discontinuous propagation patterns strongly associated with prolonged electrogram duration, increased fractionation, and complex waveform morphologies [91, 94, 111].

Fibrosis topology and spatial organisation

Fibrosis topology, not just fibrosis burden, determines activation dynamics and electrogram morphology. Random fibrosis distributions, Gaussian conductivity fields, MRI-derived fibrosis maps, and statistically generated fibrosis textures have each been used to capture different aspects of clinically realistic conduction heterogeneity [111, 103, 112]. Histological analyses show that fibrotic structures span multiple spatial scales and frequently exhibit elongated and anisotropic organisation within the myocardial wall [112]. Different fibrosis topologies can therefore produce substantially different propagation patterns and electrogram morphologies despite similar overall fibrosis density, with patchy fibrosis producing markedly more fractionated electrograms and more heterogeneous conduction than diffuse distributions [94, 95, 111]. Fibrosis density alone is therefore insufficient to predict electrophysiological behaviour, and accurate substrate modelling requires explicit representation of fibrosis microstructure and spatial connectivity.

Electrogram morphology and modelling considerations

Fibrotic remodelling is associated with reduced electrogram amplitude, prolonged duration, waveform asymmetry, and increased fractionation, driven by asynchronous activation and discontinuous propagation [79, 90, 42]. In regions with increased fibrosis density and conduction heterogeneity, fractionation duration, waveform irregularity, and phase singularity density during AF are further elevated [95, 89, 42]. Electrogram morphology is also strongly influenced by electrode configuration, tissue thickness, catheter orientation, and wavefront directionality; however, abnormal waveform patterns cannot serve as direct surrogates for fibrosis. Different tissue configurations and propagation mechanisms can produce overlapping electrogram morphologies, complicating the distinction between structural fibrosis and functional slow-conduction regions from waveform characteristics alone [45, 40, 43].

Bidomain and cell-resolved models reproduce extracellular potentials and clinical electrograms with high fidelity but carry substantially increased computational cost [53, 54, 71]. Cell-resolved approaches may better capture microstructural conduction discontinuities in small tissue samples, whereas continuum monodomain and bidomain approaches remain more suitable for large-scale simulations [113, 52]. Realistic electrogram simulation additionally requires careful consideration of electrode geometry, near-field and far-field contributions, measurement noise, and signal preprocessing [82, 84, 74]. Accurate estimation of local conduction properties and fibrosis topology from electrogram morphology therefore, remains an open problem, particularly in heterogeneous tissue and clinically realistic mapping environments.

1.5.4 Analysis of EGM signals

Conventional substrate characterisation methods, such as voltage mapping and phase mapping, provide simplified representations of cardiac electrical activity that do not uniquely characterise local electrophysiological properties [33, 37]. Low-voltage regions do not necessarily correspond to fibrotic tissue, and abnormal conduction frequently occurs in regions with relatively preserved EGM amplitudes [37, 40, 99, 34]. These limitations have driven interest in complete EGM morphology analysis and the extraction of multiple signal descriptors capable of characterising conduction heterogeneity, slow conduction, anisotropic propagation, and arrhythmogenic substrate beyond conventional voltage thresholds.

Feature extraction approaches combining amplitude descriptors, temporal parameters, fractionation measures, entropy metrics, frequency-domain descriptors, waveform derivatives, and non-linear complexity measures have been investigated using both clinical recordings and computational models [100, 74]. Simulated patient-specific ventricular models derived from DE-MRI demonstrated strong separability between healthy and abnormal tissue using features including signal range, inflexion point count, signal energy, dominant frequency, mean slope, fractionation index, and extrema amplitudes [100]. Hybrid *in silico* and clinical atrial datasets with explicit fibrosis modelling showed that combinations of peak-to-peak amplitude, activity duration, sample entropy, spectral entropy, Shannon entropy, Kolmogorov complexity, and fractal dimension enabled accurate characterisation of fibrotic atrial substrate [74]. Together, these studies establish that EGM morphology encodes multidimensional information about conduction properties that voltage mapping alone cannot capture.

EGM morphology and conduction velocity

EGM morphology reflects local conduction properties across both structural and functional substrates. Regional conduction velocity estimation from multichannel atrial EGMs revealed strong correlations with peak-to-peak voltage and moderate correlations with fractionation duration [87, 42]. High-density mapping further showed that fractionated EGMs with relatively preserved voltage amplitudes frequently correspond to slow conduction regions rather than exclusively low-voltage scar [99]. In patients with multiple atrial tachycardias, functional conduction block regions contained fractionated and prolonged EGMs associated with anisotropic propagation and conduction slowing [114]. The precise relationship between tissue conductivity, fibrosis organisation, and extracellular signal morphology remains incompletely understood. Computational models address this gap by allowing systematic investigation of EGM origins in controlled settings.

Simulated fibrosis models allow tissue conductivity, fibrosis density, transmural, and structural organisation to be systematically varied while directly observing the resulting EGM morphologies. Patient-specific fibrosis simulations demonstrated that fibrotic tissue arrangement and

density strongly influence EGM fractionation, activation discontinuities, and waveform complexity [111, 94, 91]. Conductivity-based models further showed that heterogeneous propagation and reentrant activity emerge even without fully non-conductive scar regions, highlighting the role of conduction heterogeneity beyond tissue inactivation [101, 108]. EGM morphology depends strongly on the spatial scale and organisation of fibrosis, and averaged or homogeneous fibrosis representations may therefore fail to reproduce clinically realistic conduction patterns [95, 102].

Dimensionality reduction and feature analysis

Correlations and overlap between EGM morphology metrics make interpretation based on individual features increasingly difficult as the number of proposed descriptors grows [74, 100, 42], motivating multivariate statistical analysis and dimensionality reduction approaches. Principal component analysis (PCA) identifies linear combinations of features that capture the dominant variance in EGM morphology [115, 116], enabling visualisation of high-dimensional signal relationships and revealing latent feature combinations associated with conduction properties or fibrotic remodelling. More broadly, dimensionality reduction facilitates the interpretation of large EGM datasets and reduces the risk of overfitting in subsequent classification or regression tasks.

Machine learning for substrate characterisation

Supervised learning approaches have demonstrated promising performance for automated characterisation of fibrotic substrate and conduction abnormalities. A hybrid *in silico* and *in vivo* atrial dataset combining simulated fibrosis patterns with clinical electrograms enabled decision-tree classifiers to distinguish fibrosis density and transmuralities with accuracies above 95%, substantially outperforming voltage-threshold approaches [74]. Deep neural networks trained on simulated EGM datasets with randomised scar distributions successfully estimated tissue conductivity and localised scar regions with high accuracy [117]. Frameworks based on deep learning have additionally been applied to automatically delineate local and far-field EGM components and identify decremental conduction patterns associated with accessory pathways and atrioventricular re-entrant tachycardias, achieving sensitivities and specificities above 95% in clinical validation [118]. These results demonstrate that supervised integration of multiple EGM-derived features substantially improves substrate characterisation beyond single-feature analysis.

Limitations and open challenges

Several limitations constrain the current state of EGM-based substrate characterisation. Many machine learning approaches rely heavily on simulated training data, and the electrophysiological realism of fibrosis models and electrogram simulations remains an open question [32, 102]. EGM morphology is strongly influenced by electrode size, catheter orientation, wavefront directionality, tissue contact, and near-field and far-field contributions [45, 38, 43]. Catheter geometry and electrode configuration additionally affect derived quantities such as local activation time, CV, fractionation measures, dominant frequency, and entropy-based descriptors [92, 36]. These dependencies complicate the establishment of universally applicable feature definitions and thresholds across mapping systems and recording conditions. Despite the large number of proposed EGM descriptors, systematic assessment of their latent structure and relative contribution to substrate characterisation remains limited [74, 34, 33].

The interpretation of abnormal EGM morphology itself poses an additional challenge. Frac-

tionated, prolonged, or low-voltage EGMs may originate from structural fibrosis, anisotropic propagation, functional conduction block, wavefront collision, pivot points, or far-field activity, making it difficult to attribute abnormal waveforms to a single underlying mechanism [85, 99, 42]. Distinguishing structural scar from functional slow-conduction regions based solely on EGM morphology therefore remains challenging [99, 37, 34, 33], and accurate estimation of local conduction properties from EGM signals remains an open problem in heterogeneous tissue and clinical mapping environments [119, 99, 33, 34]. The extent to which different fibrosis topologies produce distinguishable EGM morphologies is similarly incompletely understood [95, 102, 111]. Current research therefore focuses on combining electrophysiological modelling, high-density mapping, multivariate feature analysis, and machine learning to improve substrate characterisation and better relate EGM morphology to underlying tissue properties [34, 33, 120].

Chapter 2

Introduction

Cardiac electrograms (EGMs) are an important tool for identifying arrhythmogenic and scar tissue [34, 75], guiding electrophysiologists during ablation procedures. In particular, ablation therapy often targets re-entry, a mechanism in which electrical activation repeatedly circulates through cardiac tissue and sustains an arrhythmia [27]. Voltage mapping and electrogram morphology analysis are widely used for substrate characterisation [40, 99, 37], with features such as low voltage [86], prolonged duration [97], and fractionation [93] commonly associated with pathological remodelling. However, EGM signals are difficult to interpret visually, as factors such as electrode configuration, tissue contact, and wavefront direction can substantially alter recorded amplitudes, durations, and waveform morphologies [33, 34, 43]. Several diagnostic measures are furthermore poorly defined, complicating comparisons between patients and studies [66]. Most fundamentally, distinguishing scar tissue from functional slow-conduction regions based solely on electrogram morphology remains challenging, as similar EGM abnormalities can arise from fundamentally different electrophysiological mechanisms [45, 99]. This distinction is essential in ablation therapy, as re-entrant circuits of electrical activity often pass through narrow, slow conductive pathways adjacent to non-conductive scar tissue. Moreover, electrogram fractionation can occur in normal voltage atrial tissue in patients with atrial fibrillation and hence is not uniquely indicative of infarcted tissue [99].

Computational studies can help address these challenges by enabling controlled investigation of the factors that modify electrogram morphology and amplitude [32, 79, 94]. Different fibrosis representations produce different activation dynamics and electrogram morphologies, but partially overlapping waveform characteristics can arise from distinct tissue configurations, complicating electrogram-based substrate inference [102, 95, 111]. To enable quantitative analysis, features such as amplitude [45, 85, 86], asymmetry [90, 91], fractionation [93, 87], timing [88, 89], and signal area [98] are commonly extracted, and computational studies have demonstrated strong associations between these features and local conduction properties [79, 91, 94]. Building on this, supervised machine learning models using multidimensional EGM features have shown promise for inferring fibrosis properties and conduction abnormalities, often outperforming conventional voltage thresholds [74, 117, 118], although external validation remains to be confirmed [34]. Nevertheless, the relative contribution, redundancy, and latent structure of EGM-derived features remain insufficiently characterised [74, 33, 34]. Furthermore, many computational studies rely on study-specific sets of EGM features, typically guided by commonly used clinical descriptors or features reported in previous work, rather than by a systematic evaluation of their relative diagnostic value [74, 34, 33]. In addition, the extent to which different fibrosis topologies and tissue configurations produce distinguishable EGM morphologies remains incompletely understood [95, 102, 111].

To address these issues, this thesis investigates whether electrogram-derived features can distinguish scar tissue from functional slow-conduction regions in simulated atrial cardiac tissue. The analysis progresses from simplified homogeneous slow-conduction regions to increasingly heterogeneous tissue configurations, investigating the qualitative effect of the simulation parameters on the EGM signal morphology. Quantitatively, 11 electrogram features are extracted from each signal, and principal component analysis (PCA) is applied to investigate feature redundancy and dominant modes of variation. Finally, supervised machine learning approaches are used to evaluate whether the extracted features can predict local conduction velocity. The simulated datasets reproduced key EGM characteristics, while PCA revealed that EGM morphology is largely described by a low-dimensional feature space dominated by signal-strength-related features, with the resulting principal components correlating well with conduction velocity and enabling its prediction with a mean absolute error of 1.2%.

Chapter 3

Methods

This section will describe the methodology used in detail. The first subsection outlines the basis of the setup in Finitewave, which entails all of the setup elements that are the same for all simulation types. In the next subsection, the different types of simulations and their parameters will be explained, and it will be emphasised how each type of simulation lowers the symmetry and more closely approaches realistic infarcted tissue. Some of the stored properties related to the setup are also discussed. The final subsection will explain the analysis of the EGM signals and is subdivided into three sections. One of these will be dedicated to explaining both the calculation and relevance of all the features, while the second section outlines the basics of the PCA analysis, and the third section will describe the supervised machine learning methods.

3.1 Setup of the simulation in Finitewave

All simulations were performed with the Finitewave simulation package developed by the biophysics group at Ghent University. Finitewave is a Python package for simulating cardiac electrophysiology using finite-difference methods. It provides tools for modelling and visualising the propagation of electrical waves in cardiac tissue [65]. The following sections will explain the general setup of the simulation and the reasoning behind each choice. The integration scheme, cardiac model, prepacing, cell mesh, stimuli and EGM tracking will all be discussed in detail.

3.1.1 Integration scheme

Finitewave uses the monodomain approach (see section 1.4.1) and solves the equations by use of the FTCS (forward in time and centred in space) scheme. As with any partial differential equation, numerically solving the monodomain equations with the FTCS scheme requires careful consideration for temporal and spatial discretisation. Choosing the timestep too large relative to the spatial step can cause the algorithm to diverge, resulting in an unstable integration scheme. The stability criterion for this particular case is [121]:

$$\frac{D\tau}{h^2} \leq \frac{1}{2} \quad (3.1)$$

where D is the diffusion coefficient, h is the spatial discretisation step, and τ is the temporal discretisation step. In the simulations below, the used timestep was $\tau = 0.025$ ms while the used spatial step was $h = 0.3$ mm. Entering these parameters into equation 3.1, one finds that the maximum allowed diffusion coefficient is $1.8 \text{ mm}^2/\text{ms}$. Since the maximum diffusion coefficient

for the simulations outlined below was $0.154 \text{ mm}^2/\text{ms}$ (see table A.1 in appendix A.1), the integration scheme is always well inside the stable regime. Larger values for both h and τ could have been chosen, however, this resulted in slightly non-uniform amplitudes for EGM signals recorded in a uniform, healthy tissue and was hence not done. Each simulation ran for a total of 250 ms.

3.1.2 The cardiac model

All simulations were performed using the Courtemanche-Ramirez-Nattel model with all parameters given in Table A.1, located in Appendix A.1. These parameters are the same as those in the original paper [59]. The Courtemanche-Ramirez-Nattel model is a realistic atrial model, as it explicitly models all of the ionic currents, and one can hence expect it to render realistic EGM signals. Since this model is a single-cell model, it does not provide a value for the diffusion coefficient or background current parameters, which are tissue-level parameters and hence needed for tissue-level modelling. These two parameters are shown at the bottom of A.1 and are closely aligned with existing literature [52, 60].

3.1.3 The cardiac cell mesh

One can describe cardiac myocytes as 'fibres' because of their distinct shape (see section 1.2.2). Anisotropy in these fibres can significantly reduce the speed at which electrical waves propagate [18], and it is hence important to take this into account in the simulations. Finitewave represents the fibre field as a 3D float array, which specifies the 2D vector of the fibre direction at every node in the mesh. All simulations discussed below used an isotropic mesh with all fibres aligned in the same direction. The cardiac mesh itself consists of a 100×100 grid with a spatial discretisation of 0.3 mm, hence this corresponds to $30 \text{ mm} \times 30 \text{ mm}$.

3.1.4 Prepacing

Ionic cardiac models, like the Courtemanche model, have a short-term memory in the sense that the ion concentrations and recovery variables depend on previous simulation states [62, 63]. Starting the simulation from arbitrary initial conditions can hence cause several issues, such as gating variables not being in equilibrium, the action potential not being representative, and the APD being unstable. To circumvent these issues, one can prepace the cardiac tissue mesh, which entails repetitively stimulating the tissue such that all relevant variables (gating variables, APD, ...) can reach an equilibrium value, i.e. they do not vary from one heartbeat to the next [64]. For the Courtemanche model in particular, it has been shown that it never truly reaches an equilibrium state, and the current state also depends on the pacing of the beats [61]. This means that applying prepacing will only get the model to a somewhat realistic state, and care must be taken to ensure that the prepacing is exactly the same for all performed simulations, such that the results can be compared against each other.

In the simulations below, prepacing was performed by applying 10 stimuli to a rectangle on the left of the tissue with a width of 3 ($= 0.9 \text{ mm}$) across the full height of the cardiac tissue mesh. These ten beats were given at an interval of 250 ms, and a 500 ms cool-down period was implemented to ensure that all propagating signals had died out. The simulation state after these 3000 ms was then saved with Finitewave's *StateSaver* to a given directory. From there, it can be loaded with the *StateLoader* at the start of every simulation, after which infarcted tissue can be placed on the cardiac mesh. This way, the prepacing needs to only be performed

once, which lightens the computational load of the simulations. This method also ensures that all simulations start from the exact same initial state.

3.1.5 Applying stimuli

Excitation of the cardiac tissue was done by means of current stimulation, which is the most physically correct approach [122, 123]. Finitewave applies these stimuli through the *StimCurrent* class hierarchy. The applied stimulus had a magnitude of 2 nA and a duration of 2 ms, which are the same values as those used in the original Courtemanche model [59]. Since the program units for time are milliseconds, the applied stimulus lasts for 2 program units of time. For current however, the program units are actually a current density for the Courtemanche model:

$$I_{dens} = \frac{I_{total}}{C_m} \quad (3.2)$$

with units of pA/pF . Since C_m is equal to 100 pF, 2 nA of actual current corresponds to 20 program units of current density.

In the simulations below, the stimuli were always applied on the left side of the cardiac mesh. Since the Courtemanche model is a detailed ionic model, the applied stimulus must be sufficiently strong and distributed over a sufficiently large area to ensure successful initiation and propagation of the action potential [124, 125]. For all simulations except for *applying stimuli at an angle*, a rectangular stimulus of width 3 (= 0.9 mm) and the height of the entire cardiac mesh (30 mm) was applied with duration and amplitude given above. For the *applying stimuli at an angle* simulations, a circular stimulus of radius 3 (= 0.9 mm) was used. For all of these stimulations, it was visually verified that the stimulus actually propagates.

3.1.6 Setting up the EGM electrodes

The extraction of the EGM signals happens through Finitewave's *ECG2DTracker*. This tracker allows one to extract the extracellular potential at a given location by use of equation 1.7. This extraction was done every 10 timesteps (= 0.25 ms) for the duration of the simulation on an evenly spaced grid of electrodes, where the separation was 4 program units (= 1.2 mm). This separation was chosen because it closely resembles the current accuracy of catheters used in clinical settings, which is around 1 mm [126, 127]. Furthermore, no EGM electrodes were placed near the boundaries, such that boundary effects did not influence the measured results and no electrode was located directly in the stimulated area. Concretely, this means that the grid starts at x (or y) equal to 6 and runs up to x (or y) 94 program units with steps of 4 program units. Also note that all EGM electrodes are located at the same height of 1 program unit (= 300 μm). Although an electrode height of 300 μm is used in some specialised catheters, it is not the norm, and most catheters use a height of several millimetres [43]. However, for a systematic analysis, it is most important that the height of all electrodes is the same, such that they can be compared against each other. The exact height is then of minimal importance and can be seen as an arbitrary decision.

3.2 5 Datasets of EGM signals in increasingly complex S1 simulations

As stated in the introduction, the goal of this thesis is to assess which EGM features are most important to distinguish between slow conduction zones and scar tissue. Because EGM signals

can become very complex, this thesis will start from a very simple simulation and gradually increase the complexity by breaking symmetries in the setup. The first simulation type is the simplest and uses a rectangle of uniform conduction with a symmetric stimulus. The other four simulation types then introduced 1 layer of complexity on top of this basic simulation. However, the combination of these different complexities in one simulation was not explored. Hence, only one thing was changed in each of the more complex simulations.

The simplest performed simulation consists of one rectangle of infarcted tissue with modified conduction. Secondly, instead of having the conduction jump at the boundary, a gradient in the conduction was introduced. Next, the stimulus was applied from a variable height such that the timing between the different electrodes would differ over several simulations. After this, a random boundary was added to the infarcted tissue such that the wave would break up in unpredictable ways. Lastly, a fully randomised infarcted tissue was created. The following subsections will discuss each of these scenarios in detail, together with their choice of parameters.

3.2.1 Dataset 1: Modified conduction simulations

Finitewave allows for the conductivity to be altered by locally modifying the diffusion coefficient. The conductivity is a 2D array containing values between 0 (no conduction) and 1 (normal conduction), which is then multiplied by the diffusion coefficient of the model to get the actual diffusion of the region [128]. Note that tissue with a conduction of 0 is essentially the same as scar tissue since no electrical signals can propagate through. A visual inspection of a simple simulation and its produced electrogram signals showed that a conduction of 0.01 also did not show any wave propagation or a non-zero EGM signal. A conduction of 0.02 on the other hand did show a propagating wave and a non-zero EGM signal. Hence, the definition of non-conductive tissue can be set to be a conduction lower than or equal to 0.01.

Modelling infarcted tissue as a rectangle is not a terrible approximation since scars are often elongated [112]. Furthermore, blocking or slowing down of the wavefront happens regardless of the complexity of the infarcted tissue, and this simple setup helps keep the recorded signals as simple as possible for a first analysis and interpretation [114].

The infarcted tissue was always placed at $y = 10$ and had a height of 80 program units (= 24 mm). The width was always varied symmetrically across the middle of the cardiac tissue, so $x = 50$ was always the middle of the infarcted tissue. Smaller scars are harder to detect with currently available techniques, such as phase mapping or analysis of EGM signals [129, 130, 131, 132], and hence, this thesis focuses on smaller scars. Because the goal is to differentiate between scar tissue and slow conduction zones, lower conduction velocities were sampled more densely than higher conduction velocities. The exact values of both the *width* and the *conduction velocities* used in the simulations can be found in table 3.1. Note that both the used values of the width and the conduction velocities are arbitrary. They were chosen to sample the phase space as well as possible for the intended goal while also keeping the total compute time feasible. Two illustrations of the modified conduction setup are shown on the top left and top middle of figure 3.1.

3.2.2 Dataset 2: Conduction gradient simulations

The next type of simulation introduced a spatial gradient in conduction velocity along the propagation direction to avoid unrealistically abrupt changes and to generate a broader range of local conduction states, thereby producing more heterogeneous tissue and potentially more

fractionated EGMs [94]. This approaches real infarcted tissue more closely since conduction velocity changes happen more gradually than in section 3.2.1 [133].

Once again, the height of the infarcted tissue was 80 program units (= 24 mm) and the simulation was completely symmetric about $x = 50$. Conduction gradient simulations have three parameters: the *total width*, the *minimum conduction* in the centre, and a *minimum conduction width* which indicates the width along which the conduction is constant, with the value of this conduction being the minimum conduction. The used values of these three parameters can be found in table 3.1.

As an illustration of the meaning of these parameters, consider a simulation with a total width of 9 mm, a constant width of 1.2 mm and a minimum conduction of 0.1, as shown in the top right panel of figure 3.1. The conduction in the 1.2 mm in the centre of the simulation will be 0.1. Subtracting this constant region from the total width gives 7.8 mm, which is the total width over which the gradient takes place. The gradient will first decrease on the left of the constant area and then increase again on the right of the constant area, hence the full gradient falls over half of this width, which is 3.9 mm. The value of the conduction velocity will hence drop continuously from 1.0 to 0.1 over an area of 3.9 mm, resulting in a gradient of 0.231/mm.

To summarise, the excitation wave starts from the left of the cardiac mesh and will first encounter a conduction velocity gradient of -0.231/mm for 3.9 mm, after which the conduction will have reached 0.1. This value is then constant for 1.2 mm, after which the conduction velocity will increase again at a rate of 0.231/mm for 3.9 mm. After this, the excitation wave is now in healthy tissue, and the conduction velocity is equal to 1.

3.2.3 Dataset 3: Modified stimulus height simulations

Up until now, the stimulus has always been applied in a highly symmetric way, meaning that the timing in several EGM electrodes is often identical. For example, all electrodes located on the same x value on the left of the infarcted tissue will register virtually the same signal. In clinical practice or more advanced simulations, however, one will have no or only very little control over the angle under which the excitation wave hits the infarcted area, since it can have an arbitrary orientation.

To test the effect of a different stimulus orientation, a circular stimulus with radius 3 (= 0.9 mm) will now be applied from various heights. This type of stimulation will ensure that EGM electrodes located in the same x coordinate will now develop a different timing of the recorded signal, since the stimulus will reach them at different times.

One important note for this type of simulation is that the recorded signal at an electrode located at $y = 75$ for a stimulus in the top left corner and the signal recorded at an electrode located at $y = 25$ for a stimulus in the bottom left corner will be exactly the same because the setup is symmetric around $y = 50$. That is why the stimuli will only be applied with $y \geq 50$. Since the radius of the applied stimulus is equal to 3, steps of 6 are the logical choice. The used values of all three parameters can be found in table 3.1. Note that the conduction velocity and width sampling were done less densely than for other simulations in order to reduce the total number of simulations needed and keep the total compute time feasible. A detailed setup of this simulation type is shown in the bottom left panel of figure 3.1. As before, the emphasis lies on very low conductivities and smaller infarcted areas. Also note that a conduction of 0.0 is equivalent to scar tissue.

3.2.4 Dataset 4: Probabilistic boundary simulations

So far, all of the generated infarcted areas have been perfect rectangles, either with or without a gradient. The next step in the complication process is to add a degree of asymmetry or randomness to the infarcted area since real infarcted tissue has irregular shapes. Furthermore, real scars often contain 'patches' or a mix between healthy and infarcted tissue around their boundary [112, 101].

In the simulations performed here, this was done by adding a random boundary to the existing rectangle. The width, height, and conduction were still specified beforehand, but now the boundary of the infarcted region was added in a probabilistic way. This boundary extends 5 nodes deep into the tissue, where each node has a decreasing probability of being infarcted. Infarcted nodes then get the same conductivity as the main rectangle of infarcted tissue. The probabilities for infarction were: 0.7 for the first node, 0.6 for the second, 0.5 for the third, 0.4 for the fourth and 0.3 for the fifth. An easy implementation for this algorithm is to draw a random number r from a uniform distribution between 0 and 1 for every node that is located at a distance smaller than or equal to 5 from the infarcted area. Each node is then added to the infarcted tissue if r is smaller than the given probability for that node. In all of these simulations, the height of the infarcted tissue was 80 (= 24 mm). The values of the other parameters can be found in table 3.1. The setup of the probabilistic boundary simulations is shown in the bottom left panel of Figure 3.1.

3.2.5 Dataset 5: Random infarcted tissue

Instead of only randomising the boundary of the infarcted tissue, one can also fully randomise the affected area in an attempt to construct more complex EGM signals. Doing this further reduces the symmetry of the problem and more realistically replicates scars encountered in clinical practice. Similar approaches have already been used to simulate realistic EGM signals and to make meaningful predictions about the underlying tissue [117].

Implementing random infarcted tissue can be done by once again drawing a random number r from a uniform distribution between 0 and 1 for every node that is located inside the given boundary. Every node is then set to the given conduction velocity if its corresponding r is lower than the given probability. For these simulations, both the height and width were kept constant. The height amounted to 80 program units (= 24 mm) and the width was 50 program units (= 15 mm). The used values of the other parameters can be found in table 3.1. An example of a setup for the probabilistic infarct simulation is shown in the bottom right panel of figure 3.1.

3.2.6 Defining properties of the setup

In order to improve the quality of the post-processing of the data, it is desirable to save several properties of every simulation and EGM electrode. Useful information that one can extract from the setup is, for example, all of the varied parameters that were discussed in the previous subsections. Desirable properties to keep track of for the EGM electrodes are their position, simulation type with the parameters, an index to identify them by and the local conduction.

Here, the local conduction is defined as the conduction in the exact node where the EGM is placed. Since EGM electrodes behave like dipoles (see section 1.5.1), they will also pick up signals from neighbouring nodes. Hence, a physically better defined quantity is the averaged

Table 3.1: Values of all the varied parameters for each simulation type.

Simulation Type	Width [mm]	Minimum conduction []	Minimum Conduction Width [mm]	Stimulus Height [mm]	Infarction Probability []
Substrate	0.6 - 9.0, $\Delta 0.6$ 10.2 - 15.0, $\Delta 1.2$	0.00 - 0.10, $\Delta 0.01$ 0.15 - 1.00, $\Delta 0.05$	N.A.	N.A.	N.A.
Conduction Gradient	3.0 - 15.0, $\Delta 1.2$	0.00 - 0.10, $\Delta 0.01$ 0.15 - 1.00, $\Delta 0.05$	1.2 and 2.4	N.A.	N.A.
Variable Stimulus	0.6 - 3.0, $\Delta 0.6$ 5.4 - 15.0, $\Delta 2.4$	0.00 - 0.10, $\Delta 0.02$ 0.20 - 0.40, $\Delta 0.10$ 0.60 - 1.00, $\Delta 0.20$	N.A.	15.0 - 98.0, $\Delta 1.8$	N.A.
Probabilistic Boundary	0.6 - 9.0, $\Delta 0.6$ 10.2 - 15.0, $\Delta 1.2$	0.00 - 0.10, $\Delta 0.01$ 0.15 - 1.00, $\Delta 0.05$	N.A.	N.A.	N.A.
Diffuse Infarct	15.0	0.00 - 0.10, $\Delta 0.01$ 0.15 - 1.00, $\Delta 0.05$	N.A.	N.A.	0.05 - 1.00, $\Delta 0.05$

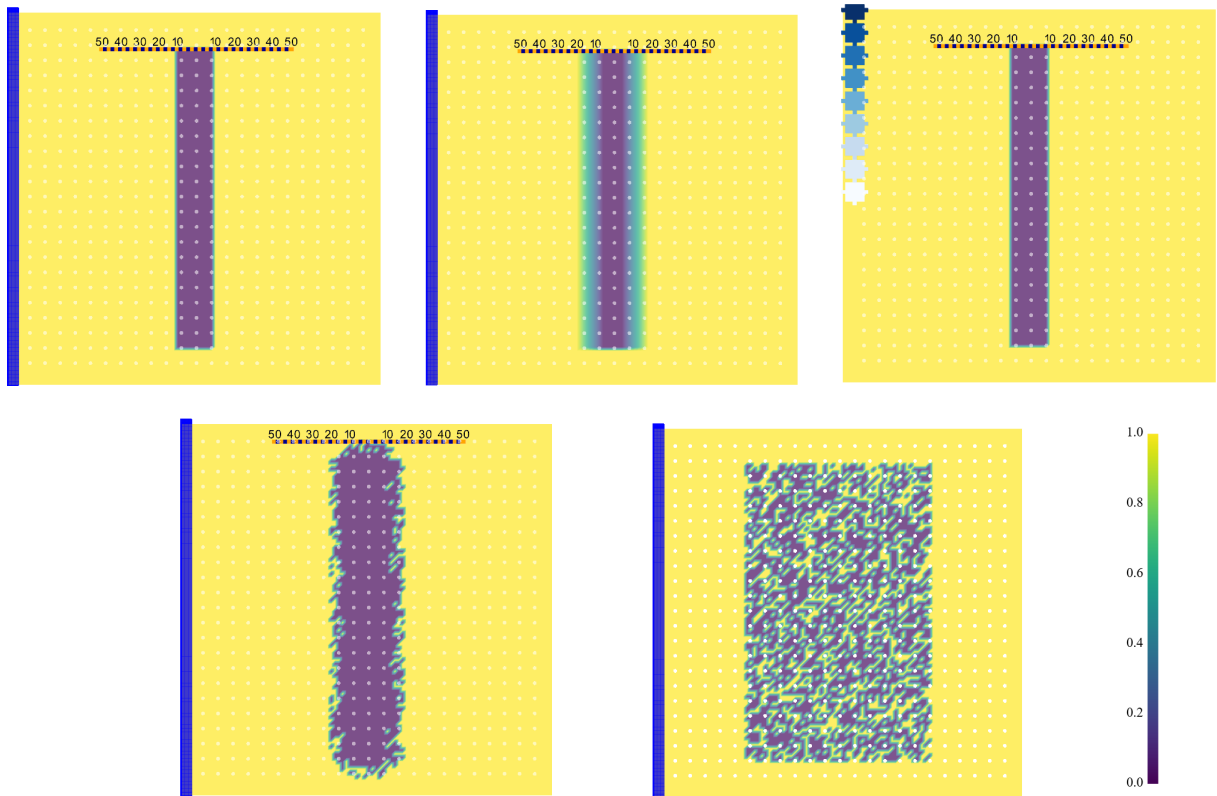


Figure 3.1: Overview of the substrate (top left), conduction gradient (top middle), variable stimulus (top right), probabilistic boundary (bottom left), and probabilistic infarct (bottom right) configurations. The colorbar on the bottom right shows the conduction velocities for all figures. The applied stimuli are represented by the blue bar on the left or by white/blue circles. Each white dot represents an EGM electrode

conduction, which can be calculated through the following equations:

$$w_i = \frac{1}{(n_0 - n_i)^2 + 1} \quad (3.3)$$

$$v(n_0) = \frac{\sum_{n_i \leq 4} w_i v_i}{\sum_{n_i \leq 4} w_i} \quad (3.4)$$

where w_i defines the weights for the weighted average, n_0 is the node at which the EGM is located, n_i are the neighbouring nodes that contribute to the average and v is the conduction velocity. A visual representation of this formula can be found in figure 3.2. The weights in equation 3.3 follow a $1/r^2$ dependence, inspired by the dipole-like behaviour of the EGM electrodes. The 1 can be seen as the vertical coordinate above the tissue and circumvents singularities in case $n_i = n_0$. Equation 3.4 cuts off after 4 nodes ($= 1.2$ mm), both to keep the calculation numerically tractable and to represent the accuracy of about 1 mm of EGM catheters in clinical practice [126, 127]. Furthermore, EGM electrodes themselves have a dimension of the order of 1 mm [33] while the simulated electrodes have the dimension of 1 mesh node (300 μ m). Hence, the use of the averaged conduction more closely represents the accuracy that real-world electrodes can accomplish, as electrode position information is typically as accurate as 1 mm [134].

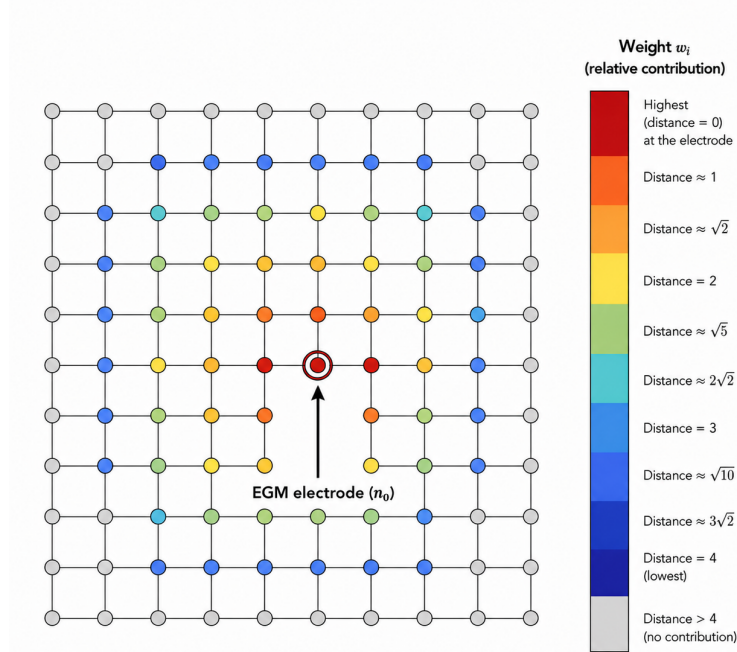


Figure 3.2: Illustration of the calculation of the averaged conduction. Each dot represents one node on the cardiac mesh and the EGM electrode is placed in the middle. All surrounding nodes with a distance less than 4 grid units contribute to the averaged conduction based on the inverse square distance to the electrode.

Introducing the averaged conduction is essentially a smoothing of the conduction matrix, and it will influence how boundaries and narrow regions of slow conduction zones are represented. Several examples of this smoothing for different simulation setups are shown in Figure A.1 in Appendix A.1. Consider for example an extreme case of this smoothing where an EGM electrode is located in the bottom right corner of a rectangular scar tissue of width 2 ($= 0.6$ mm) with conductivity 0.0. Calculating the averaged conduction, one finds a value of 0.666, which is a lot higher than the value of local conduction where the electrode is located, which is zero. However, one should keep in mind that an EGM electrode behaves as a dipole, and hence its entire

immediate neighbourhood will inevitably contribute to the recorded signal. So even though in this example, the electrode is located in scar tissue, it will still record signals that propagate through healthy tissue. Hence, when using the averaged conduction, one can only aim to predict the conduction with a resolution that is comparable to the smoothing area.

The distance cutoff is a free parameter of the model and different values will produce different results. A smaller cutoff value will incorporate fewer electrodes and hence retain more details of the initial cardiac mesh. On the other hand, averaging over fewer mesh points will take into account a smaller neighbourhood around each electrode and will miss parts of the tissue that also contribute to the EGM signal. This then makes it harder to relate the recorded EGM to an underlying averaged conduction. Conversely, a higher distance cutoff value will account for a much larger neighbourhood and improve the interpretability of the EGM signal, but it will inevitably lose the finer details of the cardiac mesh. Hence, when choosing the distance cutoff, one must weigh the importance of relating the EGM signal to the averaged conduction and keeping the resolution of the original mesh, as one cannot do both.

A final useful property is the distance of the EGM electrode to the closest infarcted tissue. One can calculate this by first calculating the Euclidean distance to all of the affected cells and then extracting the minimum with Python's built-in `.min()` method.

3.3 Analysis of EGM signals

In order to be able to distinguish between scar tissue and slow conduction zones, a systematic and quantifiable approach is needed. In clinical practice, EGM signals are still analysed visually, which is quite a demanding process and furthermore also difficult to quantify [33]. Moreover, diagnostic measures such as double potentials and fractionation do not have a precise definition, which again complicates the use of EGM signals in simulations to distinguish between scar tissue and slow conduction zones [66]. Furthermore, there is only very limited software available that can aid in the interpretation of EGM signals [135], although AI-based methods are becoming more widespread [120].

The following subsection will introduce 11 different features that can be derived from EGM signals, which can then be used for a more quantifiable approach. The subsection thereafter will then introduce PCA, a data reduction technique which can help remove existing redundancies between the features. The final subsection will then discuss two supervised machine learning techniques: the random forest regressor and the logistic regression classifier.

3.3.1 11 features extracted from the EGM signals

The first part of this systematic approach is to define several features that can be extracted from the EGM signals. All of the features discussed below were chosen such that they could also be applied in clinical situations. That is, they can either be read off from the EGM directly or be straightforwardly calculated. Several features, like the number of peaks or the range, are already used in clinical practice [32], while others are taken from other branches of cardiac electrophysiology. Furthermore, each feature should have a goal behind it, i.e. there should be a motivation as to why that feature could possibly distinguish between scar tissue, slow conduction zones, and healthy tissue. All of the features are: maximum amplitude, minimum amplitude, range, absolute area, signed area, number of peaks, time difference between minimum and maximum, maximum upstroke, maximum downstroke, Poincaré SD1/SD2 ratio, and maximum Teager-Kaiser energy. Figure 3.3 gives an overview of all features on both a simple EGM signal recorded in healthy tissue and a complicated signal recorded in a conduction gradient simulation.

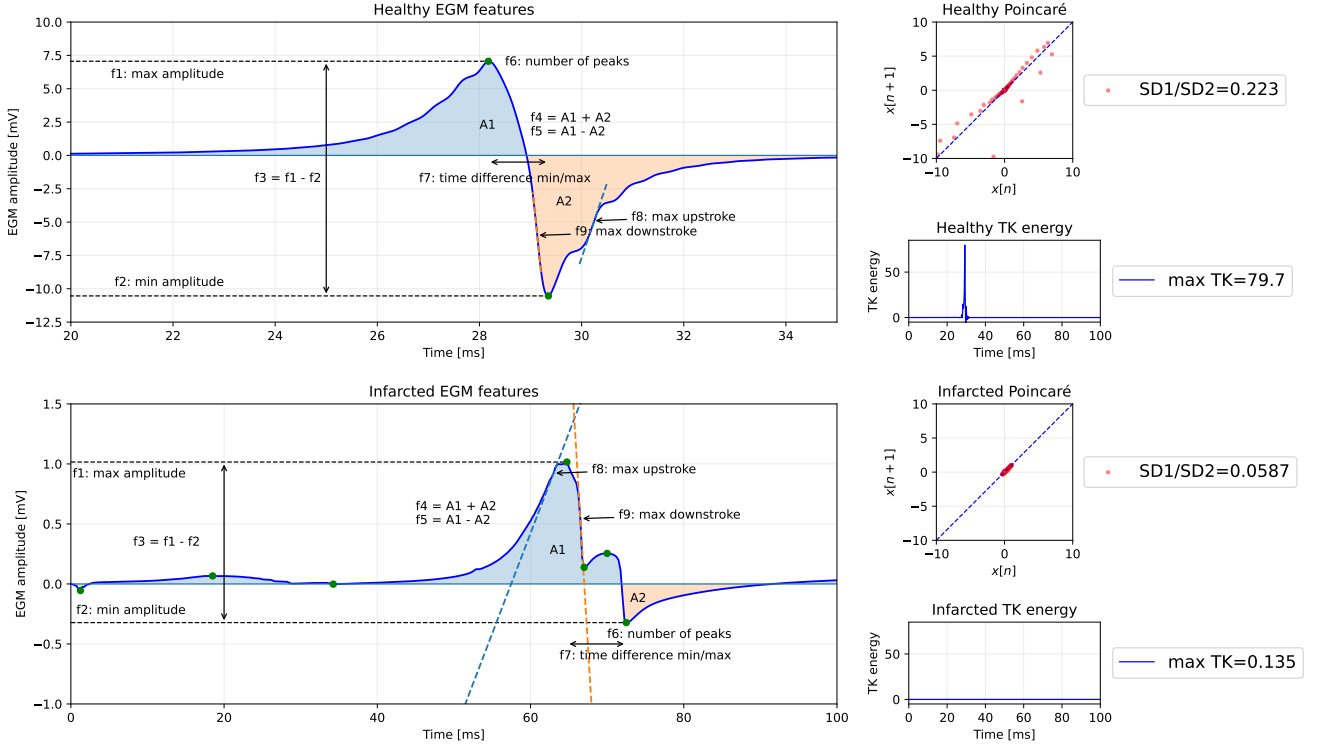


Figure 3.3: Visual representation of the 11 EGM features. The top panel shows a healthy EGM signal, while the bottom panel shows an infarcted EGM signal extracted from a conduction gradient simulation with a width of 10.2 mm and a conduction of 0.01. The blue line represents the EGM signal, the blue area is the total area above the $V=0$, and the orange area is the total area below $V=0$. The green dots indicate the identified peaks, while the orange and blue dashed lines indicate the maximum upstroke and downstroke. The features themselves are indicated with black arrows and/or black text. The Poincaré plot with descriptors is shown on the top right, and the Teager-Kaiser energy is shown below it. Note that both the x and y scales are different in the top and bottom panels.

Feature 1: maximum amplitude [mV]

This feature straightforwardly extracts the maximum of the EGM signal using Python's built-in `.max()` function. Since lower extracellular potentials are associated with scar tissue or slow conduction zones, the maximum amplitude is a useful feature [86]. The exact value will of course depend on several factors, such as size, conduction velocity and quality of the electrode contact. Slower conduction zones and larger infarcted areas will generally have lower maxima [74, 88].

Feature 2: minimum amplitude [mV]

The minimum amplitude of the EGM signal can straightforwardly be extracted with Python's built-in `.min()` function. Similarly to the maximum amplitude, the minimum will also depend on the size and the conduction value of the infarcted tissue. However, it has been shown that unipolar EGM signals can be asymmetric [90, 89, 91], so the maximum and the minimum need not necessarily be the same around infarcted tissue. Hence, including the minimum can help capture the asymmetry in the signal.

Feature 3: Range [mV]

The range is defined as the difference between the maximum and the minimum of the signal and is calculated here using Python’s built-in *.max()* and *.min()* functions. The signal range is already used in clinical analysis of EGM signals to identify scar tissue and conduction blocks, although it is sometimes also called the peak-to-peak voltage [76]. Hence, the range is a relevant feature besides the maximum and minimum, but it cannot detect asymmetries.

Feature 4: Absolute area [mV · ms]

This feature represents the total area underneath the absolute value of the EGM signal, such that all areas contribute positively towards the total sum. The calculation was done using NumPy’s *trapezoid* function on the absolute value of the EGM signal. The absolute area provides a different measure of the signal strength, besides the minimum, maximum and range. Evaluating the strength across the entire signal is useful because slow conduction and scar-related substrate tend to reduce the strength of the waveform everywhere, increase fractionation and double potentials, and prolong the signal duration [87]. Hence, the area under the absolute EGM signal provides an additional integrated measure of signal strength that is not yet covered in the maximum, minimum or range of the signal [98].

Feature 5: Signed area [mV · ms]

The signed area measures the total area underneath the EGM signal, but now with the recorded signal instead of the absolute value of the signal such that signals below zero now contribute negatively towards the total sum. The integral was once again performed with NumPy’s *trapezoid* function. Similar to the absolute area, the signed area is a measure of the strength of the signal, and it can help detect double potentials and fractionation. More importantly, it can also capture asymmetries in the signal since the signed area is zero for a perfectly symmetric signal. This is important because asymmetries in the signal are a major indicator of infarcted tissue. The relevancy of this feature is supported by earlier work, which has identified lower area and increased asymmetry as a consequence of infarcted tissue [89, 91, 90].

Feature 6: Number of peaks []

The number of peaks identifies the number of local maxima and minima that are present in the signal and serves to identify fractionation of the EGM signal, which is one of the most important diagnostic measures in cardiac electrophysiology [93, 42, 76, 88, 91]. The peak detection algorithm identifies local extrema by analysing changes in the sign of the discrete first derivative of the signal. First, the input signal is converted to a NumPy array and its discrete derivative is computed using *np.diff*, which calculates the difference between consecutive samples. The sign of this derivative is then determined using *np.sign*, resulting in an array containing values of +1 for increasing regions, -1 for decreasing regions, and 0 for flat segments. Since flat regions can interrupt the detection of extrema, zero-valued entries are removed by propagating neighbouring non-zero signs through the plateau. The indices of all non-zero derivative signs are obtained using *np.nonzero*; if no non-zero entries exist, the signal is completely flat and contains no extrema. Leading zeros are replaced by the first non-zero digit, after which all remaining zeros are forward-filled using the preceding sign value. This procedure ensures that flat plateaus inherit the direction of the surrounding trend. Finally, extrema are identified by counting sign reversals in the processed derivative-sign array. This is performed by comparing neighbouring entries and summing all locations where consecutive signs differ. A transition from +1 to -1 corresponds to

a local maximum, while a transition from -1 to +1 corresponds to a local minimum. In this way, the algorithm robustly counts both maxima and minima while remaining insensitive to short flat regions in the signal.

Feature 7: Time difference between minimum and maximum [ms]

This feature enables one to gain information about the timing of the EGM signals, which was not previously addressed. Generally speaking, the time difference between the minimum and maximum of the signal will grow when the infarcted tissue becomes larger and/or the conduction velocity is lower [96, 88, 97]. It was calculated by using NumPy's *argmin()* and *argmax()* to identify the indices of the minimum and maximum voltages. These indices were then used on the time array to find the times of the maximum and minimum, which were then subtracted from each other.

Feature 8: Maximum upstroke [mV/ms]

The maximum upstroke or maximum positive derivative of the signal indicates how fast the signal changes. The maximum upstroke hence combines the timing difference feature (feature 7) and the maximum amplitude (feature 1), and can also help identify rapid fractionation, measured by the number of peaks (feature 6). Time derivatives of the EGM signal have already been used in the past to try and identify infarcted tissue [77]. Determining the maximum upstroke was done by first calculating the derivative of the signal with NumPy's *gradient* function, after which the maximum was extracted using NumPy's *.max()* function.

Feature 9: Maximum downstroke [mV/ms]

Similarly to the maximum upstroke, the maximum downstroke can provide additional information about the timing of the signal. Once again, using both the upstroke and downstroke can help identify any asymmetries present in the signal. The maximum downstroke was calculated by first determining the derivative with NumPy's *gradient* function, after which the minimum of this derivative was determined with NumPy's *min()* function.

Feature 10: Poincaré SD1/SD2 ratio []

A Poincaré plot is a visual representation of a time series that shows the relation between consecutive signal amplitudes by plotting x_{n+1} as a function of x_n for every n . An example of a Poincaré plot is shown in figure 3.4, which also shows two statistical measures that can be calculated for this type of plot. The first of these is *SD1* which defines the spread in the data perpendicular to the $y=x$ line and is defined as [136]:

$$SD1 = \frac{1}{\sqrt{2}} SD(x_{n+1} - x_n) \quad (3.5)$$

where SD indicates the standard deviation. The other statistical measure is *SD2*, which measures the spread in data along the $y = x$ line and is defined as [136]:

$$SD2 = \frac{1}{\sqrt{2}} SD(x_{n+1} + x_n) \quad (3.6)$$

with SD once again indicating the standard deviation.

Poincaré plots are already being used in clinical practice to analyse ECG signals [137], where the SD1 and SD2 measures mark the short-term heart rate variability (HRV) and long-term HRV, respectively [136]. Since EGMs also measure electrical signals from the heart, a Poincaré plot could also prove useful in this situation. However, the connection with HRV does not hold for EGM signals recorded in this thesis, because this requires multiple heartbeats to be recorded. SD1 and SD2 are still useful since they allow one to quantify the distribution between the consecutive data points. However, they both depend on the size of the signals and since the EGM signals near or in infarcted tissue have a lower amplitude, neither SD1 nor SD2 can be used directly. To address this problem, one can use the ratio SD1/SD2 such that the magnitude of the signal is divided away and hence has no impact on the final result, which has been done before for ECG signals [138].

Calculating the SD1/SD2 ratio is quite straightforward in Python, as one can construct the x_{n+1} array simply by using `signal[1:]`. After this, both SD1 and SD2 can be calculated using formulas 3.5 and 3.6 and using NumPy's `std` function to calculate the standard deviation. Once both are known, their ratio can then be trivially determined.

The SD1/SD2 ratio is expected to differ between healthy and infarcted tissue due to changes in electrogram morphology caused by disrupted electrical propagation. In healthy tissue, activation remains relatively smooth and uniform, resulting in more regular signal trajectories and a lower SD1/SD2 ratio. In contrast, infarcted tissue introduces conduction heterogeneity, fractionation, and sharper local variations in the electrogram, which increase the short-term variability represented by SD1 relative to SD2. Consequently, the SD1/SD2 ratio is generally expected to increase in infarcted regions.

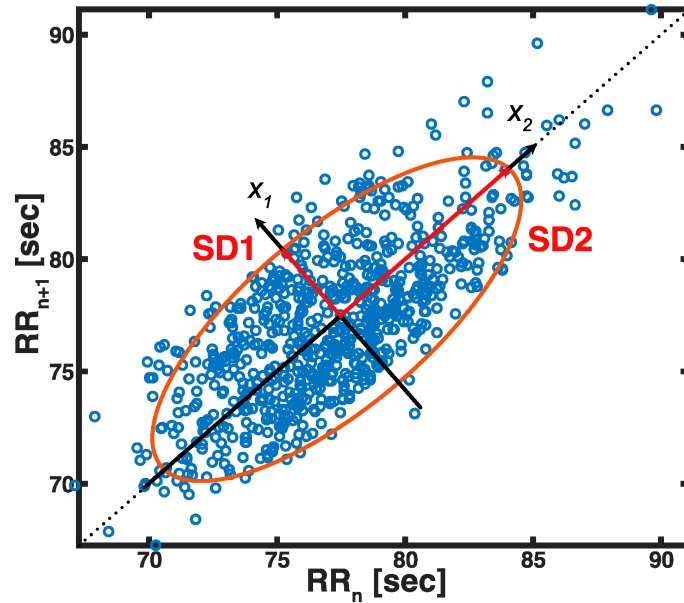


Figure 3.4: A Poincaré plot extracted from ECG data. The x-axis shows the x_n values, while the y-axis shows the x_{n+1} values for every n . The spread perpendicular to the $y = x$ line is indicated as SD1 and the spread along the $y = x$ line is indicated as SD2 [137].

Feature 11: Maximum Teager-Kaiser energy [mV^2]

The Teager-Kaiser energy is an energy measure that includes both the amplitude and the frequency of a signal and it can be interpreted as the energy a system needs to provide to create

a given signal. For continuous signals, it is defined as [139]:

$$\Psi_c[x(t)] = \left(\frac{dx(t)}{dt} \right)^2 - x(t) \left(\frac{d^2x(t)}{dt^2} \right) \quad (3.7)$$

where Ψ represents the Teager-Kaiser operator and $x(t)$ is the time series to be analysed. For a discrete time series, this becomes [139]:

$$\Psi[x[n]] = x[n]^2 - x[n-1]x[n+1] \quad (3.8)$$

Since the simulations outlined above only use one stimulus, frequency is not as important as it might be in other fields, but it can help detect fractionation of the EGM signal and a lowering of the voltage. Furthermore, the quadratic dependence on the current signal will be very sensitive to the amplitude of the signal, while the second term in equation 3.8 indicates how the signal changes via the previous and next signal. Since it combines the square amplitude of the signal and the change in the signal, it provides new information that is not yet captured in the other features, but it is heavily correlated with them. Also note that the numerical value of the energy measure will depend on the sampling rate [140], and hence care must be taken to use a uniform sampling rate across all simulations. The Teager-Kaiser energy operator has been used to classify ECG beats [141], and hence the extension to EGM signals is justified.

Because only one stimulus is given, one can intuitively expect that the TK energy will have one large spike, at least for signals recorded in healthy tissue. Hence, the maximum TK energy will be used as a feature in this analysis. To calculate the TK energy, one can transform equation 3.8 in Python as: `tk[1:-1] = signal[1:-1]**2 - signal[:-2]*signal[2:]` and then extract the maximum using NumPy's `nanmax()` function which automatically filters out all NaN values at the boundaries.

3.3.2 Principal component analysis

Principal component analysis, or PCA, is a data reduction technique that performs a coordinate transformation in order to reduce the number of basis vectors needed to retain as much of the information as possible. It is a non-parametric method that can reveal simple underlying structures in complex datasets, and its primary benefit is that it quantifies the importance of each dimension for describing the variability of the dataset [116]. Concretely, the algorithm finds the axis in the high-dimensional feature space which has the biggest variance and uses this as the first principal component (PC1). Then it repeats the process to find the axis which has the second most variance along it and uses this as the second principal component [116, 115].

A two-dimensional example is given in figure 3.5. Here, X and Y are two features that have a relatively strong correlation. PCA performs the coordinate transformation to the WZ coordinate system, where the majority of the variance lies along the Z axis. One can now use only the Z component instead of both X and Y and retain most of the variance.

PCA can be an especially promising technique when a lot of the features are mutually correlated [115, 116], as is the case here. PCA can then significantly reduce the number of components needed while retaining most of the information. Before performing PCA analysis, it is important that all the features are standardised. Here, this was done by means of z-score normalisation, which subtracts the mean for all data points and then divides by the standard deviation. The result is a dataset where all features have a mean of zero and a standard deviation of one. After this, the PCA analysis was performed using Scikit-learn's *PCA* method.

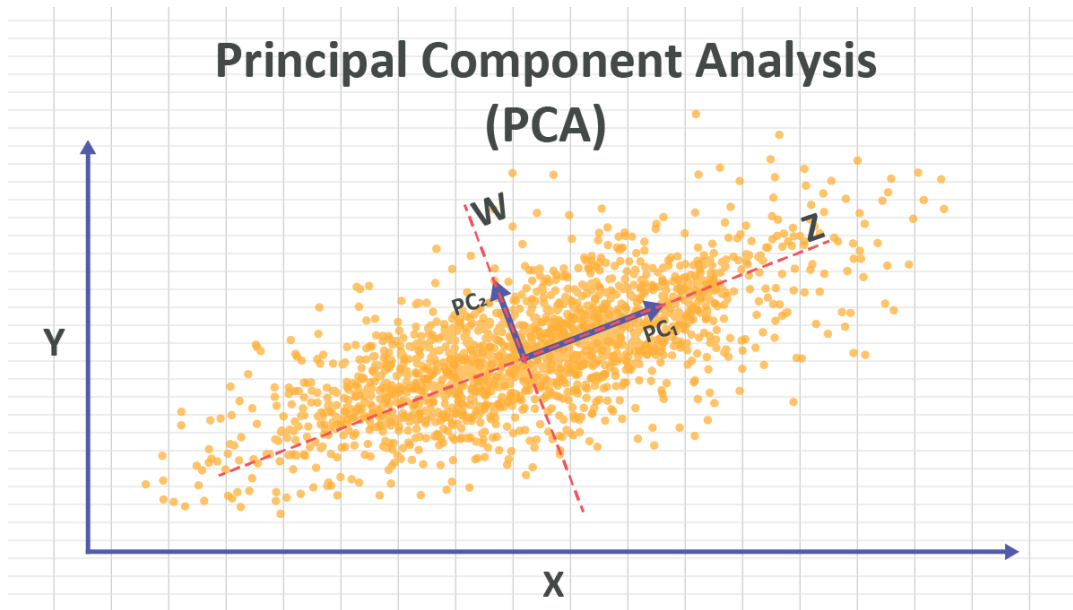


Figure 3.5: A two-dimensional example case for PCA analysis. X and Y are the original features, while W and Z are the rotated principal components. [142]

3.3.3 Supervised machine learning

Following the PCA analysis, supervised machine learning techniques can be used either to classify EGM electrodes into a set of categories or to make quantitative predictions about the underlying substrate. In this context, “supervised” means that the models are trained on examples for which the correct output is already known [143]. This condition is satisfied here because the simulations described above provide full control over the tissue substrate and model setup. Since PCA reduces the number of variables required to explain most of the variance in the data, using the resulting principal components as model inputs can improve computational efficiency and reduce redundancy and the risk of overfitting. Furthermore, where possible, the simulation datasets should be merged into a single combined dataset, so that the model receives as little explicit prior information about the substrate type as possible. The two subsections below will describe the working principles of two supervised machine learning techniques together with their implementation. The first method is a random forest regressor, which is a regression model that can make quantitative predictions [144]. The second method is logistic regression, which is commonly used for binary classification and can be extended to multiclass classification problems [145].

Random forest regressor

A random forest is an ensemble method that builds a large number of decision trees on bootstrap samples of the training data, meaning that new datasets are created by randomly sampling from the original training set with replacement. This means that some data points may appear multiple times while others may be left out. The algorithm then aggregates the predictions of all trees to make the final prediction [144]. Each tree is hence grown on a randomly drawn subset of the training data, and at each node split, only a random subset of features is considered. This double randomisation decorrelates the trees, so that averaging their predictions reduces variance without substantially increasing bias. For regression tasks, the final prediction is the mean output across all trees, which stabilises estimates relative to any individual tree. Because

no distributional assumptions are made about the relationship between inputs and response, random forests can capture non-linear dependencies that parametric models cannot.

The main advantages are robustness to overfitting, the ability to handle high-dimensional inputs without feature engineering, and a built-in estimate of generalisation error through out-of-bag samples, which are the training observations not selected in each bootstrap draw [144]. The principal drawbacks are computational cost, since training many deep trees on large datasets is memory-intensive, and limited interpretability: while feature importance is informative, the ensemble as a whole does not produce a simple, human-readable decision rule [143]. Performance is also sensitive to class imbalance in classification settings, though this is less relevant for regression.

The performance of the regression model can be indicated with several metrics. Given n observations with true values y_i and predicted values $\hat{y}_i$ and the mean of the true values $\bar{y}$, the coefficient of determination R^2 is defined as [143]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.9)$$

It measures the proportion of the variance in the target explained by the model. A value of 1 indicates perfect prediction, 0 indicates the model performs no better than predicting the mean, and a negative value means that the model is doing worse than just predicting the mean. The average magnitude of prediction errors is furthermore described by the mean absolute error defined as [143]:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.10)$$

The mean squared error penalises large errors more heavily than the MAE and is defined as [143]:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.11)$$

It is closely related to the root mean squared error [143]:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.12)$$

The random forest regressor was implemented using scikit-learn's *ensemble.RandomForestRegressor* with *n_estimators* set to 300, meaning 300 decision trees were used. Furthermore, the *random_state* was set to 42 and the input data were split into 75% training data and 25% test data using *train_test_split* from *sklearn.model_selection*.

Logistic regression

Logistic regression is a supervised classification method that models the probability of binary class membership as a function of one or more input variables. Rather than predicting a continuous output, it maps a linear combination of inputs through the logistic function, constraining predictions to the range $[0,1]$. A sample is assigned to the positive class when this probability exceeds a chosen threshold, typically 0.5. Model parameters are estimated by maximum likelihood, making no assumptions about the distribution of the input variables. Its main advantages

are computational simplicity, interpretability, and well-calibrated probability outputs. Its principal limitation is linearity: the decision boundary is a hyperplane in feature space, so logistic regression cannot capture non-linear class separations without feature engineering [145].

Classification performance is reported using several complementary metrics. Accuracy is the fraction of correctly classified samples, but it becomes misleading when class sizes are unequal. It is defined as [143]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.13)$$

where TP indicates the number of true positives, i.e. the number of cases that belong to class 1 and were classified by the model as class 1. TN is the number of true negatives, i.e. the number of class 0 cases that were classified as such. FP are the number of false positives, or hence the number of cases that were predicted to belong to class 1 while actually belonging to class 0. FNs are the false negatives, which represent the predicted class 0 cases that actually belonged to class 1. Balanced accuracy corrects for unequal class sizes by averaging sensitivity across both classes, giving equal weight to each regardless of how frequently it appears in the data [143]:

$$\text{Balanced accuracy} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (3.14)$$

Precision is the fraction of positive predictions that are truly positive, while recall (sensitivity) is the fraction of true positives that are correctly identified [143]:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.15)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.16)$$

The F1 score is the harmonic mean of precision and recall, summarising the trade-off between the two into a single value [143]:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.17)$$

Finally, the area under the receiver operating characteristic curve (ROC AUC) measures the model's ability to discriminate between classes across all possible classification thresholds: a value of 1 indicates perfect discrimination, while 0.5 corresponds to chance performance. In settings with class imbalance, balanced accuracy and ROC AUC are the most informative primary metrics, as standard accuracy and F1 can be dominated by the majority class [143].

Logistic regression is applied here not because perfect classification is expected, but because its linearity makes it a natural first test of whether the principal components encode discriminative information. Strong performance would confirm that the low-dimensional feature representation captures class-relevant structure; failure would motivate more flexible non-linear approaches.

The logistic regression was implemented using scikit-learn's *linear_model.LogisticRegression* with *max_iter* set to 5000, *class_weight* to balanced and *random_seed* to 42. The data were preprocessed in a *Pipeline* from *scikit-learn.Pipeline* and *StandardScaler* from *scikit-learn.Preprocessing*. *train_test_split* from *sklearn.model_selection* was used to split the input data into 75% training data and 25% test data. Cross-validation was furthermore implemented through *StratifiedKFold* from *sklearn.model_selection* with *cv_folds* set to 5 and *scoring* set to *balanced_accuracy*. Cross-validation is useful because it provides a more reliable estimate of model performance: the data are split into five folds, the model is trained on four folds and validated on the remaining fold, and this process is repeated until each fold has served once as the validation set.

Chapter 4

Results

This chapter will go over all the results of the outlined simulations in detail. The first section will perform a qualitative analysis of selected EGM signals and will explain how they were formed. The second section will then outline the results of a quantitative analysis on the full dataset.

4.1 Qualitative analysis of EGM signals

In this section, a subset of EGM electrodes was chosen for every simulation type and their properties and origin are discussed. The effect of changing different parameters on the signals is also discussed.

4.1.1 Dataset 1: Modified conduction simulations

Five selected EGM signals and the location of the corresponding electrodes are shown in the top left panel of Figures 4.2 and 4.1. The first thing to note is that the rightmost electrode (red) shows a much higher maximum than the leftmost electrode (green), even though these are both located in healthy tissue. However, this is due to the symmetry of the simulation, as all of the shown electrodes are located in the middle of the tissue. Since the scar will block any wavefront from passing through, the wavefront will warp around the scar at the top and bottom of the mesh, as can be seen in the top left panel of Figure 4.3. The wavefront then moves back towards the centre of the mesh from both the top and the bottom at the same time, after which the two wavefronts will collide right in the centre on the right side of the scar. It is exactly this collision that causes the spike in the top left panel of Figure 4.2, which is hence an artefact of the symmetry in the simulation setup.

Secondly, these signals show zero recorded voltage between about 20 ms and 60 ms, and a voltage peak just before and after this window. The height of the peaks at 20 ms is inversely proportional to the distance between the edge of the scar and the position of the electrode, i.e. an electrode that is located deeper into the infarcted tissue records a (much) lower maximum (and minimum) voltage. For the peak at 60 ms, only the electrode located in healthy tissue records a peak. Looking at the simulation frames, it can be seen that the wavefront reaches the scar at a certain time, after which it only propagates at the top and bottom of the simulated tissue. Both of these wavefronts then move back towards the middle once they have passed the scar to then meet back up at the halfway point. Since the plotted signals stem from electrodes in the middle row in the vertical direction, they will show a signal when the wavefront is near

this vertical row. Hence, the zero voltage part corresponds to the part of the simulation where the excitation wave is travelling around the scar, and the two peaks correspond to the wavefront arriving at the left and right sides of the scar, respectively. Furthermore, the leftmost electrode (green) shows a slight local minimum just after the main peak, even though it is located in healthy tissue.

A selection of 5 EGM signals from a slow conduction simulation is shown in the top middle panel of figure 4.2, with the corresponding electrodes shown in the top middle panel of figure 4.1. As in the previous simulation type, the high amplitude peak for the rightmost electrode (red) is again an artefact of the symmetry in the simulation, since a part of the wavefront still travels around the infarcted tissue, as is visible in figure 4.3. Unlike the previous simulation, there are now recorded signals between 20 ms and 50 ms, as the wave can now also travel through the infarcted tissue. These signals are lower in amplitude than the signals recorded in healthy tissue, but the amplitude does not decrease with depth in the infarcted tissue as it did before. Once again, the leftmost electrode (green) shows a shallow second minimum, even though it is located in healthy tissue.

All of the signals recorded in the infarcted tissue show a fractionation or break-up of the healthy EGM shape as well as a reduced amplitude. Both the fractionation and amplitude are similar for all electrodes in the infarcted tissue. Furthermore, the signals from the two rightmost electrodes that are located fully in the slow conduction zone (blue and orange) show their largest signal later than the electrode located on the right of the slow conduction zone. Looking at the top middle panel of figure 4.3, it is clear that this is caused by conduction velocity in the slow conduction zone, since the wavefront is quicker to travel around the infarcted tissue to the red electrode than through it to the blue and orange electrode.

4.1.2 Dataset 2: Conduction gradient simulations

Five selected EGM signals from the conduction gradient simulation are shown in the top right panel of figure 4.2 with the corresponding electrodes visualised in figure 4.1. Once again, the high peaks recorded by the two rightmost electrodes are an artefact of the symmetry of the simulation, as can be seen in the top right panel of figure 4.3. The amplitude of the electrodes decreases as the local conduction (i.e. the conduction at the node where the electrode is placed) decreases, and then increases again when the local conduction increases. The fractionation of the signals also seems to follow this pattern, where a lower local conduction velocity causes a more fractionated signal with the maximum and minimum amplitudes lying further apart. Overall, the signals show a wider range of morphology and a higher level of complexity than before due to a more continuous change in conduction velocity compared to dataset 1.

4.1.3 Dataset 3: Applying stimuli at an angle

Five EGM signals from the variable stimulus simulation are shown in the bottom left panel of figure 4.2, with the corresponding electrodes visualised in figure 4.1. Contrary to the previous simulations, the rightmost electrode now does not show an extremely high maximum because the two wavefronts going around the infarcted tissue do not meet in the middle, but more towards the bottom right of the mesh, as can be seen in the bottom left panel of figure 4.3. The two electrodes in healthy tissue (the leftmost and the rightmost) now both show a similar morphology, although the red electrode has a slightly lower maximum and a slightly higher minimum. The other electrodes are located completely inside the slow conduction zone, and all show a similar fractionated signal with the maximum and the minimum being further apart. As one moves deeper into the slow conduction zone (i.e. more to the right), the recorded signals

become more and more delayed. The two rightmost electrodes inside the slow conduction zone (blue and orange) record the majority of their signal well after the electrode on the right side of the slow conduction zone (red), and both reach their maximum and minimum at a similar time. The bottom left panel of figure 4.3 shows that this is caused by the wavefront travelling around the slow conduction zone faster than through it.

4.1.4 Dataset 4: Adding randomness to the boundaries of affected tissue

Five selected signals from the probabilistic boundary simulations are shown in the bottom middle panel of figure 4.2, with the corresponding electrodes shown in figure 4.1. The red peak once again shows a high maximum amplitude because both wavefronts travelling around the infarcted tissue meet here, as can be seen in the bottom middle panel of figure 4.3. However, unlike the previous simulation types, the signal of the red electrode now looks slightly different, as it is now located inside the boundary of the infarcted tissue. As for the previous simulations, the electrodes located inside the slow conduction zone recorded highly fractionated signals with an increased time difference between their minimum and maximum. However, unlike the previous simulation types, the morphology of the different signals now changes and is not exactly the same. This can be attributed to a more complex propagation of the wavefront, as it breaks up slightly when hitting the random boundary, as can be seen in the bottom middle panel of figure 4.3. Once again, the wavefront travels around the slow conduction zone quicker than it travels through it, and the orange and blue signals reach their maximum amplitude well after the red signal.

4.1.5 Dataset 5: Generating random affected tissue

Five selected EGM signals for the probabilistic infarct simulation are shown in the bottom right panel of figure 4.2, and the corresponding electrodes are shown in figure 4.1. The bottom right panel of figure 4.3 shows that the wavefront now experiences significant breakup once it reaches the infarcted tissue, and the propagation now happens non-homogeneously. The recorded signals reflect this, as every one is different and no clear pattern can be seen. All electrodes located inside the infarcted area show multiple positive and negative peaks, and all of them are heavily fractionated and very complex. It should also be noted that the time order of the signals is the same as the order of the EGM electrodes as seen from the wavefront. The bottom right panel of figure 4.3 also shows that the propagation through the infarcted tissue is, in this case, quicker than the propagation around it.

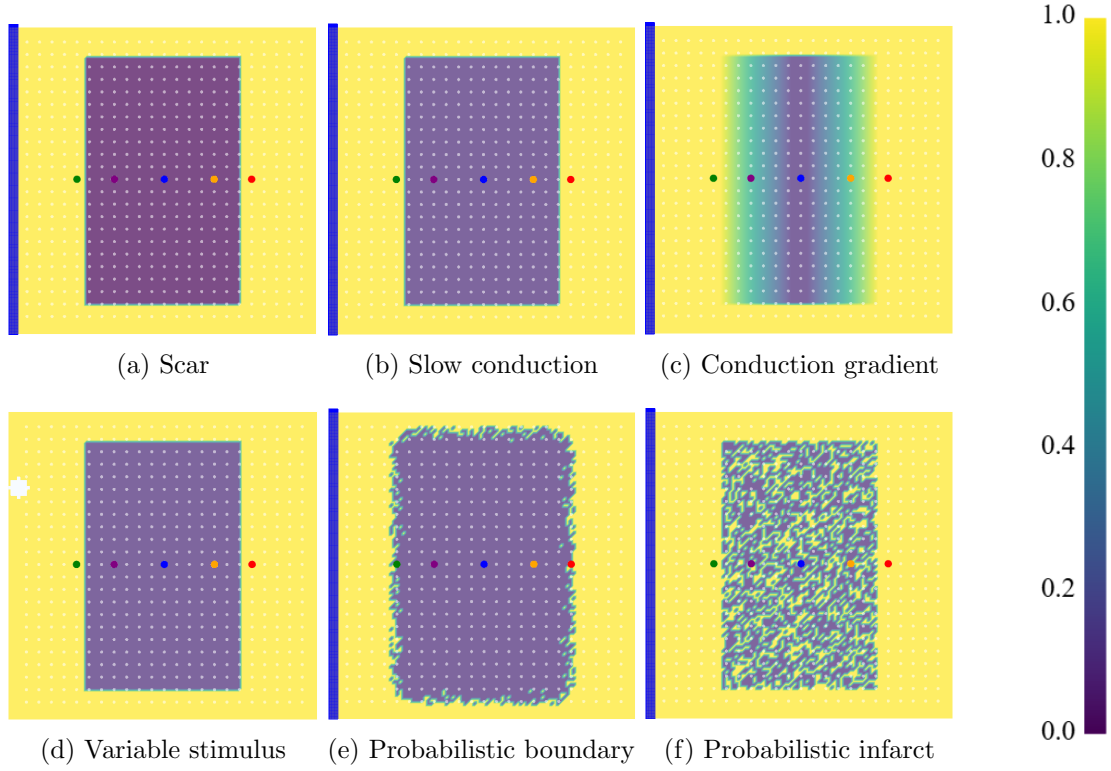


Figure 4.1: The electrodes that recorded the signals shown in figure 4.2. The stimulus is shown in blue or white on the left, and the used electrodes are shown with their corresponding colours. All infarcted areas have a height of 24 mm and a width of 15 mm. The conduction is 0.10 in all simulation types except for the scar simulation type, where the conduction is 0.00. The probabilistic infarct has a coverage of 60%.

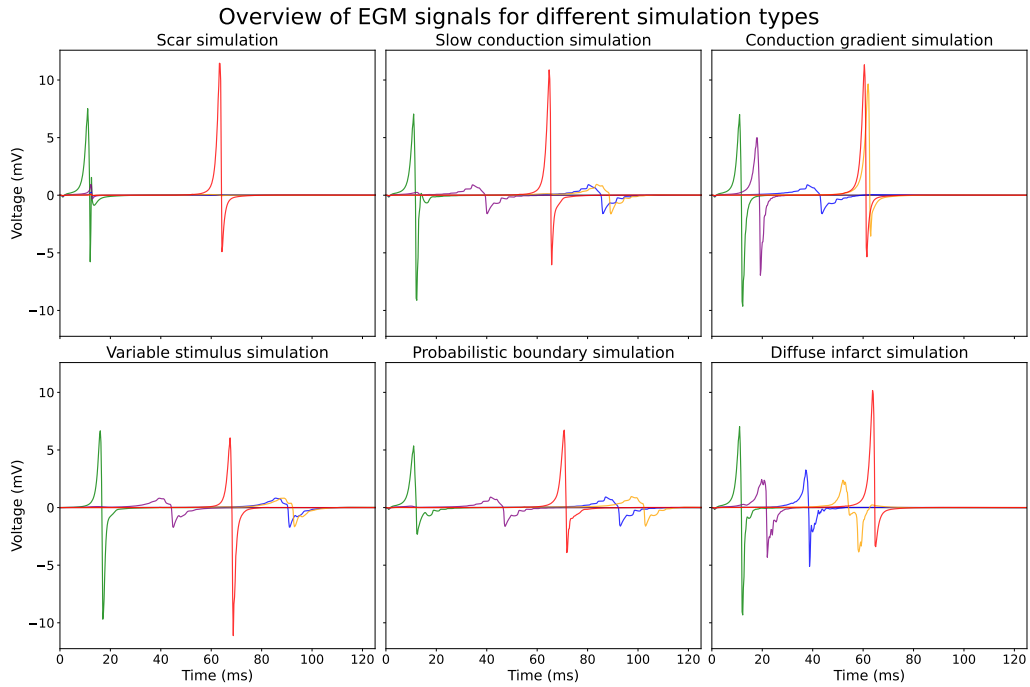


Figure 4.2: EGM signals extracted from different simulation types. The colours of the signals correspond to the coloured electrodes in figure 4.1.

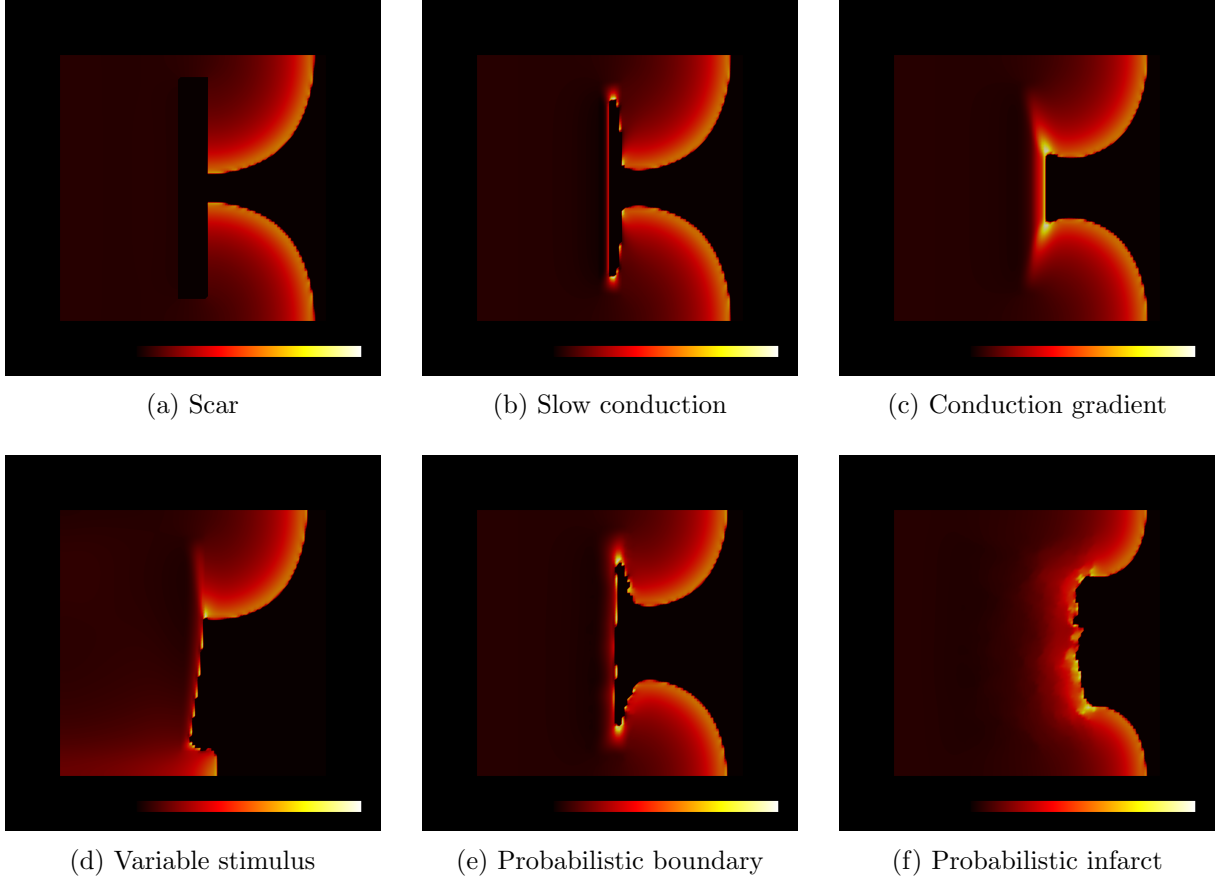


Figure 4.3: One simulation frame for each of the simulation types. The colourmap shows the value of the extracellular voltage, where yellow indicates a high extracellular voltage, and black a low one. Also note that all these frames display the same simulation time, but in different simulation types.

4.1.6 Effect of changes in the parameters on the recorded EGM signals

This subsection will discuss the effect of changes in the simulation parameters on the recorded EGM signals. The first section will discuss the effect of increasing the width of the infarcted tissue on recorded EGM signals at electrodes inside, on the boundary, and outside of the tissue. The second paragraph will then outline the same results for reducing the conductivity of the slow conduction zone. Lastly, an overview of signals recorded inside the infarcted tissue for changes in all simulation parameters will be given.

Effect of changing the width on EGM signals

The effect of changing the width of the scar in the scar simulations can be seen in figure 4.4. This figure shows the recorded signal of three different EGM electrodes for several widths of the simulated scar. The electrode in the left column is only located inside the scar when the width reaches around 40 ($= 12$ mm), and is hence located on the left of the infarcted tissue for most of the simulations. The timing of the recorded peak does not change as the width is increased, and also the shape remains the same for widths smaller than 30 ($= 9$ mm). However, as the scar creeps closer towards the electrode, the signal starts to deform, even when it is not yet located inside the infarcted tissue. Once it is fully located inside the scar, the amplitude decreases strongly, but only one peak is present.

The electrode in the second column is located inside the scar for all the simulated widths. The top row shows the recorded signal for a width of 2 ($= 0.6$ mm), where two distinct peaks are clearly visible. These peaks remain visible for somewhat larger widths, but eventually disappear completely as the widths keep increasing. For a width of 46 ($= 13.8$ mm), the signal is almost completely flat and almost no extracellular voltage is detected. At this point, the electrode is located 6.9 mm away from the nearest healthy tissue. The timing of the first peak does not change as the width of the scar increases. The second peak is only visible in the top two rows, but the timing also does not change here.

The right column shows an electrode that is located on the right of the scar tissue for most of the simulations and only enters the scar tissue for a width of around 40 ($= 12$ mm). As for the other electrodes, the timing of the peak does not change for an increasing width, but the amplitude strongly decreases once the electrode is located inside the scar tissue. Furthermore, the asymmetry of the signal increases from the top to the bottom, i.e. as the width of the scar increases. In particular, the maximum amplitude remains the same as long as the electrode is located outside of the scar tissue, but the minimum amplitude becomes smaller as the width increases.

Looking at all the signals in the top row, i.e. for a width of 2 ($= 0.6$ mm), one can see that the peak for the left electrode occurs first and the peak for the right electrode occurs last. The two peaks that are present in the graph in the middle seem to coincide with the peak of the left electrode and the peak of the right electrode, respectively. This pattern corresponding to the timing of the different signals continues for larger widths, although the second peak in the central figure is not visible for widths larger than 14 ($= 4.2$ mm).

Effect of changing the conduction on EGM signals

Figure 4.5 shows EGM signals in and around a slow conduction zone for varying conductivities. The signal in the column on the left, recorded on the left side of the slow conduction zone, shows no changes when the conduction is varied. The peak in the signals always occurs at the same time and both the minimum and the maximum remain constant. Contrary to figure 4.4, there are no deformations in the signal, even though the electrode is relatively close to the infarcted tissue.

The middle electrode is located inside the slow conduction zone. For large conduction velocities, the amplitude looks like that of a healthy EGM. Reducing the conduction velocity first causes a decrease in amplitude, but for conductivities above 0.5, the signal looks qualitatively similar to the signal recorded in healthy tissue. Further reducing the conduction velocity causes fractionated signals that also become wider, i.e. the time difference between the minimum and the maximum increases. Furthermore, the maximum amplitude moves to the right and is reached much later. Hence, for lower conduction velocities, the timing of the strongest part of the signal shifts towards the timing of the rightmost electrode. For the smallest values of the conduction velocities, the EGM electrode registers zero voltage everywhere, except for an extremely small bump at around 20 ms.

The right column shows the recorded signals from the right electrode, i.e. the electrode located behind the slow conduction zone. Also here, the maximum amplitude is reached later for slower conductions, and it moves towards the left as the conduction velocity increases. Furthermore, the recorded signals for slow conductions show a minimum that is much smaller than the maximum, and the signal is hence very asymmetric. The minimum gradually becomes larger as the conduction velocity increases, and at a conduction of around 0.5, the asymmetry is almost completely gone.

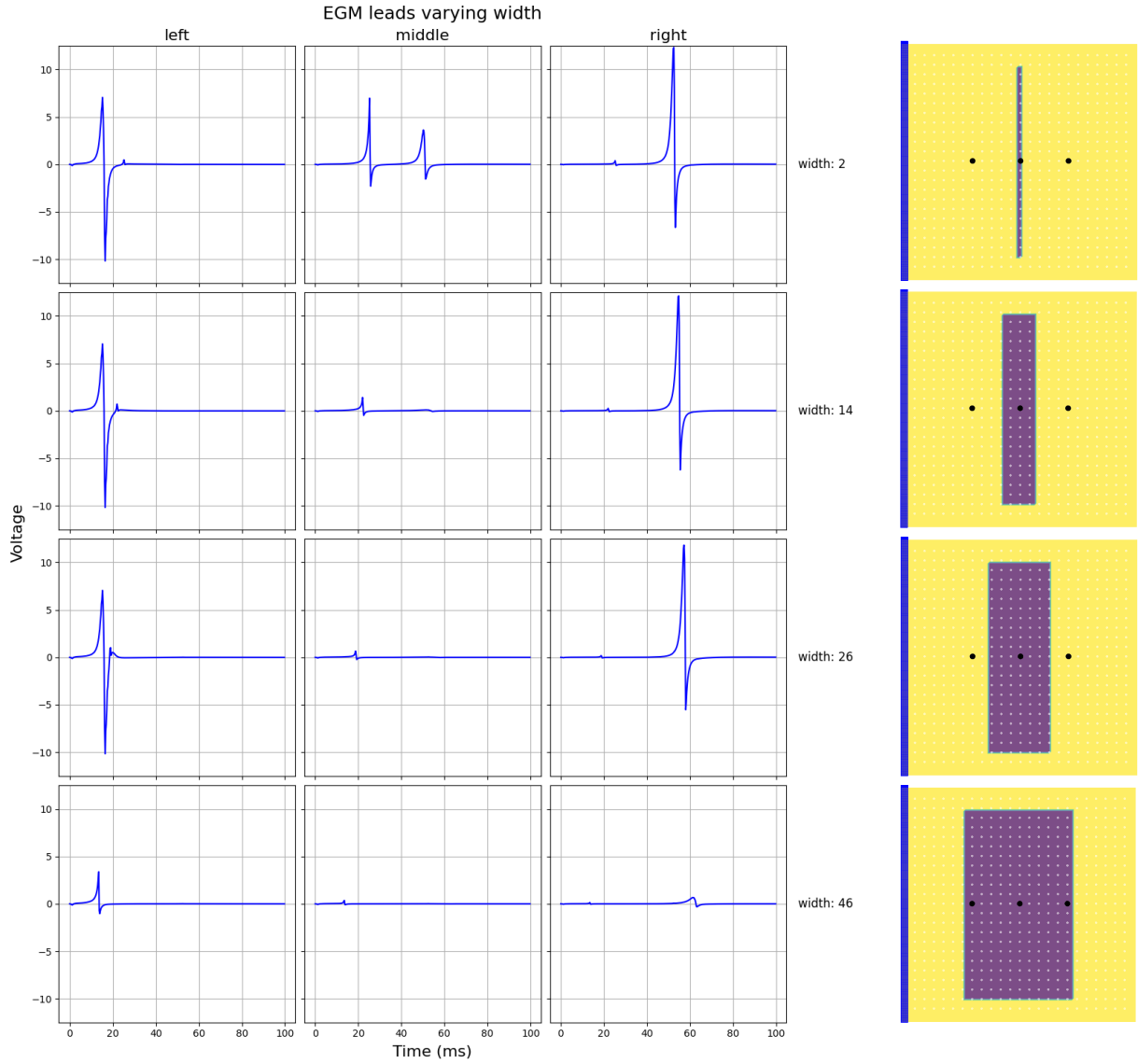


Figure 4.4: The figure on the left shows recorded EGM signals for three different electrodes for varying widths in the scar simulations. The three electrodes at which the signals were recorded are shown in black on the figures on the right, which show setups for the increasing widths. Note that the widths were varied symmetrically from the centre. The applied stimulus is shown in blue on the left.

Comparison of changes in all simulation parameters on the EGM signals

Figure 4.6 displays an overview of recorded EGM signals from the electrode in the centre of the mesh, which was always located inside the infarcted tissue. The figure shows signals at varying conduction velocities and simulation parameters for all the different types of simulations. Note that the conduction velocity here represents the conduction through the infarcted tissue, and not the averaged conduction at the electrode recording the signals. This figure allows one to investigate the influence of the different parameters on the recorded signals and to compare the signals between different simulation types.

A first global result of this figure is that a decreasing conduction velocity causes a reduction in the amplitude of the signal and an increase in complexity, with lower conduction velocities often showing clear fractionation or double potentials. A decrease in the conduction velocity also

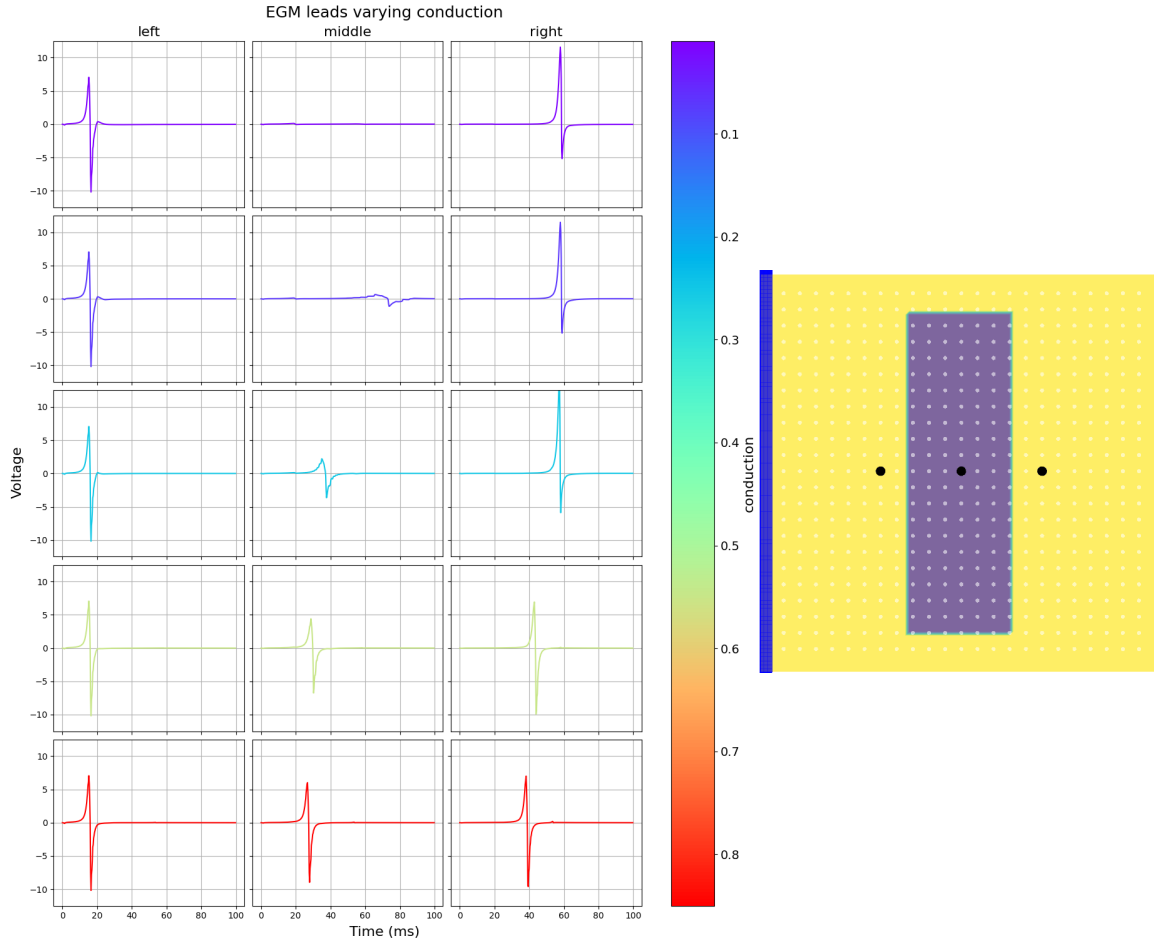


Figure 4.5: The figure on the left shows recorded EGM signals at three different electrodes for varying conductivities in the slow conduction simulations. The colorbar shows the value of the conduction velocity, where purple and blue colours indicate a slow conduction, and red colours point to a high conduction. Note that this conduction represents the conduction velocity of the infarcted tissue, and not the averaged conduction that was introduced in section 3.2.6. The three electrodes that recorded these signals are shown in black in the figure on the right. The applied stimulus is shown in blue on the left.

causes a delay in the timing of the signals, as the wavefront will reach the electrode at a later simulation time. Another effect of a lowering of the conduction velocity is an increase in the time difference between the minimum and the maximum of the recorded EGM signal. On the other hand, a decrease in conduction velocity does not always cause asymmetry in the signal. Overall, signals recorded in high conduction velocity (0.6 and 1.0) areas display a sharp maximum and minimum with a large amplitude. On the contrary, low conduction signals generally show a decrease in amplitude and an increase in complexity.

The signals recorded in the substrate simulation (red) show that increasing the width of the infarcted tissue causes a decrease in both the positive and negative amplitude of the signal. For low conduction velocities (0.1), a larger width also results in a broadening of the signal. At zero conduction velocity, there is only a slight bump for a width of 26 (= 7.8 mm) and 50 (= 15 mm). However, the simulation with a width of 2 (= 0.6 mm) shows a clear double potential with two distinct positive peaks but only shallow negative peaks. This double potential is also present for a conduction velocity of 0.1 and a width of 2 (= 0.6 mm), but the positive peaks are slightly lower, and the negative peaks are even less present. At a conduction velocity of 0.6,

there is a little broadening of the signal and a slight decrease in amplitude.

The gradient simulation (blue) with a gradient width of 3 (= 0.9 mm) and a conduction velocity of 0.0 also shows a double potential, albeit with a very small positive amplitude and no negative peaks. At this conduction velocity, the simulations of gradient width 11 (= 3.3 mm) show an even smaller amplitude double potential, although it is hard to see. With a gradient width of 23 (= 6.9 mm) and a conduction velocity of 0.0, the signal becomes completely flat. At a conduction velocity of 0.1, there is more fractionation and an overall more complex signal than in the substrate simulations. The time difference between the maximum and minimum is generally speaking also larger than in the substrate simulations. However, increasing the width of the infarcted tissue now results in smoother signals with a decreasing time difference between the minimum and the maximum. The recorded signals at a conduction velocity of 0.6 all show a reduced amplitude and an increased width compared to the signals recorded for a conduction velocity of 1.0. Also here, increasing the width results in a lower amplitude, although the effect is less pronounced. Overall, the changes are less abrupt than before.

The displayed signals from the variable stimulus simulations (green) all have a width of 2 (= 0.6 mm), and the height of the applied stimulus is varied. Note that $y = 50$ is exactly halfway up the mesh, and hence a stimulus at this height would produce a highly symmetric simulation, similar to those of the substrate type. For a conduction velocity of 0.0, the signal shows clear double potentials with a slightly stronger minimum amplitude than in the other simulation types. The timing between the two peaks also decreases as the stimulus height is increased. A conduction velocity of 0.1 shows EGM signals that are very similar to those recorded in tissue of zero conductivity. There are slight changes in both positive and negative amplitude, but the timing of both peaks is almost identical. The signals recorded in a conduction velocity of 0.6 again show a little broadening and a slightly reduced amplitude compared to the signals recorded in a conduction velocity of 1.0, which are recorded in healthy tissue. Also for these signals, the wavefront takes longer to reach them as the stimulus height is increased, hence they shift to the right for increasing values of the stimulus height.

A very slight double potential is visible for the probabilistic boundary (yellow) simulation with a width of 2 (= 0.6 mm) and conduction velocity of 0.0. The positive amplitudes are very small and no negative peaks are visible. The signals at widths 26 (= 7.8 mm) and 50 (= 15 mm) for a conduction velocity of 0.0 are completely flat. At a conduction velocity of 0.1, the simulation with a width of 2 (= 0.6 mm) shows an increased complexity compared to the previous simulation type and is comparable to the gradient simulation with a gradient width of 3 (= 0.9 mm) and the same conduction velocity. As the width is increased, the complexity of the signal and the time difference between the minimum and the maximum decrease again. Once again, the signals recorded at 0.6 conduction velocity show a small broadening and a slight decrease in amplitude, but the effect of increasing the width is rather limited.

The signals recorded in the diffuse infarct simulations (purple) show a strong dependence on the infarction probability. For a probability of 10%, the diffuse infarct signals show only a small effect of the infarcted tissue. For the 0.0 conduction velocity, there is a clear asymmetry, and for the 0.1 conduction velocity, there is a slight decrease in amplitude. The signals at 0.6 and 1.0 conduction velocity look the same. For a probability of 50%, the effect is more pronounced. The zero conduction simulation shows only a very small bump and the 0.1 conduction simulation shows a clear asymmetry with an increased time difference between the minimum and the maximum. Furthermore, the 0.6 conduction velocity simulation shows a slight decrease in amplitude. For the 90% probability simulation, the zero conduction velocity signal is completely flat. At a conduction velocity of 0.1, the signal is highly deformed and fractionated. At 0.6 conduction velocity, the amplitude has now markedly decreased and the signal is also broader than before.

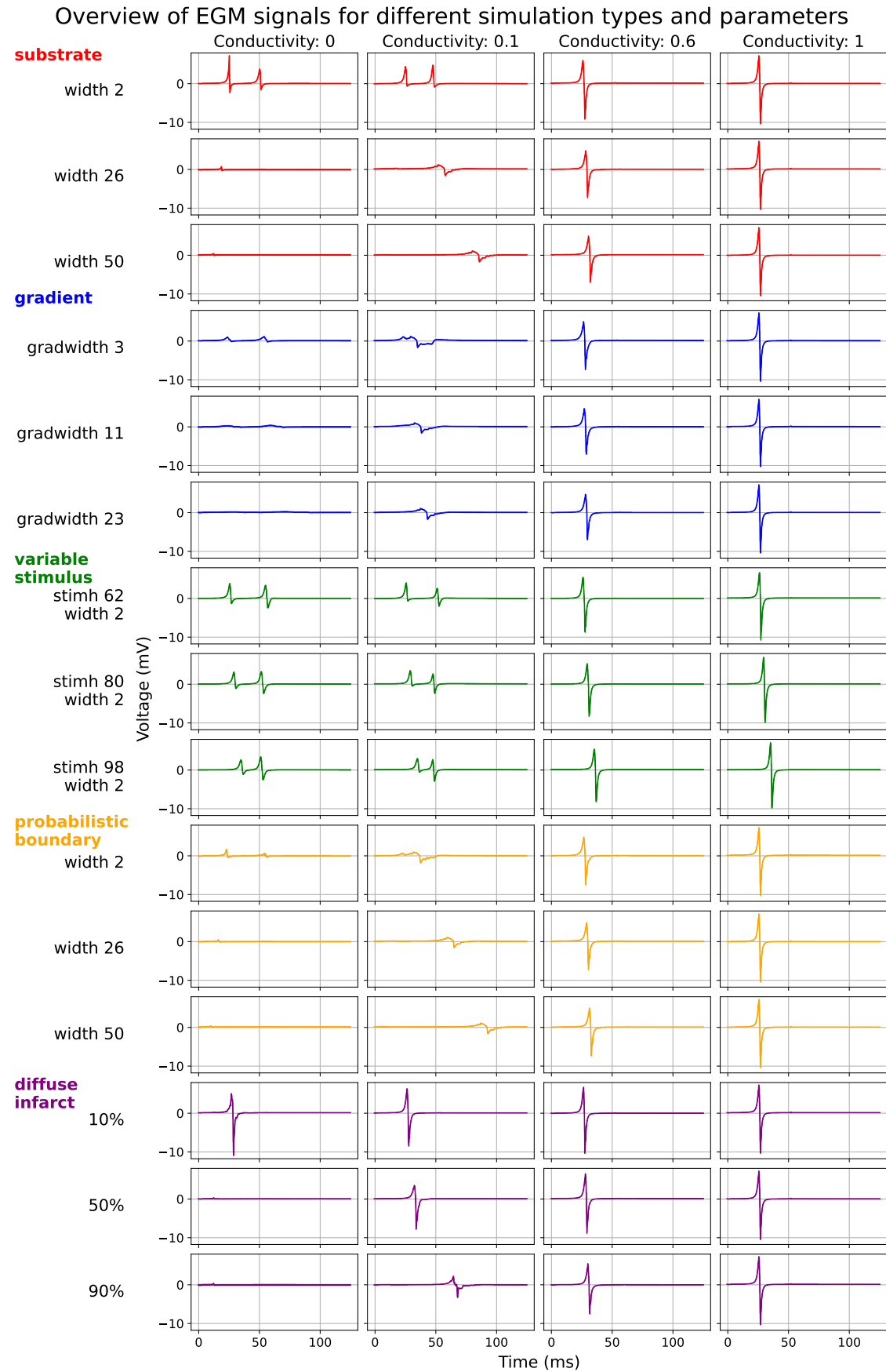


Figure 4.6: Recorded EGM signals at the electrode in the centre of the mesh for different simulation types with varying parameters. The used electrode is always located inside the infarcted tissue. The conductivities represent the actual conduction value inside the infarcted tissue and are indicated at the top, while the other varied parameters are shown on the left of each row. Every colour represents a different simulation type.

4.2 Quantitative analysis of EGM signals

This section describes the results of all analyses from the recorded EGM signals for every simulation. The first subsection will focus on the features and discuss how well they are able to distinguish averaged conductions. The second section will outline the results of the PCA analysis and describe which features explain the most variance. Finally, the third subsection will present the results of the supervised machine learning approaches.

4.2.1 EGM feature extraction

The correlation between features for every simulation type is shown in the correlation matrices in Appendix A.2.1, which demonstrate that the features are highly correlated. An overview of the correlation between each feature and the averaged conduction across simulation types is shown in Figure 4.7. All the features plotted on the y-axis have been normalised, that is, they have been divided by the largest value of that feature for that simulation type, and hence all have a value between -1 and 1. Each column shows a different simulation type. The x-axis shows the averaged conduction, which reflects the local neighbourhood of the electrodes and hence allows one to show the trend between a given feature and the averaged conduction for a specific simulation type.

Globally, it can be seen that the features all follow a very tight trend. That is, the 1σ and 2σ areas lie very close to the median, and strong deviations are only seen in the 3σ area, which incorporates nearly all data points. Furthermore, the trends for a single feature do generally not vary a lot over the different simulation types.

The *signal_max* median differs only slightly across simulation types and shows a very tight relation. The median value of the maximum increases as the averaged conduction increases, however the slope is not that steep, so a single value of the *single_max* cannot single-handedly determine the averaged conduction. A similar discussion can be held for the *signal_min*. The slope of this feature is steeper than for *signal_max*, indicating that it can narrow down the averaged conduction to a smaller interval. *Signal_range* combines *signal_min* and *signal_max* into one feature and hence also shows a tight relationship across all simulation types. Also here, the slope is quite steep, and the *signal_range* can hence narrow down the averaged conduction quite a bit. The *absolute_area* also has a narrow spread. The slope is somewhat shallow, but there is still a clear rise in absolute area as the averaged conduction becomes larger. Both the *max_upstroke* and *max_downstroke* show narrow trends as a function of the averaged conduction. Furthermore, they seem to be each other's mirror image, with the maximum upstroke having a strong positive slope, and the maximum downstroke showing a strong negative slope. Both of these can hence provide a relatively narrow range of averaged conductivities. Both the *TK_max* and the *sd1/sd2* features globally show a narrow spread and a decently pronounced slope, which indicates that they are both useful features for determining the averaged conduction velocity.

On the other hand, the *signed_area* has a nearly flat median, with only a very slight downwards slope. However, the spread in the standard deviation areas has a more pronounced downwards slope, indicating that the signed area can still help narrow down the value of the averaged conduction, albeit a bit less than for the previous features. The case for *n_peaks* is even worse than for *signed_area*, as the median is now completely flat. There are some variations and a slightly downward trend in the areas of the standard deviations, so some quantitative prediction for the averaged conduction can still be made, but the range will be very broad. The *time_difference_min_max* also has a completely flat median, however there is a large spread in the 2σ and 3σ trends for low averaged conductivities, again allowing some quantitative prediction to be made.

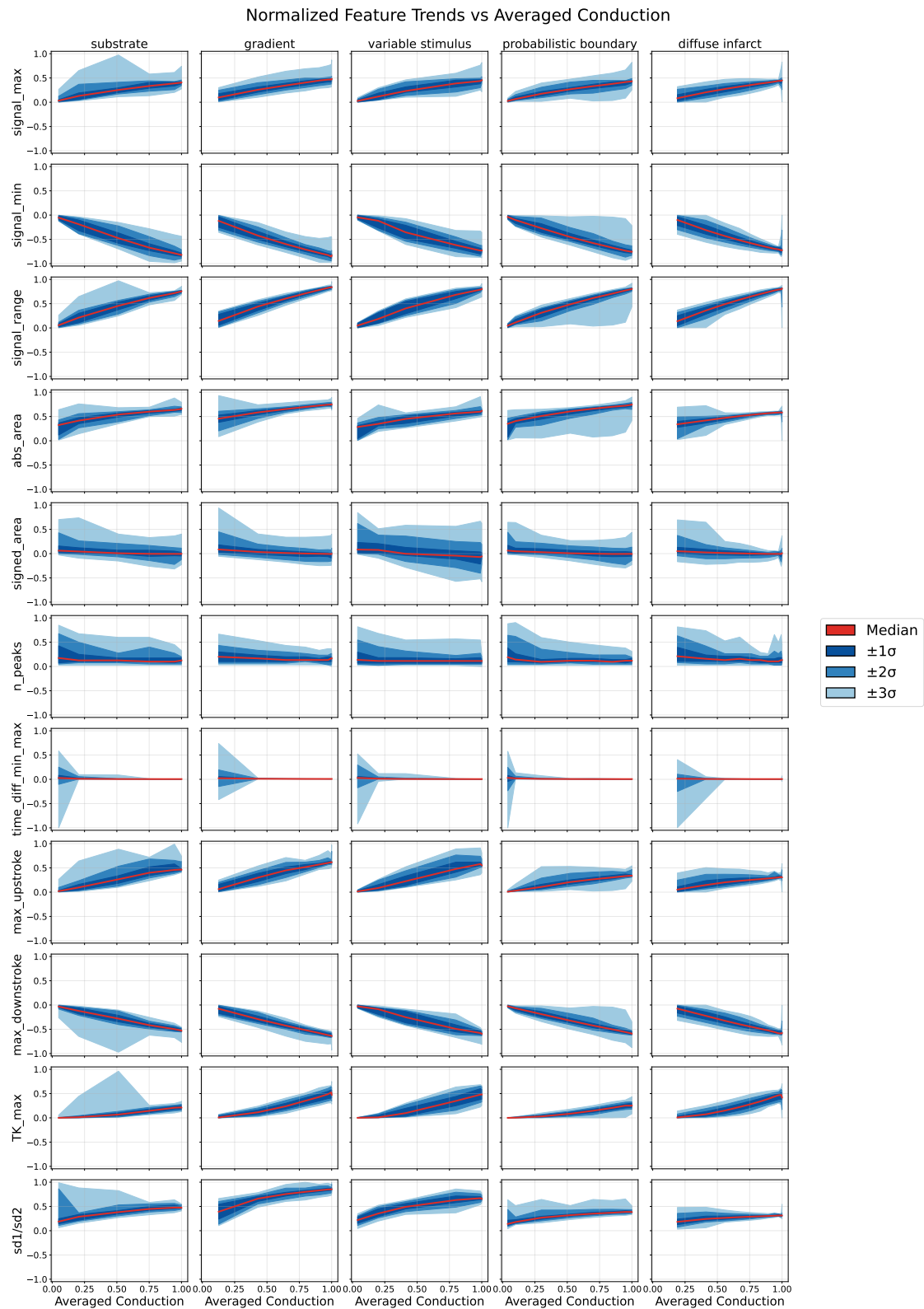


Figure 4.7: Overview figure of all features as a function of averaged conduction for all simulation types. Every column shows a different simulation type; their labels can be found at the top. The rows represent the features, with their labels visible on the left. All features have been normalised, i.e. divided by the largest value of that feature for that simulation type. All y axes hence run from -1 to +1. The x-axis always contains the averaged conduction and hence runs from 0 to 1. The red lines show the median trend, with the shades of blue showing the 1σ , 2σ and 3σ deviations. The lighter the shade of blue, the larger the deviation shown.

4.2.2 Principal component analysis (PCA)

This section is divided into two paragraphs. The first paragraph will focus on the principal components themselves, i.e. it will discuss the loadings and describe how much information each principal component contains. The second paragraph will then outline how these principal components are related to the averaged conduction and the distance towards the affected tissue.

The principal components

The results of the principal component analysis are summarised in Figure 4.8, the exact numerical values of the loadings per simulation type can be found in Appendix A.2.2. The right column of this figure shows the explained variance per principal component for every simulation type. This explained variance is relatively constant across different simulations, with the first principal component representing 65% to 70% of the variance. The second component then represents around 10% of the variance for all simulations, and components three and four sit just below this value at around 7% to 9% of explained variance for every type. The fifth principal component only contains a few per cent of the variance for all simulation types and hence does not contribute a lot of new information that is not yet contained in the other principal components. Hence, the first four principal components capture over 90% of the variance for all simulation types and significantly reduce the redundancy in the dataset. Additional components do not provide substantial further information.

The loadings of each principal component are shown in the left column of Figure 4.8, where the original features are shown on the x-axis and the loadings per simulation type per principal component are shown on the y-axis. The loadings of the first principal component are very stable across different simulation types. That is, they vary only slightly, if at all, across different simulation types. 8 out of the 11 features contribute with a similar weight to the first principal component: *signal_max*, *signal_min*, *signal_range*, *absolute_area*, *maximum_upstroke*, *maximum_downstroke*, *maximum_TK_energy*, and *sd1/sd2*. The loadings of all of these features towards PC1 are nearly constant across simulation types, except for *sd1/sd2*, which shows a slight spread. Furthermore, the absolute loadings of all these features are very similar, ranging closely between 0.35 and 0.4. *Signed_area*, *n_peaks*, and *time_difference_min_max* contribute the least to the first principal component and also have a small spread between different simulation types. Note that these last three features are also the features with the flattest slope from figure 4.7.

The loadings of the second principal component are in some sense exactly opposite to those of the first principal component. First of all, the features that provided a high loading for the first principal component now have a loading that hovers around zero. Conversely, *Signed_area*, *n_peaks*, and *time_difference_min_max* contributed very little to PC1, but now make up the majority of the second principal component, albeit with completely different loadings for different simulation types. Overall, the spread over simulation types within one loading is much higher than for the first principal component.

The third and fourth principal components are similar to the second principal component, in the sense that the same features, *Signed_area*, *n_peaks*, and *time_difference_min_max*, contribute very strongly to these principal components, while the other loadings tend to lie around zero. Also for these principal components, it is the case that the exact loading for each strongly contributing feature differs widely per simulation type, and no clear pattern can be established. The loadings of the features that have a low loading, on the other hand, show only a small spread between simulation types. As stated before, the fifth principal component only explains a few per cent of the variance, and it will hence not be discussed in detail.

The results of the principal component analysis can be summarised as follows. The first principal component accounts for approximately 65–70% of the total variance and has roughly uniform contributions from most features, with the exception of *signed_area*, *n_peaks*, and *time_difference_min_max*. These features instead contribute more strongly to the second through fourth principal components, with the exact loadings varying across simulation types. Together, PC2–PC4 explain an additional 25–30% of the total variance.

Principal components as a function of averaged conduction and distance

The previous paragraph discussed the significance and contents of all principal components, but did not yet illustrate how the principal components correlate with the averaged conduction and the distance to the affected tissue. That will be the topic of this paragraph.

A correlation plot of PC1 against PC2 for all the different simulation types is shown in figure 4.9, once with the averaged conduction as a colourmap and once with the distance to affected tissue as a colourmap. Only PC1 and PC2 are shown here since, as shown in the previous paragraphs, PC2–PC4 are similar in contents and explained variance. The distribution of the principal components looks quite similar for all simulation types: a wide range of PC1 values with a relatively narrow band of PC2 values, except for very low conductivities, which have a broader range of PC2 values. Note that before the PCA analysis, all features had been regularised with z-score normalisation, and all PCA loadings lie between -1 and +1. Since there are 11 original features, PC values below -11 or above +11 hence indicate a strong deviation from the mean for one or multiple original features for that EGM signal and hence represent statistical outliers.

The colourmaps show a strong trend in averaged conduction, especially when considering PC1. Note that the figure shows the averaged conduction and hence reflects the conduction properties in the local neighbourhood of each electrode. PC2 shows much less of this behaviour, as all averaged conduction seem to have similar values of PC2, with only small averaged conduction showing a wider range of values. For both *probabilistic boundary* and *diffuse infarct*, there is a much wider spread on conduction velocities around 1 for PC1.

The bottom panel of this figure shows the same principal components, but with *distance to affected tissue* as a colourmap. These figures show that a lot of electrodes are either in or very near the affected tissue. Furthermore, the values of the principal components for electrodes that are further away coincide with the values of the electrodes that are located inside the affected tissue, as can be seen by the strong overlap of different coloured dots. This indicates that making a prediction of the distance to affected tissue will be difficult based on the principal components alone.

The cross-correlations of PC1–PC4 for the full dataset are shown in Figure 4.10. That is, the datasets of all simulation types have been merged into one, and this one was used to construct the figure. These figures show that the range of the values of the principal components is very similar across different simulations, as the amount of overlap between different conduction velocities has remained constant. However, there is now a larger overlap between clusters. The strong trends for the averaged conduction in the figures containing PC1 is still visible, but this is not the case for the other principal components. PC2–PC4 show much more overlap for different conduction velocities than PC1. The distance to the affected tissue colourmap also did not change dramatically compared to the previous figure, with a lot of the points still having a zero or close to zero distance. Again, both the zero and non-zero distance points show a large overlap. Correlations between all principal components for each simulation type can be found in Appendix A.2.3

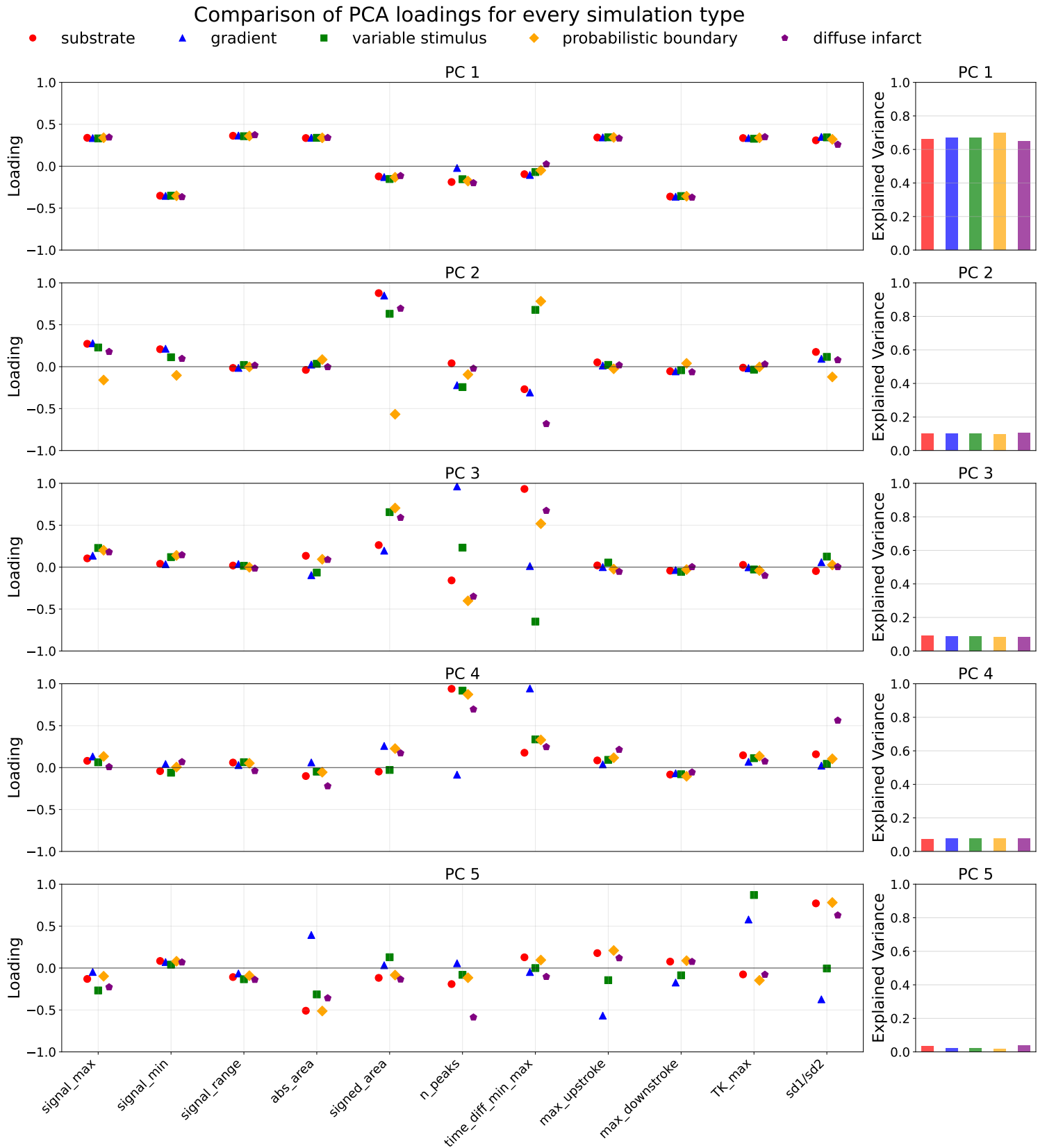


Figure 4.8: Results of the PCA analysis. Every row shows a separate PCA component. The left column shows the loadings of the PCA components in terms of the original features, with each simulation type having a separate colour and symbol. The right column shows the explained variance for every principal component, again with a different colour for each simulation type.

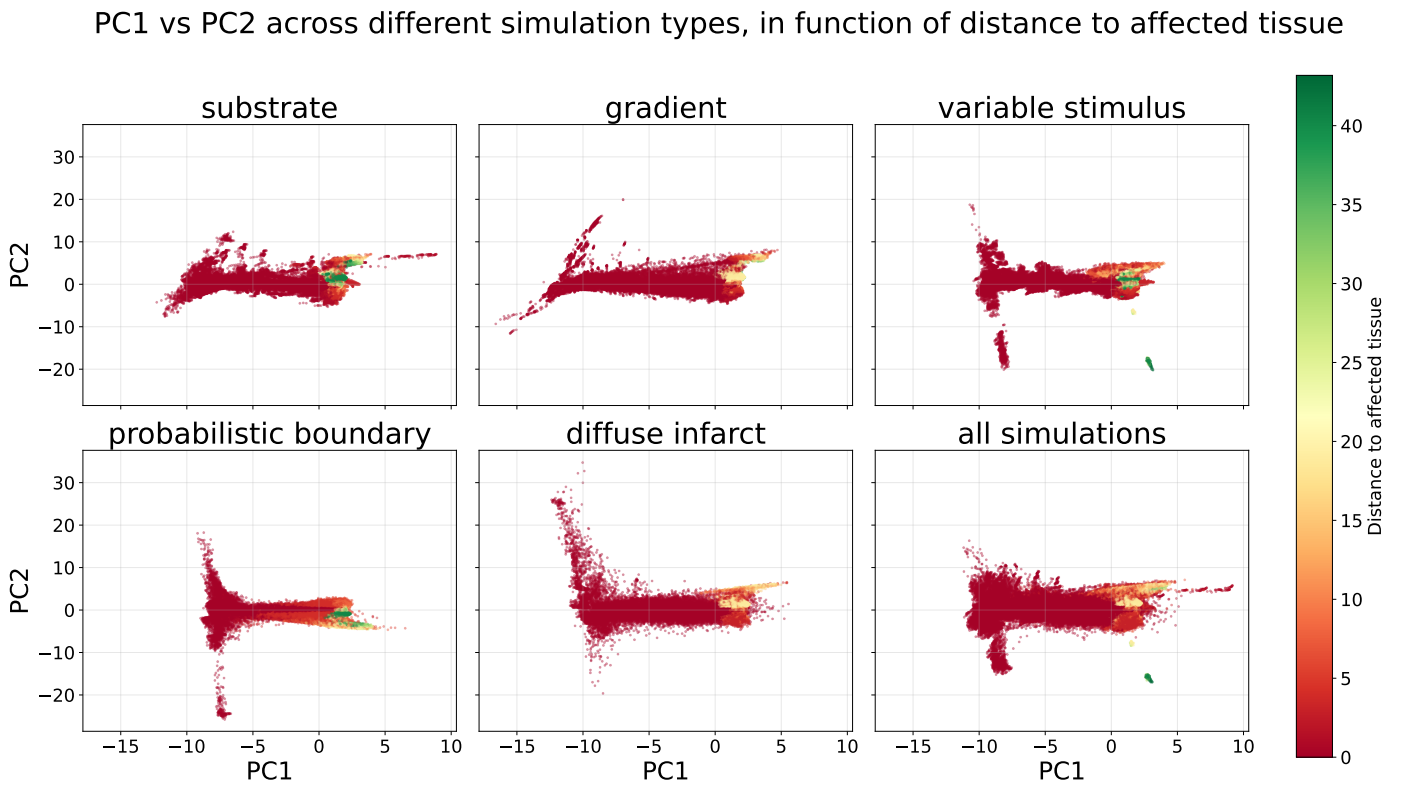
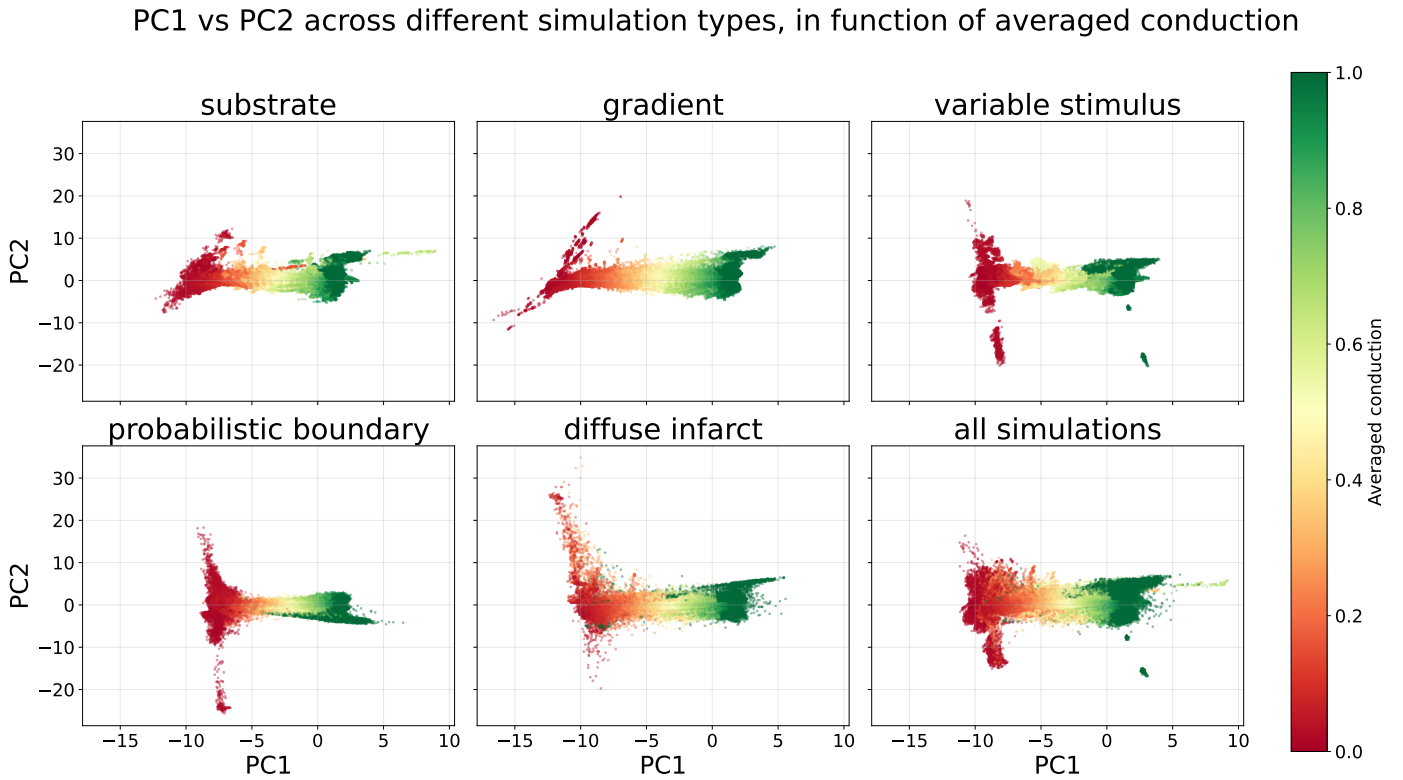
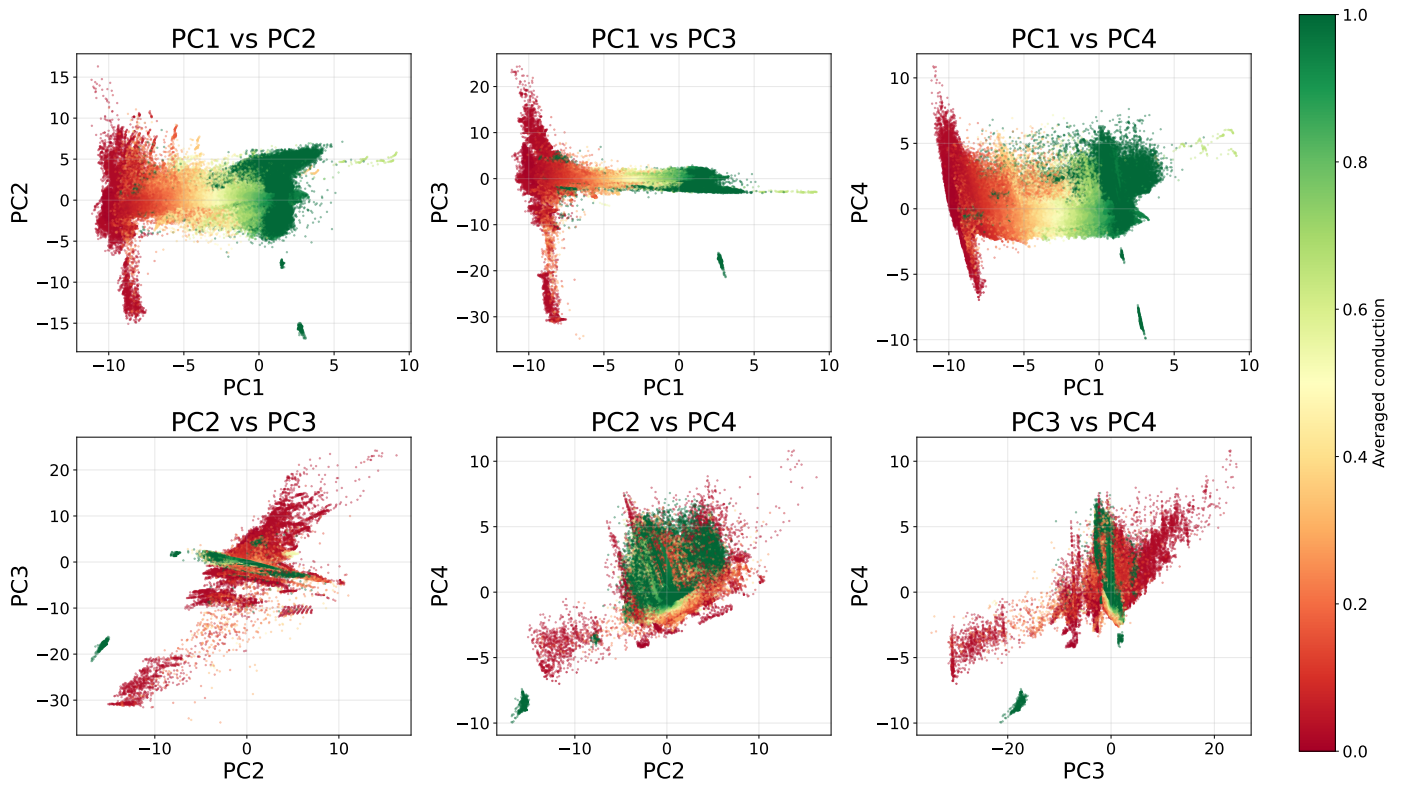


Figure 4.9: The correlation between PC1 and PC2 as a function of the averaged conduction (top panel) and distance to affected tissue (bottom panel) for all simulation types.

Correlation between PCA components for all_simulations in function of averaged conduction



Correlation between PCA components for all_simulations in function of distance to affected tissue

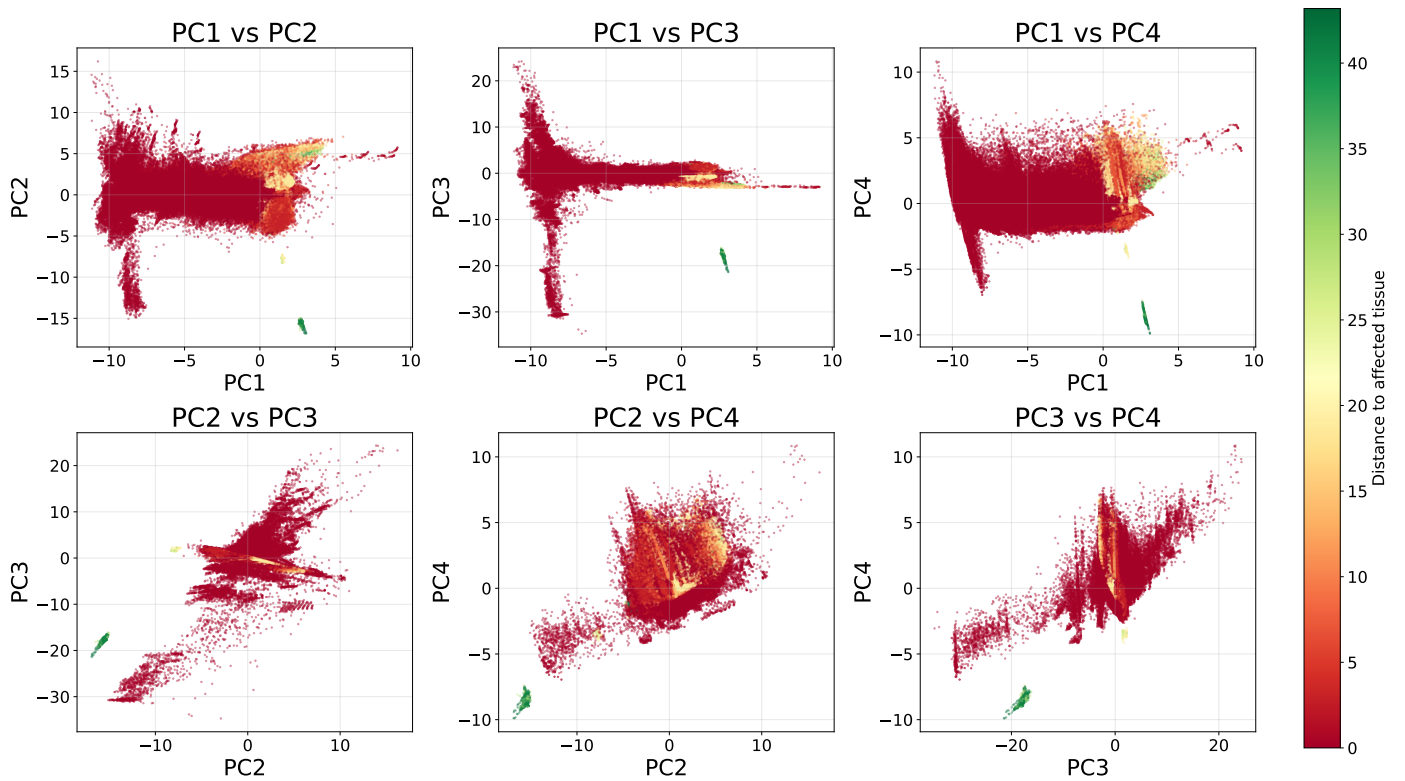


Figure 4.10: The correlation between all principal components as a function of the averaged conduction (top panel) and distance to affected tissue (bottom panel) for the combined dataset of all simulations.

4.2.3 Supervised machine learning

This section will discuss the performance of two supervised machine learning techniques for predicting the distance to affected tissue and the averaged conduction. The datasets of all 5 simulation types were merged into one, such that the model has as little information about the underlying substrate as possible. Only the first four principal components were used as the inputs, since these were the only ones contributing a significant amount to the explained variance. Section 4.2.2 showed that the principal components showed a clear trend as a function of averaged conduction and only a small clustering as a function of distance to affected tissue. The first subsection below will hence try to predict the averaged conduction by use of a random forest regressor, while the second subsection will try to classify zero and non-zero distances by use of a logistic regression.

Random forest regressor for averaged conduction prediction

Table 4.1: Regression performance of the random forest regressor for averaged conduction using the full dataset.

Descriptor	Value
R^2	0.981
Mean absolute error (MAE)	0.012
Root mean squared error (RMSE)	0.038
Mean squared error (MSE)	0.00147

The random forest regressor was used to predict the averaged conduction from the four PCA components on the full dataset, and the results are displayed in Table 4.1. The model achieved a high coefficient of determination, with an R^2 score of 0.981, indicating that most of the variance in the target variable was captured by the regression model. The mean absolute error was 0.012, while the root mean squared error was 0.038, corresponding to a mean squared error of 0.00147. Since the conduction takes values between 0 and 1, the mean absolute error indicates that the random forest regressor is accurate up to 1.2%.

The relatively small difference between the mean absolute error and root mean squared error indicates that most prediction errors remained limited in magnitude, although the larger RMSE shows that some higher-error predictions were present. Overall, the results show that the random forest regressor was able to reproduce the averaged conduction values with high accuracy across the full dataset.

Logistic regression for distance classification

The classification performance of the logistic regression for classifying zero distance and finite distance to affected tissue is summarised in Table 4.2, and the confusion matrix is shown in Figure 4.11. All 1,83 million electrodes were used, and the zero-distance category made up 28.6% of all cases. The overall accuracy is 92.6%, but when accounting for class imbalance, this drops slightly to 89.3% balanced accuracy. These results indicate that the model can distinguish between the two classes quite well, despite the imbalance between them. The ROC AUC value of 0.946 shows that the model separates zero-distance and finite-distance very well, while the F1 score of 94.9% demonstrates that the classifier has a strong balance between correctly identifying finite-distance samples and avoiding false zero-distance predictions.

The per-class breakdown shows that the precision of both classes lies close together, indicating that the classifier does not strongly favour one class over the other when assigning positive

Table 4.2: Classification performance of the logistic regression classifier for distance to affected tissue using the full dataset.

Classification performance	
Accuracy	92.6%
Balanced accuracy	89.3%
ROC AUC	0.946
F1 score – class 1	94.9%
Per-class breakdown	
Precision – zero distance	91.4%
Recall – zero distance	81.6%
Precision – non-zero distance	92.9%
Recall – non-zero distance	96.9%
5-fold cross-validation	
balanced accuracy mean	89.3%
balanced accuracy standard deviation	± 0.001

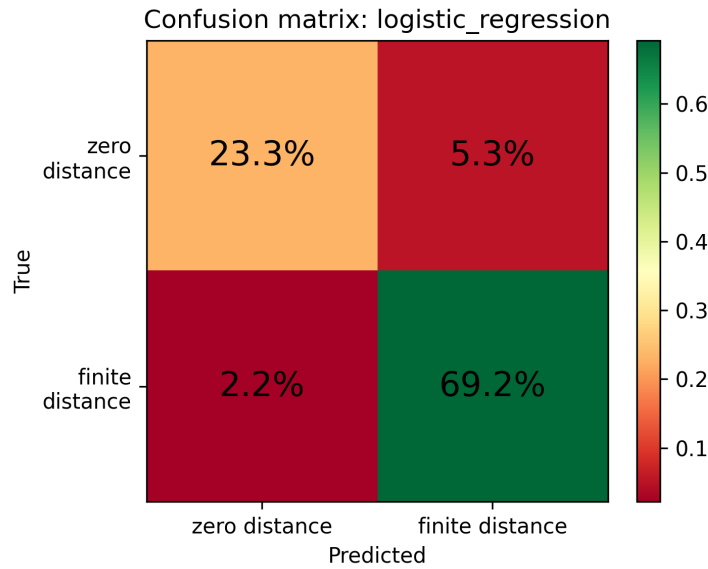


Figure 4.11: Confusion matrix for the logistic regression. Diagonal elements show correct classifications, while off-diagonal elements show wrong classifications.

predictions. However, the large difference between the recalls of both classes indicates an asymmetry in the classifier performance, with non-zero-distance samples being detected more reliably than zero-distance samples. This asymmetry is also visible in the confusion matrix, where 5.3% of zero-distance cases are misclassified as finite-distance cases. Finite-distance electrodes are misclassified less frequently, with only 2.2% being classified as zero-distance cases.

The 5-fold cross-validation score was obtained by splitting the dataset into five subsets, training the model on four subsets and evaluating it on the remaining subset in turn. The mean balanced accuracy of 89.3% with a standard deviation of 0.001 indicates that the classifier performance was stable across different validation splits, which indicates that the model is not strongly overfitting on the training set.

Chapter 5

Discussion

5.1 Overview of current literature on EGM signal analysis

EGM morphology reflects a combination of activation sequence, conduction properties, tissue geometry, and electrode configuration, and should therefore be interpreted as a multidimensional signal rather than reduced to isolated markers [100, 74, 42]. Voltage amplitude and fractionation are widely used clinically, but their physiological meaning is ambiguous: bipolar voltage depends on measurement factors beyond tissue properties [37, 78, 88, 96], fractionation is inconsistently defined [66], and both can occur in tissue without structural remodelling [99, 114].

Computational models have been instrumental in establishing the mechanistic links between tissue properties and EGM morphology, since fibrosis density, conductivity, and structural organisation can be varied in isolation [111, 94, 91]. These studies show that complex fractionated electrograms can arise from slow or heterogeneous propagation rather than from scar alone [110, 108], and that partially overlapping morphologies can arise from distinct tissue configurations [95, 102]. Translating simulation-derived relationships to clinical recordings remains difficult, however, as electrode configuration, wavefront direction, tissue contact, and noise all substantially influence extracted features [45, 43, 33, 34, 38, 84].

Multidimensional feature analysis and machine learning have been proposed to address the limitations of isolated EGM markers. Models combining amplitude, timing, asymmetry, and fractionation features have shown promise for characterising fibrotic substrate and estimating conduction properties. Amplitude- and area-based features in particular have been related to conduction velocity [87, 98, 119, 74, 117, 118, 76], often outperforming voltage thresholds alone. However, feature redundancy is common and rarely explicitly examined, and dimensionality reduction approaches such as PCA can help identify the dominant modes of variation in the EGM feature space [115, 116]. Furthermore, the relationship between qualitative electrogram morphology and simulation configurations remains incompletely understood [32], which is the gap the present analysis addresses.

5.2 Main findings

5.2.1 Qualitative analysis of EGM signals

The simulated datasets reproduced several characteristic EGM morphologies associated with conduction slowing and scar tissue, including reduced voltage, fractionation, and prolonged

activation. Electrodes inside the non-conductive region showed reduced voltage, though for narrow scars (0.6 mm), amplitudes as high as 6 mV were still recorded, depending on electrode proximity to the scar boundary. This is consistent with the clinical observations of Frontera et al., who questioned the universality of fixed voltage thresholds, and shows that even in a controlled simulation setting, voltage amplitude is strongly modulated by substrate geometry and electrode position, confirming that fixed voltage thresholds for identifying fibrotic tissue are not universally transferable [99, 37, 40].

The narrowest scars reproduced double potential electrograms [85, 96], originating from the wavefront reaching one scar boundary, propagating around the non-conductive obstacle, and subsequently activating the opposite boundary. The first peak had a higher amplitude than the second, suggesting an asymmetry between proximal and distal activation, consistent with wavefront splitting around the obstacle. Because this morphology requires the electrode to be sufficiently close to both boundaries, it only appeared for narrower scars (figure 4.4). Critically, narrow slow-conduction zones (0.6 mm) with sufficiently low conduction velocity (0.1) produced morphologically similar double potentials, distinguished by two comparable positive amplitudes and a very small negative component, demonstrating that double potentials are not specific to non-conductive scar tissue [99, 88]. This non-specificity extends to fractionation more broadly: although multiple peaks were associated with slow and heterogeneous propagation, peak count alone did not reliably indicate conduction severity [66, 76, 42], and at conduction velocities above 0.5, the signals showed only reduced amplitude and slight peak broadening relative to healthy tissue.

Unlike scar tissue, slow-conduction regions still permitted propagation through the affected area. Many signals exhibited clear fractionation, with multiple local minima and maxima reflecting delayed and non-uniform propagation caused by the abrupt change in conduction [91, 110, 66]. In addition, electrodes distal to the slow-conduction region activated before electrodes inside it, indicating that propagation through the affected region could become slower than propagation around it. These timing delays are directly analogous to the delayed activation signatures targeted in LAT mapping [119, 130].

Scar tissue also influenced electrodes located in neighbouring healthy tissue, which underscores that EGM interpretation cannot be restricted to purely local properties. As shown in figure 4.4, proximal electrodes showed a small deflection after the main activation, corresponding to the excitation wave reaching the scar boundary, while distal electrodes were strongly asymmetric, with much larger positive than negative amplitudes. This asymmetry reflects a larger contribution from curved wavefronts approaching the distal electrode than from wavefronts moving away from it, as illustrated in figure 4.3. Infarcted signals were also asymmetric around $V = 0$, consistent with findings from other studies [89, 90, 91], supporting the inclusion of features such as signed area to capture this asymmetry. A similar non-local effect was observed in the slow-conduction simulations. Electrodes located distal to the affected region showed increased asymmetry compared with healthy EGMs, as well as delayed activation relative to electrodes inside the slow-conduction zone, suggesting that wavefront propagation around the affected region could outpace propagation through it. This demonstrates that the recorded EGM is shaped not only by the local tissue beneath the electrode but also by the surrounding conduction pattern. As a result, classification should not rely solely on local signal morphology, but should also account for neighbouring electrode responses and the broader conduction environment [88, 33, 99]. This observation motivates the use of averaged conduction as a more robust descriptor of the tissue region influencing each EGM.

Conduction-gradient simulations introduced more spatially variable morphologies than the homogeneous substrate simulations. EGMs displayed stronger fractionation and pronounced asym-

metry, reflecting the presence of spatially varying conduction properties. As in the homogeneous slow-conduction simulations, signal amplitude decreased as conduction velocity decreased. In addition, electrodes located closer to the distal side of the infarcted tissue produced high positive amplitudes and relatively small negative amplitudes. This is consistent with the collision of wavefronts travelling around the affected region and suggests a larger contribution from incoming than outgoing activation. These observations indicate that signal asymmetry is sensitive to wavefront directionality and may therefore contain information about local conduction heterogeneity. This conclusion has also been made in previous literature, however these focused on bipolar electrograms [45, 38, 78]. Changing the stimulus angle mainly affected relative activation timing while preserving the overall signal morphology. In contrast, probabilistic boundaries and random infarct geometries produced substantial signal distortion and heterogeneity, generating a much broader range of EGM morphologies more closely resembling the irregular spatial heterogeneity expected in pathological tissue [32, 95, 108]. The diffuse infarct simulations in particular produced highly diverse EGMs across neighbouring electrodes, indicating that structural disorder strongly increases local electrophysiological variability. The probabilistic simulations therefore establish whether the systematic relationships between conduction abnormalities and signal structure observed in simpler geometries persist under realistic structural disorder.

5.2.2 Quantitative analysis of EGM signals

The extracted EGM features showed strong quantitative relationships with averaged conduction, as shown in figure 4.7, which supports their use as a physically meaningful descriptor. Most amplitude- and slope-related features displayed clear monotonic trends, indicating that they constrain the estimated local conduction state. In particular, signal maximum, signal minimum, and signal range were informative across datasets, showing robust correlations with relatively narrow spreads. This confirms their established relevance as indicators of conduction abnormalities and tissue pathology [87, 98, 119]. On the other hand, features related to asymmetry, fractionation, and timing showed weaker global trends, but remained informative in lower conduction regimes. Surprisingly, the number of peaks showed little direct correlation with conduction velocity, apart from an increasing spread at lower averaged conduction. Although fractionation was clearly associated with slow conduction in the qualitative analysis, these results indicate that peak count alone does not provide a reliable quantitative estimate of conduction severity. It should therefore be interpreted as a complementary rather than primary marker [66, 76, 42].

The relatively narrow feature distributions observed here reflect the controlled nature of the simulations, in which measurement noise and uncontrolled structural variability are absent. In clinical recordings, both sources of variability would broaden these distributions and reduce the apparent separability between conduction regimes [33, 84, 34]. The persistence of systematic trends across increasingly complex datasets confirms that the extracted features capture reproducible electrophysiological structure, though the broader spreads in probabilistic and diffuse infarct simulations indicate that infarct topology contributes substantially to residual signal variability.

Principal component analysis further revealed that EGM variability is governed by a low-dimensional structure, consistent with the substantial interdependence between the extracted features, as shown in Appendix A.2.1. The first principal component presented in figure 4.8 was highly robust across simulation types, with dominant loadings on signal strength-related features, which were also those with the most clearly defined trends in figure 4.7, indicating that PC1 captures overall conduction integrity as a single dominant axis. The secondary principal components (PC2-PC4) mainly captured asymmetry, fractionation, and timing, which were the

features with the least defined trends in figure 4.7. Their loadings varied more substantially between simulation types, indicating that these components are more sensitive to the underlying substrate topology. The largest deviations occurred in the probabilistic boundary and diffuse infarct simulations, suggesting that the secondary components are particularly affected by the detailed spatial arrangement of the infarcted tissue.

The secondary components each explain similar fractions of the total variance and therefore span a weakly constrained low-dimensional subspace. When several principal components explain nearly equal amounts of variance, small perturbations in the dataset can rotate the PCA basis vectors or alter their ordering [116, 115]. Consequently, differences between datasets in the exact ordering of PC2-PC4 are expected to occur, and these components should be interpreted collectively as a shared representation of asymmetry, fractionation, timing dispersion, and conduction heterogeneity rather than as three fully independent physiological axes. That these secondary features capture conduction heterogeneity is consistent with their established role in characterising slow conduction zones [99, 114, 42], and their collective sensitivity to substrate topology suggests they may carry discriminative information precisely where the classification problem is hardest.

Projection into PCA space as a function of averaged conduction and distance to affected tissue, presented in figures 4.9 and 4.10, revealed clear electrophysiological organisation. PC1 formed an approximately continuous electrophysiological severity axis: low averaged conduction was positioned at strongly negative PC1 values, whereas healthier conduction states progressively shifted toward more positive PC1 values. PC1 therefore, reflects not just signal strength but the overall integrity of local electrical propagation. The continuous organisation of datapoints along this axis suggests that conduction abnormalities form gradual transitions rather than discrete categories, which supports regression-based prediction of conduction velocity from EGM features.

The PCA projections also suggest that distance to pathological tissue is partially, but not uniquely, encoded in the feature space. Electrodes further away from affected regions tended to occupy relatively restricted regions of PCA space, whereas electrodes close to or in pathological tissue were distributed across a broader range of PCA values. However, this relationship was less continuous than the relationship with averaged conduction, suggesting that distance influences the EGM features indirectly through its effect on local signal morphology and neighbouring activation patterns.

Surprisingly, the logistic regression classified electrodes as inside or outside affected tissue with 89.3% balanced accuracy, despite substantial overlap between the two classes in PCA space. The misclassified electrodes were likely located at tissue boundaries, where contributions from both healthy and pathological tissue produce only minor EGM modifications. The random forest regressor predicted averaged conduction with a mean absolute error of 1.2%, indicating that the four major principal components encode substantial information about the conduction environment surrounding the recording electrode. These results demonstrate that supervised machine learning applied to EGM-derived features can reduce the ambiguity between scar and functionally slow conducting tissue that isolated signal markers cannot resolve.

Different simulation types produced distinct geometric distributions in PCA space, but preserved a similar global organisation. Substrate, gradient, and variable stimulus simulations formed relatively smooth trajectories, whereas probabilistic boundary and diffuse infarct simulations generated broader and more fragmented point clouds. Despite these differences, pathological regions remained partially localised and showed systematic separation from healthier conduction states, although overlap persisted between neighbouring conduction regimes. The feature space

retained physically meaningful organisation across all simulation types, though overlap between conduction regimes increased with substrate complexity.

The use of averaged conduction is both a strength and a limitation of the present analysis. From a physical perspective, it is more appropriate than the local conduction at a single mesh node, because the simulated uEGM does not reflect only the tissue directly underneath the electrode. Instead, the recorded signal is influenced by a finite surrounding region, especially near sharp boundaries where healthy, slowly conducting, and non-conductive tissue can all contribute to the same electrogram. The systematic relationships observed between several EGM features and averaged conduction suggest that this quantity captures relevant electrophysiological information in the electrode neighbourhood. In this sense, averaged conduction provides a more realistic target for feature-based prediction than the exact local conduction, particularly given the finite spatial resolution of clinical mapping electrodes and the non-local nature of electrogram formation.

At the same time, averaging inevitably smooths the underlying conduction field and can obscure sharp transitions or narrow pathological structures. This is especially important at scar borders and in narrow slow-conduction regions, where an electrode located inside abnormal tissue may still receive substantial contributions from nearby healthier tissue. Consequently, averaged conduction should not be interpreted as a direct measurement of the true local tissue property at the electrode position, but rather as an effective descriptor of the conduction environment sampled by the electrogram. This has important implications for the interpretation of the machine learning results: successful prediction of averaged conduction would not necessarily imply that the model can resolve microscopic conduction patterns or identify exact tissue boundaries. Instead, it would indicate that EGM-derived features contain information about conduction deviations at the spatial scale imposed by the electrode sensitivity, the chosen weighting kernel, and the cutoff radius. The results should therefore be interpreted as predictions of an electrogram-weighted conduction substrate, rather than as pointwise reconstructions of the underlying cardiac mesh.

The chosen cutoff radius is therefore an additional modelling assumption that may influence the apparent relationship between EGM features and conduction. A smaller cutoff would preserve sharper substrate boundaries and make the target closer to the local conduction, but it could also underestimate the spatial region contributing to the recorded electrogram. Conversely, a larger cutoff would better reflect broad spatial averaging by the electrode, but would further smooth narrow scars or slow-conduction channels and could artificially improve feature-conduction correlations by removing fine-scale heterogeneity. The cutoff used here should therefore be interpreted as defining the spatial scale at which conduction is predicted, rather than as a uniquely determined physiological parameter. Future work could test the sensitivity of the feature trends, PCA structure, and machine learning performance to this cutoff choice.

5.3 Limitations

First, the representation of pathological tissue is simplified. This thesis modelled conduction abnormalities either by reducing the diffusion coefficient or by introducing non-conductive regions representing scar tissue. Although this makes the signals easier to interpret, these substrates remain idealised abstractions of infarcted myocardium that cannot fully reproduce the complex behaviour of real pathological tissue. In particular, the used representations of pathological tissue introduced sharp boundaries that are not typically found in real infarcted tissue. The results therefore demonstrate how electrogram morphology can become ambiguous under controlled conditions, but they likely underestimate the full ambiguity present in biological tissue.

Second, the use of averaged conduction as the target variable for quantitative analysis and machine learning has downsides. Although this choice is physically motivated and provides a more realistic local descriptor than the diffusion at a single mesh node, averaging inevitably smooths the underlying conduction field and hence results in a loss of resolution. Consequently, narrow slow conduction zones and scars, sharp borders and small isolated patches of abnormal tissue are partly obscured. The cutoff radius used to compute averaged conduction is therefore an important modelling choice which requires a fine balance between the resolved resolution scale and the correlation with the recorded EGM signal. The present analysis used a single cutoff radius, and the robustness of feature-conduction relationships across spatial scales therefore remains untested.

Third, the simulation protocol was fixed and does not exactly reflect realistic situations. The tissue models used in this thesis do not include the full three-dimensional structure of the atrial wall, which can modify recorded signal morphology. Extending the models to three-dimensional or coupled epicardial-endocardial geometries would be important to determine whether the same morphology-substrate relationships remain visible when transmural activation and realistic anatomy are included. Furthermore, the analysis was based on electrograms generated from single activation sequences, which improves interpretability. However, arrhythmogenic tissue may show other effects that only appear under repeated or premature stimulation. This is particularly important for functional conduction abnormalities, which depend on activation history.

Fourth, the electrode configuration was fixed throughout the simulations, which made comparisons between substrates more controlled but did not capture the variability of clinical mapping. These variations alter the recorded feature distributions. Moreover, the simulated signals did not include other noise contributions that typically affect clinical recordings. Including these would likely broaden the feature distributions and reduce the separability between different conduction regimes. Furthermore, the current analysis only looked at unipolar electrograms, which further limits clinical transferability, where bipolar, multipolar and omnipolar electrodes are the norm. These signals have different sensitivities, which would alter the feature values, PCA structure and supervised learning performance.

Fifth, the extracted feature list is not exhaustive, and the PCA analysis revealed strong redundancy between several features, suggesting that a smaller set might achieve similar results. The present work therefore supports multidimensional feature analysis without identifying an optimal minimal feature set. Furthermore, PCA only captures linear feature relationships, so a low-dimensional PCA representation does not imply that the underlying mechanisms are intrinsically low-dimensional; it only reflects the dominant linear structure in the selected features under the simulated conditions. Non-linear dimensionality reduction methods would likely reveal additional structure, at the cost of interpretability.

Finally, the results were not directly validated against patient-specific clinical mapping data. Although comparisons with the literature suggest that several simulated findings are consistent with clinical observations, such comparisons do not establish quantitative clinical generalizability. Clinical validation would require high-density electroanatomic mapping data together with independent information about the underlying substrate. Until such validation is performed, the results should be viewed as mechanistic support rather than direct clinical evidence.

5.4 Future prospects

Future work should first test the robustness of the extracted features under more realistic recording conditions. Controlled noise, variable electrode height, and far-field contributions could

be added to the simulated EGMs to determine which features remain stable under acquisition variability. This would also clarify whether the PCA structure and machine learning predictions remain robust when local and non-local signal components are mixed.

A second direction is to refine the signal representation. Derivative-based features could be used to study whether fractionation is related to abrupt changes or irregularities in the first or second derivative of the EGM, providing a more mechanistic alternative to peak-count-based measures. In parallel, systematic feature removal, feature selection, or ablation studies could identify which descriptors are essential and which are redundant. Extending the analysis to bipolar, multipolar, and omnipolar electrograms would also be necessary to determine which findings are specific to unipolar recordings and which generalise across clinically used signal types.

The simulation framework itself could also be extended. The effect of the averaged conduction cutoff radius should be tested explicitly, since this parameter defines the spatial scale of the prediction target. More realistic tissue models could include anisotropy, patient-specific fibrosis patterns, curved three-dimensional atrial geometries, epicardial-endocardial differences, and multiple-beat or dynamic pacing protocols. These extensions would test whether the structural-functional ambiguity observed in this thesis persists under more realistic substrate and activation conditions.

Future work could also investigate whether more advanced machine-learning approaches improve the prediction and classification performance obtained in this study. The present analysis relied on relatively standard models, namely random forest regression for averaged conduction prediction and logistic regression for distance classification, which provide robust and interpretable baselines. However, more flexible models such as gradient-boosted decision trees, support vector machines, or neural networks may be able to capture non-linear relationships between EGM features and the underlying tissue properties more effectively. Such models should be evaluated carefully using cross-validation and independent test data to determine whether any improvement in performance outweighs the reduced interpretability and increased risk of overfitting.

Finally, clinical validation is essential. A promising approach would be to extract the same features from clinical CARTO electrograms, interpolate them onto the anatomical mesh used for electroanatomic mapping, and compare the resulting feature maps with voltage maps, local activation time maps, and independently defined scar or slow-conduction regions. This would determine whether the multidimensional feature relationships observed in simulation can support substrate interpretation in patient-specific data.

Chapter 6

Conclusion

This thesis systematically analysed simulated atrial electrograms from increasingly complex tissue models using both qualitative signal interpretation and quantitative feature analysis. The simulations reproduced key EGM morphologies associated with conduction abnormalities, largely independent of simulation type. Voltage amplitude was strongly modulated by substrate geometry and electrode position, supporting the view that fixed voltage thresholds are not universally transferable. Importantly, slow conduction regions influenced not only electrodes inside the affected tissue, but also neighbouring electrodes in healthy tissue.

Quantitatively, the extracted EGM features showed strong relationships with averaged conduction, supporting its use as a physically meaningful descriptor of the tissue region sampled by the electrode. PCA further showed that EGM variability was governed by a low-dimensional structure where the first principal component explained 65-70% of the variance and was dominated by amplitude- and slope-related features. This component was highly robust across different pathological tissue configurations and formed an approximate electrophysiological severity axis associated with averaged conduction. The secondary principal components explained an additional 25-30% of the variance and captured complementary information related to timing, asymmetry, and fractionation.

The reproducibility of this low-dimensional structure across datasets was confirmed by supervised machine learning applied to a merged dataset spanning all simulation types, such that the models had no knowledge of the underlying substrate. A logistic regression classified electrodes as located inside or outside affected tissue with 89.3% balanced accuracy, and a random forest regressor predicted averaged conduction with a mean absolute error of 1.2%. Together, these results support the feasibility of using feature-based analysis to investigate and potentially reduce the ambiguity between scar tissue and functional slow-conduction regions.

Overall, the results show that EGM-derived features preserve physically meaningful structure across increasingly complex simulated substrates. However, they also demonstrate why isolated electrogram markers are insufficient for uniquely distinguishing scar tissue from functional slow conduction regions. Multidimensional feature analysis, PCA, and subsequent machine learning approaches therefore provide a more appropriate framework for reducing this ambiguity.

Bibliography

- [1] J. W. Calvert and D. J. Lefer, “Overview of cardiac muscle physiology,” in *Cardiovascular Physiology Concepts*, 2nd ed., R. E. Klabunde, Ed. Lippincott Williams & Wilkins, 2012, ch. 6.
- [2] J. Rice, *Medical Terminology for Health Care Professionals*, 9th ed. USA: Pearson, 2017, [Online]. [Online]. Available: <https://www.pearsonhighered.com/assets/samplechapter/0/1/3/4/0134746279.pdf>
- [3] J. E. Hall and M. E. Hall, *Guyton and Hall Textbook of Medical Physiology*, 14th ed. Philadelphia, PA: Elsevier, 2021.
- [4] A. Sirajuddin, M. Y. Chen, C. S. White, and A. E. Arai, “Coronary venous anatomy and anomalies,” *Journal of Cardiovascular Computed Tomography*, vol. 14, no. 1, pp. 80–90, 2020.
- [5] H. E. Durham Jr., *Cardiac Anatomy*. John Wiley Sons, Ltd, 2017, ch. 1, pp. 5–21. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119357407.ch1>
- [6] T. Mikawa and T. Brand, “Chapter 5.1 - epicardial lineage: Origins and fates,” in *Heart Development and Regeneration*, N. Rosenthal and R. P. Harvey, Eds. Boston: Academic Press, 2010, pp. 325–344. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780123813329000165>
- [7] Wikiversity, “Wikijournal of medicine/medical gallery of blausen medical 2014 — wikiversity,,” 2026, [Online; accessed 28-May-2026]. [Online]. Available: https://en.wikiversity.org/w/index.php?title=WikiJournal_of_Medicine/Medical_gallery_of_Blausen_Medical_2014&oldid=2785552
- [8] R. L. Maynard and N. Downes, “Chapter 8 - histology of the vascular system,” in *Anatomy and Histology of the Laboratory Rat in Toxicology and Biomedical Research*. Academic Press, 2019, pp. 91–95. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128118375000083>
- [9] P. Kohl and M. Helmes, “Cardiovascular system – the heart,” in *Lecture Notes: Human Physiology*, 5th ed., O. H. Petersen, Ed. Blackwell Publishing Ltd, 2006, vol. 5, pp. 335–371.
- [10] K. Nagao, J. Toyama, I. Kodama, and K. Yamada, “Role of the conduction system in the endocardial excitation spread in the right ventricle,” *The American Journal of Cardiology*, vol. 48, no. 5, pp. 864–870, 1981. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0002914981903519>
- [11] Z. F. Issa, J. M. Miller, and D. P. Zipes, “Chapter 27 - epicardial ventricular tachycardia,” in *Clinical Arrhythmology and Electrophysiology: A Companion to Braunwald’s Heart*

- Disease (Second Edition)*, second edition ed. Philadelphia: W.B. Saunders, 2012, pp. 608–617. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9781455712748000270>
- [12] S. B. Yeon and N. Oyama, “Chapter 36 - the pericardium: Normal anatomy and spectrum of disease,” in *Cardiovascular Magnetic Resonance (Second Edition)*, second edition ed., W. J. Manning and D. J. Pennell, Eds. Philadelphia: Churchill Livingstone, 2010, pp. 488–497. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780443066863000369>
 - [13] J. Pinnell, S. Turner, and S. Howell, “Cardiac muscle physiology,” *Continuing Education in Anaesthesia Critical Care Pain*, vol. 7, no. 3, pp. 85–88, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1743181617304900>
 - [14] C. Antzelevitch and A. Burashnikov, “Overview of basic mechanisms of cardiac arrhythmia,” *Cardiac Electrophysiology Clinics*, vol. 3, no. 1, pp. 23–45, 2011.
 - [15] B. F. Fremgen and S. S. Frucht, *Medical Terminology: A Living Language*, 6th ed. Pearson, 2018.
 - [16] Y. Wang, D. DeMazumder, and J. A. Hill, “Chapter 7 - ionic fluxes and genesis of the cardiac action potential,” in *Muscle*, J. A. Hill and E. N. Olson, Eds. Boston/Waltham: Academic Press, 2012, pp. 67–85. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780123815101000077>
 - [17] M. J. Ackerman and D. E. Clapham, “Ion channels — basic science and clinical disease,” *New England Journal of Medicine*, vol. 336, no. 22, pp. 1575–1586, 1997. [Online]. Available: <https://www.nejm.org/doi/full/10.1056/NEJM199705293362207>
 - [18] I. Kotadia, J. Whitaker, C. Roney, S. Niederer, M. O’Neill, M. Bishop, and M. Wright, “Anisotropic cardiac conduction,” *Arrhythmia & Electrophysiology Review*, vol. 9, no. 4, p. 202, 2020.
 - [19] S. A. Bernstein and G. E. Morley, “Gap junctions and propagation,” *Cardiovascular Gap Junctions*, vol. 42, pp. 71–85, 2006.
 - [20] X. Wei, S. Yohannan, and J. R. Richards, “Physiology, cardiac repolarization dispersion and reserve,” in *StatPearls [Internet]*. StatPearls Publishing, 2023.
 - [21] E. L. Kessler, M. Boulaksil, H. V. van Rijen, M. A. Vos, and T. A. van Veen, “Passive ventricular remodeling in cardiac disease: focus on heterogeneity,” *Frontiers in physiology*, vol. 5, p. 482, 2014.
 - [22] A. Yartsev, “Normal processes of cardiac excitation and electrical activity,” <https://derangedphysiology.com/main/cicm-primary-exam/cardiovascular-system/Chapter-010/normal-processes-cardiac-excitation-and-electrical-activity>, 2020, deranged Physiology. Last updated: 2025-01-09. Accessed: 2025-05-05.
 - [23] P. A. Iaizzo, *Handbook of cardiac anatomy, physiology, and devices*. Springer Science & Business Media, 2010.
 - [24] E. Prystowsky, “Tachycardia-induced-tachycardia: a mechanism of initiation of atrial fibrillation,” *Atrial Arrhythmias: State of the Art. Armonk, NY: Futura*, pp. 81–95, 1995.
 - [25] Z. Chu, D. Yang, and X. Huang, “Conditions for the genesis of early afterdepolarization in a model of a ventricular myocyte,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 4, 2020.

- [26] A. L. Wit and P. A. Boyden, “Triggered activity and atrial fibrillation,” *Heart rhythm*, vol. 4, no. 3, pp. S17–S23, 2007.
- [27] G. Larraitz, M. F. E., and B. B. P., “Mechanisms of cardiac arrhythmias,” *Revista Española de Cardiología (English Edition)*, vol. 65, no. 2, pp. 113–201, 2012.
- [28] N. Vandersickel, S. Hendrickx, R. Van den Abeele, and B. Verstraeten, “Impact of topology on the number of loops during macro-re-entrant atrial tachycardia,” 2024.
- [29] N. Vandersickel, “Revelation after three decades of ablation therapy: two loops instead of a single re-entry loop drive atrial tachycardia,” *Project Repository Journal*, vol. 18, pp. 20–23, 2023, accessed: 30 May 2026. [Online]. Available: <https://www.europeandissemination.eu/article/revelation-after-three-decades-of-ablation-therapy-two-loops-instead-of-a-single-re-entry-loop-drive-atrial-tachycardia/20794>
- [30] H. Lin, R. Liu, and Z. Liu, “Electrocardiogram signal denoising,” Encyclopedia, accessed: 30 May 2026. [Online]. Available: <https://encyclopedia.pub/entry/52174>
- [31] A. K. Gupta, A. Maheshwari, R. Thakur, and Y. Y. Lokhandwala, “Cardiac mapping: utility or futility?” *Indian Pacing and Electrophysiology Journal*, vol. 2, no. 1, p. 20, 2002.
- [32] J. Sánchez and A. Loewe, “A review of healthy and fibrotic myocardium microstructure modeling and corresponding intracardiac electrograms,” *Frontiers in Physiology*, vol. 13, p. 908069, 2022.
- [33] S. Saha, D. Linz, D. Saha, A. McEwan, and M. Baumert, “Overcoming uncertainties in electrogram-based atrial fibrillation mapping: A review,” *Cardiovascular Engineering and Technology*, vol. 15, pp. 52–64, 2023.
- [34] N. M. S. de Groot, D. Shah, P. M. Boyle, E. Anter, G. D. Clifford, I. Deisenhofer, T. Deneke, P. van Dessel, O. Doessel, P. Dilaveris *et al.*, “Critical appraisal of technologies to assess electrical activity during atrial fibrillation: a position paper from the european heart rhythm association and european society of cardiology working group on ecardiology in collaboration with the heart rhythm society, asia pacific heart rhythm society, latin american heart rhythm society and computing in cardiology,” *EP Europace*, vol. 24, no. 2, pp. 313–330, 2022. [Online]. Available: <https://doi.org/10.1093/europace/euab254>
- [35] S. Saha, D. Linz, P. Sanders, and M. Baumert, “Beamforming-inspired spatial filtering technique for intracardiac electrograms,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 4254–4257.
- [36] B. Abdi, M. S. van Schie, N. M. de Groot, and R. C. Hendriks, “Analyzing the effect of electrode size on electrogram and activation map properties,” *Computers in Biology and Medicine*, vol. 134, p. 104467, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482521002614>
- [37] E. Anter and M. E. Josephson, “Bipolar voltage amplitude: What does it really mean?” *Heart Rhythm*, vol. 13, no. 1, pp. 326–327, 2016. [Online]. Available: <https://doi.org/10.1016/j.hrthm.2015.09.033>
- [38] S. Gaeta, T. D. Bahnson, and C. Henriquez, “Mechanism and magnitude of bipolar electrogram directional sensitivity: Characterizing underlying determinants of bipolar amplitude,” *Heart Rhythm*, vol. 17, no. 5, Part A, pp. 777–785, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1547527119311026>

-
- [39] J. M. Stinnett-Donnelly, N. Thompson, N. Habel, V. Petrov-Kondratov, D. D. C. de Sa, J. H. Bates, and P. S. Spector, “Effects of electrode size and spacing on the resolution of intracardiac electrograms,” *Coronary Artery Disease*, vol. 23, no. 2, pp. 126–132, March 2012. [Online]. Available: <https://doi.org/10.1097/MCA.0b013e3283507a9b>
 - [40] D. Nairn, H. Lehrmann, B. Müller-Edenborn, S. Schuler, T. Arentz, O. Dössel, A. Jadidi, and A. Loewe, “Comparison of unipolar and bipolar voltage mapping for localization of left atrial arrhythmogenic substrate in patients with atrial fibrillation,” *Frontiers in Physiology*, vol. 11, p. 575846, 2020.
 - [41] L. J. van der Does, P. Knops, C. P. Teuwen, C. Serban, R. Starreveld, E. A. Lanthers, E. M. Mouws, C. Kik, A. J. Bogers, and N. M. de Groot, “Unipolar atrial electrogram morphology from an epicardial and endocardial perspective,” *Heart Rhythm*, vol. 15, no. 6, pp. 879–887, Jun. 2018.
 - [42] Z. Ye, M. S. van Schie, L. Pool, A. Heida, P. Knops, Y. J. H. J. Taverne, B. J. J. M. Brundel, and N. M. S. de Groot, “Characterization of unipolar electrogram morphology: a novel tool for quantifying conduction inhomogeneity,” *Europace*, vol. 25, no. 11, p. euad324, 2023.
 - [43] V. Schlageter, A. Luca, P. Badertscher, P. Krisai, T. Kueffer, D. Spreen, J. Katic, S. Osswald, B. Schaer, C. Sticherling, M. Kühne, and S. Knecht, “Effect of electrode size and distance to tissue on unipolar and bipolar voltage electrograms and their implications for a near-field cutoff,” *Scientific Reports*, vol. 14, p. 27184, 2024. [Online]. Available: <https://www.nature.com/articles/s41598-024-78627-5>
 - [44] M. Takigawa, T. Kitamura, S. Basu, M. Bartal, C. A. Martin, R. Martin, G. Cheniti, K. Vlachos, X. Pillois, A. Frontera, G. Massoulié, N. Thompson, F. Bourier, A. Lam, J. Duchateau, T. Pambrun, A. Denis, N. Derval, H. Cochet, M. Haïssaguerre, F. Sacher, M. Hocini, and P. Jaïs, “Effect of electrode size and spacing on electrograms: Optimized electrode configuration for near-field electrogram characterization,” *Heart Rhythm*, vol. 19, no. 1, pp. 102–112, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1547527121021147>
 - [45] M. Hwang, J. Kim, B. Lim, J.-S. Song, B. Joung, E. B. Shim, and H.-N. Pak, “Multiple factors influence the morphology of the bipolar electrogram: An in silico modeling study,” *PLoS Computational Biology*, vol. 15, no. 4, p. e1006765, 2019.
 - [46] P. Mankad and G. Kalahasty, “Antiarrhythmic drugs: risks and benefits,” *Medical Clinics of North America*, vol. 103, no. 5, pp. 821–834, 2019.
 - [47] D. E. Krummen, G. Ho, C. T. Villongco, J. Hayase, and A. A. Schricker, “Ventricular fibrillation: triggers, mechanisms and therapies,” *Future cardiology*, vol. 12, no. 3, pp. 373–390, 2016.
 - [48] R. Parameswaran, A. M. Al-Kaisey, and J. M. Kalman, “Catheter ablation for atrial fibrillation: current indications and evolving technologies,” *Nature Reviews Cardiology*, vol. 18, no. 3, pp. 210–225, 2021.
 - [49] A. L. Hodgkin and A. F. Huxley, “A quantitative description of membrane current and its application to conduction and excitation in nerve,” *The Journal of Physiology*, vol. 117, no. 4, pp. 500–544, 1952.
 - [50] O. H. Schmitt, *Biological Information Processing Using the Concept of Interpenetrating Domains*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1969, pp. 325–331. [Online]. Available: https://doi.org/10.1007/978-3-642-87086-6_18

- [51] L. Tung, “A bi-domain model for describing ischemic myocardial d-c potentials,” PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1978. [Online]. Available: <https://dspace.mit.edu/entities/publication/375f998c-5980-4274-83de-b22fff1e5bcc>
- [52] R. H. Clayton, O. Bernus, E. M. Cherry, H. Dierckx, F. H. Fenton, L. Mirabella, A. V. Panfilov, F. B. Sachse, G. Seemann, and H. Zhang, “Models of cardiac tissue electrophysiology: Progress, challenges and open questions,” *Progress in Biophysics and Molecular Biology*, vol. 104, no. 1–3, pp. 22–48, 2011.
- [53] M. Potse, B. Dubé, A. Vinet, and R. Cardinal, “A comparison of monodomain and bidomain propagation models for the human heart,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 3895–3898.
- [54] M. J. Bishop and G. Plank, “Representing cardiac bidomain bath-loading effects by an augmented monodomain approach: Application to complex ventricular models,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 4, pp. 1066–1075, Apr. 2011.
- [55] S. Rossi and B. E. Griffith, “Incorporating inductances in tissue-scale models of cardiac electrophysiology,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, no. 9, p. 093926, September 2017. [Online]. Available: <https://doi.org/10.1063/1.5000706>
- [56] J. Sánchez, J. F. Gomez, L. Martinez-Mateu, L. Romero, J. Saiz, and B. Trenor, “Heterogeneous effects of fibroblast-myocyte coupling in different regions of the human atria under conditions of atrial fibrillation,” *Frontiers in Physiology*, vol. 10, p. 847, Jul. 2019.
- [57] E. J. Vigmond, M. Hughes, G. Plank, and L. J. Leon, “Computational tools for modeling electrical activity in cardiac tissue,” *Journal of Electrocardiology*, vol. 36, no. Supplement 1, pp. 69–74, Dec. 2003.
- [58] G. Plank, A. Loewe, A. Neic, C. Augustin, Y.-L. Huang, M. A. Gsell, E. Karabelas, M. Nothstein, A. J. Prassl, J. Sánchez, G. Seemann, and E. J. Vigmond, “The opencarp simulation environment for cardiac electrophysiology,” *Computer Methods and Programs in Biomedicine*, vol. 208, p. 106223, Sep. 2021.
- [59] M. Courtemanche, R. J. Ramirez, and S. Nattel, “Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 275, no. 1, pp. H301–H321, 1998.
- [60] C. Corrado, J. Whitaker, H. Chubb, S. Williams, M. Wright, J. Gill, M. O’Neill, and S. A. Niederer, “Personalized models of human atrial electrophysiology derived from endocardial electrograms,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 735–742, 2017.
- [61] M. Wilhelms, H. Hettmann, M. M. Maleckar, J. T. Koivumäki, O. Dössel, and G. Seemann, “Benchmarking electrophysiological models of human atrial myocytes,” *Frontiers in Physiology*, vol. 3, p. 487, 2013.
- [62] N. F. Otani and R. F. J. Gilmour, “Memory models for the electrical properties of local cardiac systems,” *Journal of Theoretical Biology*, vol. 187, no. 3, pp. 409–436, 1997.
- [63] A. Baher, Z. Qu, A. Hayatdavoudi, S. T. Lamp, M.-J. Yang, F. Xie, S. Turner, A. Garfinkel, and J. N. Weiss, “Short-term cardiac memory and mother rotor fibrillation,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 292, no. 1, pp. H180–H189, 2007.

- [64] E. M. Cherry, H. M. Hastings, and S. J. Evans, “Dynamics of human atrial cell models: Restitution, memory, and intracellular calcium dynamics in single cells,” *Progress in Biophysics and Molecular Biology*, vol. 98, no. 1, pp. 24–37, 2008.
- [65] T. Nezlobinsky and A. Okenov, “Finitewave,” n.d., mIT License, accessed 2026-04-14. [Online]. Available: <https://github.com/finitewave/finitewave>
- [66] L. J. M. E. van der Does and N. M. S. de Groot, “Inhomogeneity and complexity in defining fractionated electrograms,” *Heart Rhythm*, vol. 14, no. 4, pp. 616–624, 2017.
- [67] A. Neic, F. O. Campos, A. J. Prassl, S. A. Niederer, M. J. Bishop, E. J. Vigmond, and G. Plank, “Efficient computation of electrograms and ecgs in human whole heart simulations using a reaction-eikonal model,” *Journal of Computational Physics*, vol. 346, pp. 191–211, Oct. 2017.
- [68] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. New York: Oxford University Press, 1995.
- [69] J. P. Ugarte, A. Orozco-Duque, C. Tobón, V. Kremen, D. Novak, J. Saiz, T. Oesterlein, C. Schmitt, A. Luik, and J. Bustamante, “Dynamic approximate entropy electroanatomic maps detect rotors in a simulated atrial fibrillation model,” *PLoS ONE*, vol. 9, no. 12, p. e114577, 2014.
- [70] M. J. Bishop and G. Plank, “Bidomain ecg simulations using an augmented monodomain model for the cardiac source,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 8, pp. 2297–2307, 2011.
- [71] M. W. Keller, S. Schuler, G. Seemann, and O. Dössel, “Differences in intracardiac signals on a realistic catheter geometry using mono- and bidomain models,” in *2012 Computing in Cardiology*, 2012, pp. 305–308.
- [72] M. Walker, S. Schuler, M. Wilhelm, G. Lenis, G. Seemann, and C. Schmitt, “Characterization of radiofrequency ablation lesion development based on simulated and measured intracardiac electrograms,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2467–2478, 2014.
- [73] P. C. Franzone, L. F. Pavarino, S. Scacchi, and B. Taccardi, “Monophasic action potentials generated by bidomain modeling as a tool for detecting cardiac repolarization times,” *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 293, no. 5, pp. H2776–H2786, 2007.
- [74] J. Sánchez, G. Luongo, M. Nothstein, L. A. Unger, J. Saiz, B. Trenor, A. Luik, O. Dössel, and A. Loewe, “Using machine learning to characterize atrial fibrotic substrate from intracardiac signals with a hybrid in silico and in vivo dataset,” *Frontiers in Physiology*, vol. 12, p. 699291, 2021.
- [75] A. Jadidi, M. Nothstein, J. Chen, H. Lehrmann, O. Dössel, J. Allgeier, D. Trenk, F.-J. Neumann, A. Loewe, B. Müller-Edenborn, and T. Arentz, “Specific electrogram characteristics identify the extra-pulmonary vein arrhythmogenic sources of persistent atrial fibrillation – characterization of the arrhythmogenic electrogram patterns during atrial fibrillation and sinus rhythm,” *Scientific Reports*, vol. 10, p. 9147, 2020.
- [76] Z. Ye, M. S. van Schie, and N. M. S. de Groot, “Signal fingerprinting as a novel diagnostic tool to identify conduction inhomogeneity,” *Frontiers in Physiology*, vol. 12, p. 652128, 2021.

- [77] M. S. Spach and P. C. Dolber, “Relating extracellular potentials and their derivatives to anisotropic propagation at a microscopic level in human cardiac muscle: Evidence for electrical uncoupling of side-to-side fiber connections with increasing age,” *Circulation Research*, vol. 58, no. 3, pp. 356–371, March 1986.
- [78] M. Beheshti, K. Magtibay, S. Massé, A. Porta-Sanchez, S. Haldar, A. Bhaskaran, S. Nayyar, B. Glover, D. C. Deno, E. J. Vigmond, and K. Nanthakumar, “Determinants of atrial bipolar voltage: Inter electrode distance and wavefront angle,” *Computers in Biology and Medicine*, vol. 102, pp. 449–457, 2018. [Online]. Available: <https://doi.org/10.1016/j.compbimed.2018.07.011>
- [79] V. Jacquemet, N. Virag, Z. Ihara, L. Dang, O. Blanc, S. Zozor, J.-M. Vesin, L. Kappenberger, and C. S. Henriquez, “Study of unipolar electrogram morphology in a computer model of atrial fibrillation,” *Journal of Cardiovascular Electrophysiology*, vol. 14, no. S10, pp. S172–S179, 2003.
- [80] S. Schuler, M. W. Keller, T. Oesterlein, G. Seemann, and O. Dössel, “Influence of catheter orientation, tissue thickness and conduction velocity on the intracardiac electrogram,” *Biomedical Engineering / Biomedizinische Technik*, vol. 58, no. s1, 2013.
- [81] D. Nairn, D. Hunyar, J. Sánchez, O. Dössel, and A. Loewe, “Impact of electrode size on electrogram voltage in healthy and diseased tissue,” in *Computing in Cardiology Conference (CinC)*. Brno, Czech Republic: IEEE, 2020.
- [82] L. Schicketanz, L. A. Unger, J. Sánchez, O. Dössel, and A. Loewe, “Separating atrial near fields and atrial far fields in simulated intra-atrial electrograms,” *Current Directions in Biomedical Engineering*, vol. 7, no. 2, pp. 175–178, 2021.
- [83] T. Oesterlein, D. Frisch, A. Loewe, G. Seemann, C. Schmitt, O. Doessel, and A. Luik, “Basket-type catheters: Diagnostic pitfalls caused by deformation and limited coverage,” *BioMed Research International*, vol. 2016, p. 5340574, 2016.
- [84] L. A. Unger, T. G. Oesterlein, A. Loewe, and O. Dössel, “Noise quantification and noise reduction for unipolar and bipolar electrograms,” in *Computing in Cardiology Conference (CinC)*. IEEE, 2021, pp. 1–4.
- [85] A. Frontera, M. Takigawa, R. Martin, P. Jaïs, M. Haïssaguerre, and N. Derval, “Electrogram signature of specific activation patterns: Analysis of atrial tachycardias at high-density endocardial mapping,” *Heart Rhythm*, vol. 15, no. 1, pp. 28–37, Jan. 2018.
- [86] F. D. Ramirez, M. Meo, C. Dallet, P. Krisai, K. Vlachos, A. Frontera, M. Takigawa, Y. Nakatani, T. Nakashima, C. André, T. Kamakura, T. Takagi, A. Grapezzi, R. Tixier, R. Chauvel, G. Cheniti, J. Duchateau, T. Pambrun, F. Sacher, M. Hocini, and N. Derval, “High-resolution mapping of reentrant atrial tachycardias: Relevance of low bipolar voltage,” *Heart Rhythm*, 2023.
- [87] K. Miyamoto, T. Tsuchiya, S. Narita, T. Yamaguchi, Y. Nagamoto, S. ichi Ando, K. Hayashida, Y. Tanioka, and N. Takahashi, “Bipolar electrogram amplitudes in the left atrium are related to local conduction velocity in patients with atrial fibrillation,” *Europace*, vol. 11, no. 12, pp. 1597–1605, 2009.
- [88] A. Codreanu, F. Odille, E. Aliot, P.-Y. Marie, I. Magnin-Poull, M. Andronache, D. Mandry, W. Djaballah, D. Régent, J. Felblinger, and C. de Chillou, “Electroanatomic characterization of post-infarct scars: comparison with 3-dimensional myocardial scar reconstruction based on magnetic resonance imaging,” *Journal of the American College of Cardiology*, vol. 52, no. 10, pp. 839–842, September 2008.

-
- [89] M. Saha, C. Roney, H. Cochet, S. Niederer, E. Vigmond, and S. Nattel, “Myocardial transmural electrical disruption affects electrogram pattern,” in *2019 Computing in Cardiology (CinC)*, 2019, pp. Page 1–Page 4.
 - [90] E. Irakoze, C. H. R. Gowda, and V. Jacquemet, “Asymmetry of unipolar electrograms in a thin tissue with epicardial-endocardial activation delay,” in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4.
 - [91] V. Jacquemet and C. S. Henriquez, “Genesis of complex fractionated atrial electrograms in zones of slow conduction: a computer model of microfibrosis,” *Heart Rhythm*, vol. 6, no. 6, pp. 803–810, June 2009.
 - [92] A. N. Ganesan, P. Kuklik, D. H. Lau, A. G. Brooks, M. Baumert, P. W. Lim, S. Thanigaimani, S. Nayyar, M. R. Mahajan, J. M. Kalman, K. C. Roberts-Thomson, and P. Sanders, “Bipolar electrogram shannon entropy at sites of rotational activation: Implications for ablation of atrial fibrillation,” *Circulation: Arrhythmia and Electrophysiology*, vol. 6, no. 1, pp. 48–57, Feb. 2013.
 - [93] P. S. Cuculich, J. Zhang, Y. Wang, K. A. Desouza, R. Vijayakumar, P. K. Woodard, and Y. Rudy, “The electrophysiological cardiac ventricular substrate in patients after myocardial infarction: Noninvasive characterization with electrocardiographic imaging,” *Journal of the American College of Cardiology*, vol. 58, no. 18, pp. 1893–1902, October 2011.
 - [94] M. W. Keller, A. Luik, M. S. Abady, G. Seemann, C. Schmitt, and O. Dössel, “Influence of three-dimensional fibrotic patterns on simulated intracardiac electrogram morphology,” in *Computing in Cardiology 2013*. IEEE, 2013, pp. 923–926.
 - [95] C. H. Roney, J. D. Bayer, S. Zahid, M. Meo, P. M. J. Boyle, N. A. Trayanova, M. Haïssaguerre, R. Dubois, H. Cochet, and E. J. Vigmond, “Modelling methodology of atrial fibrosis affects rotor dynamics and electrograms,” *EP Europace*, vol. 18, no. suppl_4, pp. iv146–iv155, Dec. 2016.
 - [96] F. Bogun, S. Krishnan, M. Siddiqui, E. Good, J. E. Marine, C. Schuger, H. Oral, A. Chugh, F. Pelosi, and F. Morady, “Electrogram characteristics in postinfarction ventricular tachycardia: effect of infarct age,” *Journal of the American College of Cardiology*, vol. 46, no. 4, pp. 667–674, August 2005.
 - [97] M. Masjedian, C. Jung, P. Kuklik, F.-A. Alken, A.-K. Kahle, N. Klatt, K. Scherschel, and C. Meyer, “A novel algorithm for 3-d visualization of electrogram duration for substrate-mapping in patients with ischemic heart disease and ventricular tachycardia,” *PLOS ONE*, vol. 16, no. 7, p. e0254683, July 2021.
 - [98] C. M. Costa, G. C. Anderson, V. M. F. Meijborg, C. O’Shea, M. J. Shattock, P. Kirchhof, R. Coronel, S. Niederer, and J. Winter, “The amplitude-normalized area of a bipolar electrogram as a measure of local conduction delay in the heart,” *Frontiers in Physiology*, vol. 11, p. 465, 2020.
 - [99] A. Frontera, L. R. Limite, S. Pagani, M. Cireddu, K. Vlachos, C. Martin, M. Takigawa *et al.*, “Electrogram fractionation during sinus rhythm occurs in normal voltage atrial tissue in patients with atrial fibrillation,” *Pacing and Clinical Electrophysiology*, vol. 45, no. 3, pp. 304–313, 2022.
 - [100] R. Cabrera-Lozoya, B. Berte, H. Cochet, P. Jaïs, N. Ayache, and M. Sermesant, “Image-based biophysical simulation of intracardiac abnormal ventricular electrograms,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1446–1454, Jul. 2017.

- [101] R. Morgan, M. A. Colman, H. Chubb, G. Seemann, and O. V. Aslanidi, “Slow conduction in the border zones of patchy fibrosis stabilizes the drivers for atrial fibrillation: Insights from multi-scale human atrial modeling,” *Frontiers in Physiology*, vol. 7, p. 474, 2016.
- [102] T. A. Gokhale, E. M. Devecchi, and C. S. Henriquez, “Modeling dynamics in diseased cardiac tissue: Impact of model choice,” *Chaos*, vol. 27, no. 9, p. 093909, 2017.
- [103] M. W. Krueger, K. S. Rhode, M. D. O’Neill, C. A. Rinaldi, J. Gill, R. Razavi, G. Seemann, and O. Doessel, “Patient-specific modeling of atrial fibrosis increases the accuracy of sinus rhythm simulations and may explain maintenance of atrial fibrillation,” *Journal of Electrocardiology*, vol. 47, no. 3, pp. 324–328, 2014.
- [104] M. W. Krueger, G. Seemann, K. Rhode, D. U. J. Keller, C. Schilling, A. Arujuna, J. Gill, M. D. O’Neill, R. Razavi, and O. Doessel, “Personalization of atrial anatomy and electrophysiology as a basis for clinical modeling of radio-frequency ablation of atrial fibrillation,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 1, pp. 73–84, Jan. 2013.
- [105] B. Lim, J. Kim, M. Hwang, J.-S. Song, J. K. Lee, H.-T. Yu, T.-H. Kim, J.-S. Uhm, B. Joung, M.-H. Lee, and H.-N. Pak, “In situ procedure for high-efficiency computational modeling of atrial fibrillation reflecting personal anatomy, fiber orientation, fibrosis, and electrophysiology,” *Scientific Reports*, vol. 10, no. 1, p. 2417, 2020.
- [106] K. H. W. J. T. Tusscher and A. V. Panfilov, “Influence of diffuse fibrosis on wave propagation in human ventricular tissue,” *EP Europace*, vol. 9, no. suppl.6, pp. vi38–vi45, Nov. 2007.
- [107] R. H. Clayton, “Dispersion of recovery and vulnerability to re-entry in a model of human atrial tissue with simulated diffuse and focal patterns of fibrosis,” *Frontiers in Physiology*, vol. 9, p. 1052, 2018.
- [108] E. Vigmond, A. Pashaei, S. Amraoui, H. Cochet, and M. Haïssaguerre, “Percolation as a mechanism to explain atrial fractionated electrograms and reentry in a fibrosis model based on imaging data,” *Heart Rhythm*, vol. 13, no. 7, pp. 1536–1543, Jul. 2016.
- [109] S. Alonso and M. Bär, “Reentry near the percolation threshold in a heterogeneous discrete model for cardiac tissue,” *Physical Review Letters*, vol. 110, no. 15, p. 158101, 2013.
- [110] T. Ashihara, R. Haraguchi, K. Nakazawa, T. Namba, T. Ikeda, Y. Nakazawa, T. Ozawa, M. Ito, M. Horie, and N. A. Trayanova, “The role of fibroblasts in complex fractionated electrograms during persistent/permanent atrial fibrillation: Implications for electrogram-based catheter ablation,” *Circulation Research*, vol. 110, no. 2, pp. 275–284, 2012.
- [111] J. Sánchez, M. Nothstein, L. Unger, J. Saiz, B. Trénor, and O. Dössel, “Influence of fibrotic tissue arrangement on intracardiac electrograms during persistent atrial fibrillation,” in *Computing in Cardiology*. IEEE, 2017, pp. 47–50.
- [112] A. Okenov, T. Nezlobinsky, C. A. Glashan, K. Zeppenfeld, N. Vandersickel, and A. V. Panfilov, “Analysis and generation of the fibrosis textures based on histology data of the human hearts with non-ischemic cardiomyopathy,” *Physics in Medicine & Biology*, vol. 70, no. 13, p. 135002, 2025.
- [113] K. H. Jaeger, J. D. Trotter, X. Cai, H. Arevalo, and A. Tveito, “Evaluating computational efforts and physiological resolution of mathematical models of cardiac tissue,” *Scientific Reports*, vol. 14, p. 16954, 2024.

-
- [114] B. Zhu, G. Zhang, S. Xie, Y. Luan, W. Cao, J. Xu, S. Zhang, J. Tian, F. Wang, and S. Li, “The characterization of functional conduction block in patients with multiple types of atrial tachycardia: A discussion on the mechanism of multiple atrial tachycardia,” *Journal of Interventional Cardiac Electrophysiology*, vol. 67, pp. 1793–1806, 2024. [Online]. Available: <https://doi.org/10.1007/s10840-024-01817-8>
 - [115] I. T. Jolliffe, “Principal component analysis: A beginner’s guide — i. introduction and application,” *Weather*, vol. 41, no. 7, pp. 206–211, 1986.
 - [116] J. Shlens, “A tutorial on principal component analysis,” arXiv:1404.1100, 2014, version 3.02, April 7, 2014.
 - [117] K. Ntagiantas, E. Pignatelli, N. S. Peters, C. D. Cantwell, R. A. Chowdhury, and A. A. Bharath, “Estimation of fibre architecture and scar in myocardial tissue using electrograms: An in-silico study,” *Biomedical Signal Processing and Control*, vol. 89, p. 105746, 2024.
 - [118] G. Jimenez-Perez, J. Acosta, J. Bocanegra-Pérez, E. Arana-Rueda, M. Frutos-López, J. A. Sánchez-Brotons, H. Llamas-Gómez, R. Di Massa Pezzutti, C. González de la Portilla Concha, O. Camara, and A. Pedrote, “Delineation of intracavitary electrograms for the automatic quantification of decrement-evoked potentials in the coronary sinus with deep-learning techniques,” *Frontiers in Physiology*, vol. Volume 15 - 2024, 2024. [Online]. Available: <https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2024.1331852>
 - [119] B. Verma, T. Oesterlein, A. Loewe, A. Luik, C. Schmitt, and O. Dössel, “Regional conduction velocity calculation from clinical multichannel electrograms in human atria,” *Computers in Biology and Medicine*, vol. 92, pp. 188–196, Jan. 2018.
 - [120] G. Canino, A. D. Costanzo, N. Salerno, I. Leo, M. Cannataro, P. H. Guzzi, P. Veltri, S. Sorrentino, S. D. Rosa, and D. Torella, “Artificial intelligence in cardiac electrophysiology: A clinically oriented review with engineering primers,” *Bioengineering*, vol. 12, no. 10, p. 1102, 2025.
 - [121] S.-T. Jeong, “Stability of finite difference schemes on the diffusion equation with discontinuous coefficients,” 2018, available via MIT repository. [Online]. Available: <http://math.mit.edu>
 - [122] L. Tung and J. R. Borderies, “Analysis of electric field stimulation of single cardiac muscle cells,” *Biophysical Journal*, vol. 63, no. 2, pp. 371–386, 1992.
 - [123] V. Monasterio, E. Pueyo, J. F. Rodríguez-Matas, and J. Carro, “Cardiac cells stimulated with an axial current-like waveform reproduce electrophysiological properties of tissue fibers,” *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107121, 2022.
 - [124] J. P. Keener, “Propagation and its failure in coupled systems of discrete excitable cells,” *SIAM Journal on Applied Mathematics*, vol. 47, no. 3, pp. 556–572, 1987.
 - [125] S. F. Pravdin, “Anti-tachycardia pacing simulation in the human heart left ventricle using the Luo-Rudy model,” in *2024 IEEE SIBIRCON*, 2024, pp. 1–5.
 - [126] Z. Li, C. Lambranzi, D. Wu, A. Segato, F. De Marco, E. Vander Poorten, J. Dankelman, and E. De Momi, “Robust path planning via learning from demonstrations for robotic catheters in deformable environments,” *IEEE Transactions on Biomedical Engineering*, vol. 72, no. 1, pp. 324–335, 2025.

- [127] N. V. Belikov, I. V. Khaydukova, I. E. Poludkin, and A. S. Borde, “Evolution and current state of robotic catheters for endovascular surgery: A comprehensive review,” *Engineering Science and Technology, an International Journal*, vol. 57, p. 101789, 2024.
- [128] T. Nezlobinsky and A. Okenov, “Finitewave documentation,” 2025, accessed: 2026-03-27.
- [129] M. Masuda, Y. Matsuda, H. Uematsu, and T. Mano, “Remote entrainment pacing from multiple distant areas to identify a slow conduction isthmus of a reentrant circuit in scar-related atrial tachycardia,” *HeartRhythm Case Reports*, vol. 5, no. 10, pp. 513–517, 2019.
- [130] E. Anter, M. Duytschaever, C. Shen, T. Strisciuglio, E. Leshem, F. M. Contreras-Valdes, J. W. Waks, and A. E. Buxton, “Activation mapping with integration of vector and velocity information improves the ability to identify the mechanism and location of complex scar-related atrial tachycardias,” *Circulation: Arrhythmia and Electrophysiology*, vol. 11, no. 8, p. e006536, 2018.
- [131] S.-H. Liu, Y.-J. Lin, P.-T. Lee, J. J. Vicera, S.-L. Chang, L.-W. Lo, and Y.-F. Hu, “The isthmus characteristics of scar-related macroreentrant atrial tachycardia in patients with and without cardiac surgery,” *Journal of Cardiovascular Electrophysiology*, vol. 32, no. 6, pp. 1535–1545, 2021.
- [132] J. J. B. Vicera, Y.-J. Lin, P.-T. Lee, S.-L. Chang, L.-W. Lo, Y.-F. Hu, F.-P. Chung, C.-Y. Lin, and T.-Y. Chang, “Identification of critical isthmus using coherent mapping in patients with scar-related atrial tachycardia,” *Journal of Cardiovascular Electrophysiology*, vol. 31, no. 6, pp. 1436–1445, 2020.
- [133] P. Konovalov, D. Mangileva, A. Dokuchaev, O. Solovyova, and A. V. Panfilov, “Rotational activity around an obstacle in 2d cardiac tissue in presence of cellular heterogeneity,” *Mathematics*, vol. 9, no. 23, 2021. [Online]. Available: <https://www.mdpi.com/2227-7390/9/23/3090>
- [134] M. Shenasa, S.-M. Razavi, H. Shenasa, and A. Al-Ahmad, “The ideal cardiac mapping system,” *Cardiac Electrophysiology Clinics*, vol. 11, no. 4, pp. 739–748, 2019.
- [135] T. G. Oesterlein, J. Schmid, S. Bauer, A. Jadidi, C. Schmitt, O. Dössel, and A. Luik, “Analysis and visualization of intracardiac electrograms in diagnosis and research: Concept and application of kapavie,” *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 165–173, 2016.
- [136] M. Brennan, M. Palaniswami, and P. Kamen, “Do existing measures of poincaré plot geometry reflect nonlinear features of heart rate variability?” *IEEE Transactions on Bio-medical Engineering*, vol. 48, no. 11, pp. 1342–1347, November 2001.
- [137] E.-F. Chou, M. Khine, T. Lockhart, and R. Soangra, “Effects of ecg data length on heart rate variability among young healthy adults,” *Sensors*, vol. 21, no. 18, p. 6286, 2021.
- [138] C. Lerma, O. Infante, H. Pérez-Grovas, and M. V. José, “Poincaré plot indexes of heart rate variability capture dynamic adaptations after haemodialysis in chronic renal failure patients,” *Clinical Physiology and Functional Imaging*, vol. 23, no. 2, pp. 72–80, 2003.
- [139] A.-O. Boudraa and F. Salzenstein, “Teager–kaiser energy methods for signal and image analysis: A review,” *Digital Signal Processing*, vol. 78, pp. 338–375, 2018.
- [140] G. Deng, “A study of sampling frequency for efficient demodulation of am signals,” in *Proceedings of the 1997 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 4. Hong Kong: IEEE, Jun. 1997, pp. 2677–2680.

- [141] C. Kamath, “Ecg beat classification using features extracted from teager energy functions in time and frequency domains,” *IET Signal Processing*, vol. 5, no. 6, pp. 575–581, September 2011.
- [142] NumXL. (2016) Principal component analysis (pca) 102. Accessed: 2026-04-08. [Online]. Available: <https://numxl.com/blogs/principal-component-analysis-pca-102/>
- [143] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.
- [144] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [145] V. Bewick, L. Cheek, and J. Ball, “Statistics review 14: Logistic regression,” *Critical care (London, England)*, vol. 9, pp. 112–8, 03 2005.

Appendix A

Supplementary material

A.1 Supplementary material for methods

Table A.1: Model constants for the implementation of the Courtemanche model and tissue-level parameters in Finitewave.

Model constants			Model constants		
Symbol	Definition	Value	Symbol	Definition	Value
R	Gas constant	$8314.0 \text{ J (mmol K)}^{-1}$	T	Temperature	310.0 K
F	Faraday constant	$96485.0 \text{ C mol}^{-1}$	V_{cell}	Cell volume	$20100 \mu\text{m}^3$
V_i	Intracellular volume	$13668 \mu\text{m}^3$	V_{up}	SR uptake compartment volume	$754.4736 \mu\text{m}^3$
V_{rel}	SR release compartment volume	$65.6064 \mu\text{m}^3$	$[\text{K}^+]_o$	Extracellular K^+ concentration	5.4 mM
$[\text{Na}^+]_o$	Extracellular Na^+ concentration	140 mM	$[\text{Ca}^{2+}]_o$	Extracellular Ca^{2+} concentration	1.8 mM
g_{Na}	Maximal I_{Na} conductance	7.8 nS/pF	g_{K1}	Maximal I_{K1} conductance	0.09 nS/pF
g_{to}	Maximal I_{to} conductance	0.1652 nS/pF	g_{Kr}	Maximal I_{Kr} conductance	0.0294 nS/pF
g_{Ks}	Maximal I_{Ks} conductance	0.129 nS/pF	$g_{\text{Ca,L}}$	Maximal $I_{\text{Ca,L}}$ conductance	0.1238 nS/pF
$g_{\text{b,Ca}}$	Maximal background Ca^{2+} conductance	0.00113 nS/pF	$g_{\text{b,Na}}$	Maximal background Na^+ conductance	0.000674 nS/pF
$I_{\text{NaK,max}}$	Maximal I_{NaK}	0.6 pA/pF	$I_{\text{NaCa,max}}$	Maximal I_{NaCa}	1600 pA/pF
$I_{\text{p,Ca,max}}$	Maximal $I_{\text{p,Ca}}$	0.275 pA/pF	$I_{\text{up,max}}$	Maximal I_{up}	0.005 mM/ms
K_{Q10}	Temperature scaling factor	3	$K_{\text{m,Na(i)}}$	$[\text{Na}^+]_i$ half-saturation constant for I_{NaK}	10 mM
$K_{\text{m,K(o)}}$	$[\text{K}^+]_o$ half-saturation constant for I_{NaK}	1.5 mM	$K_{\text{m,Na}}$	$[\text{Na}^+]_i$ half-saturation constant for I_{NaCa}	87.5 mM
$K_{\text{m,Ca}}$	$[\text{Ca}^{2+}]_o$ half-saturation constant for I_{NaCa}	1.38 mM	k_{sat}	Saturation factor for I_{NaCa}	0.1
k_{rel}	Maximal release rate for I_{rel}	30 ms^{-1}	K_{up}	$[\text{Ca}^{2+}]_i$ half-saturation constant for I_{up}	0.00092 mM
$[\text{Ca}^{2+}]_{\text{up,max}}$	Maximal Ca^{2+} concentration in uptake compartment	15 mM	$[\text{Cmdn}]_{\text{max}}$	Total calmodulin concentration in myoplasm	0.05 mM/ms
$[\text{Trpn}]_{\text{max}}$	Total troponin concentration in myoplasm	0.07 mM	$[\text{Csqn}]_{\text{max}}$	Total calsequestrin concentration in SR release compartment	10.0 mM
$K_{\text{m,Cmdn}}$	$[\text{Ca}^{2+}]_i$ half-saturation constant for calmodulin	0.00238 mM	$K_{\text{m,Trpn}}$	$[\text{Ca}^{2+}]_i$ half-saturation constant for troponin	0.0005 mM
$K_{\text{m,Csqn}}$	$[\text{Ca}^{2+}]_{\text{rel}}$ half-saturation constant for calsequestrin	0.8 mM			
Tissue-level parameters					
D_{model}	Diffusion/model coefficient	$0.154 \text{ mm}^2/\text{ms}$	I_{bk}	Background current parameter	0.0 pA/pF

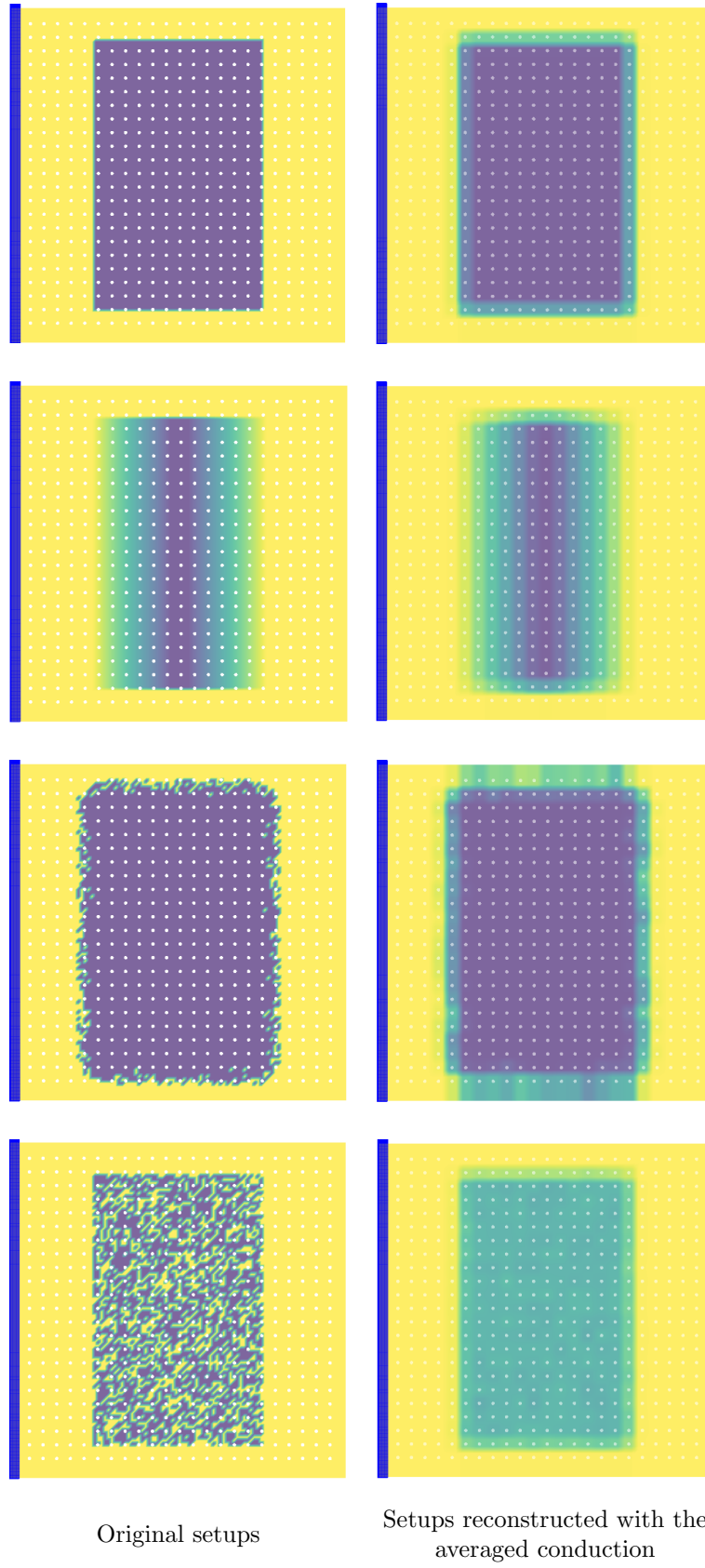


Figure A.1: The effect of the averaged conduction (right) on the original setups (left). Yellow regions indicate healthy conduction, while darker purple regions indicate slower conduction velocity.

A.2 Supplementary materials for results

A.2.1 Correlation between different features

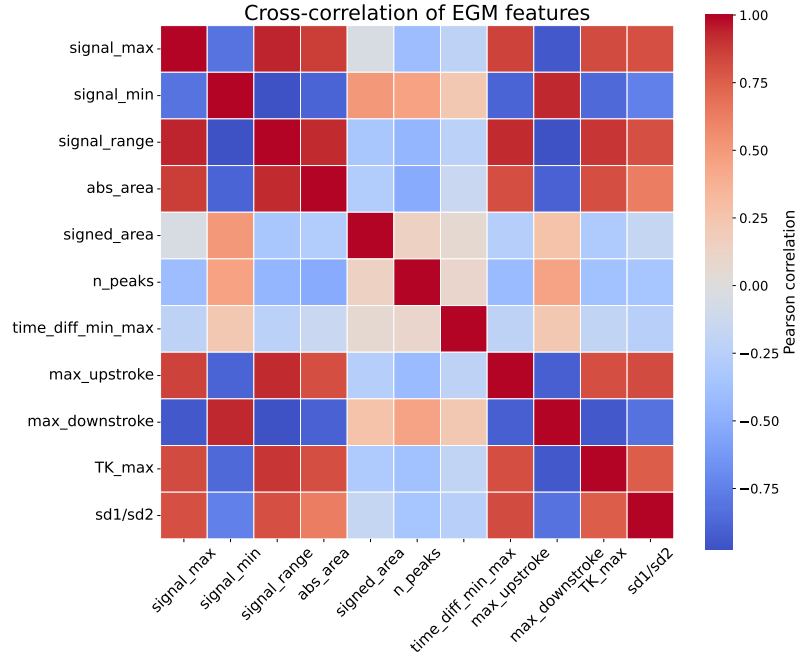


Figure A.2: Feature correlation matrix for the substrate simulations. Rows and columns represent the same 11 features, with each square showing their pairwise correlation. Red indicates correlation and blue anticorrelation, with stronger color saturation corresponding to larger absolute correlation values.

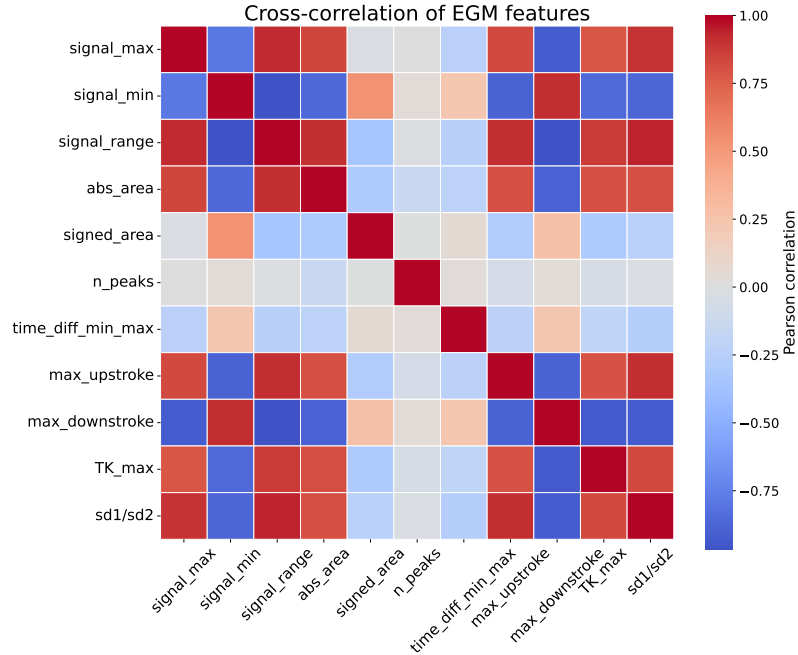


Figure A.3: Feature correlation matrix for the conduction gradient simulations. Rows and columns represent the same 11 features, with each square showing their pairwise correlation. Red indicates correlation and blue anticorrelation, with stronger color saturation corresponding to larger absolute correlation values.

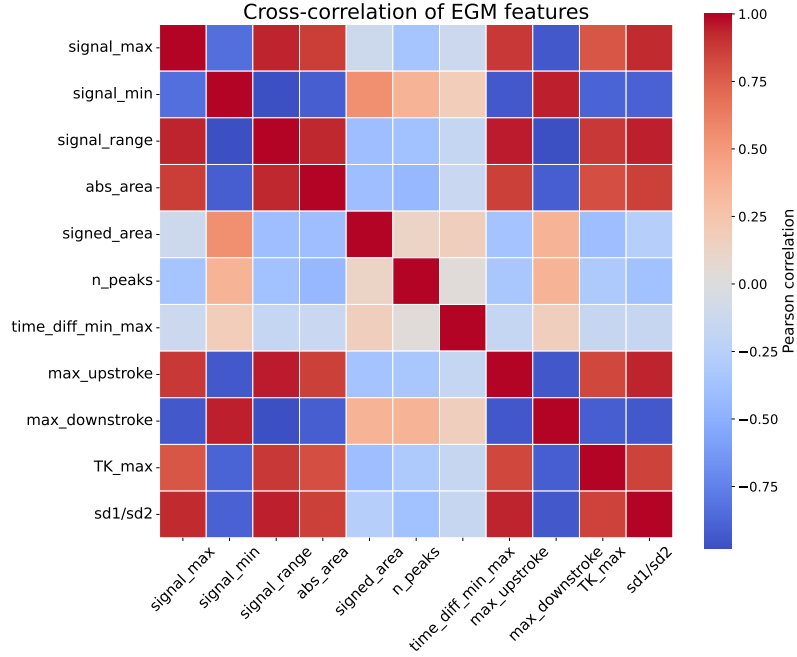


Figure A.4: Feature correlation matrix for the variable stimulus simulations. Rows and columns represent the same 11 features, with each square showing their pairwise correlation. Red indicates correlation and blue anticorrelation, with stronger color saturation corresponding to larger absolute correlation values.

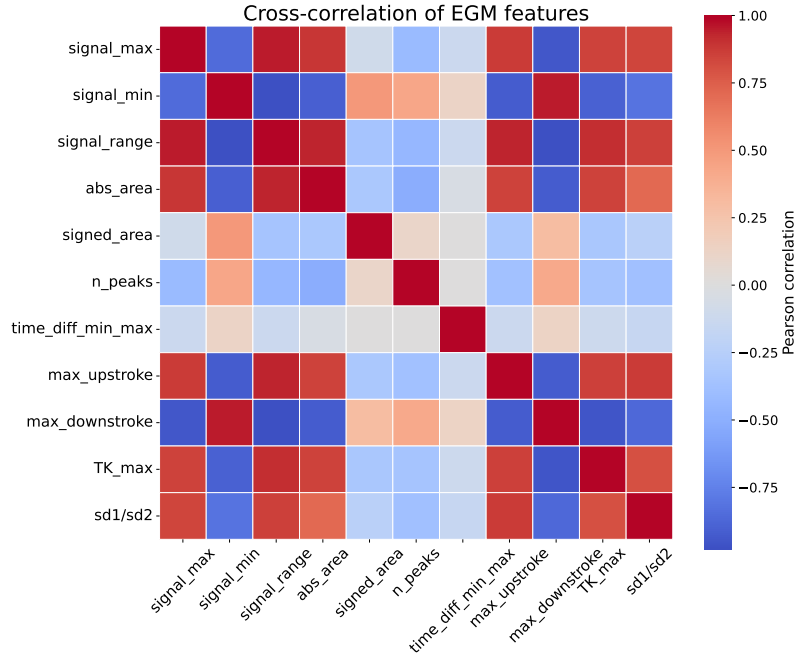


Figure A.5: Feature correlation matrix for the probabilistic boundary simulations. Rows and columns represent the same 11 features, with each square showing their pairwise correlation. Red indicates correlation and blue anticorrelation, with stronger color saturation corresponding to larger absolute correlation values.

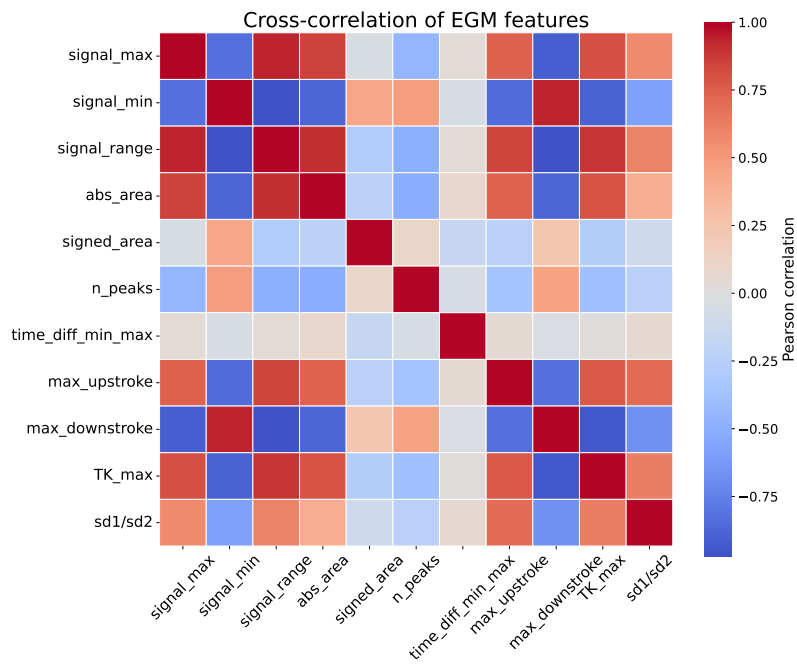


Figure A.6: Feature correlation matrix for the diffuse infarct simulations. Rows and columns represent the same 11 features, with each square showing their pairwise correlation. Red indicates correlation and blue anticorrelation, with stronger color saturation corresponding to larger absolute correlation values.

A.2.2 PCA loadings and explained variance

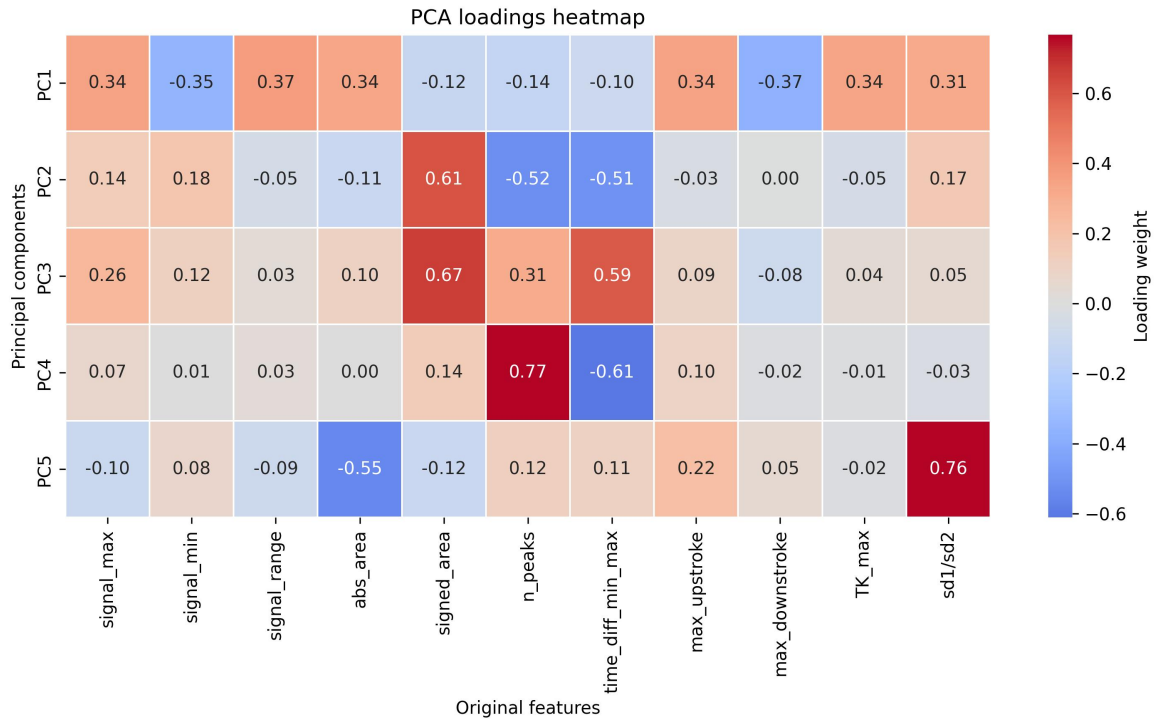


Figure A.7: PCA loadings for the substrate simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.

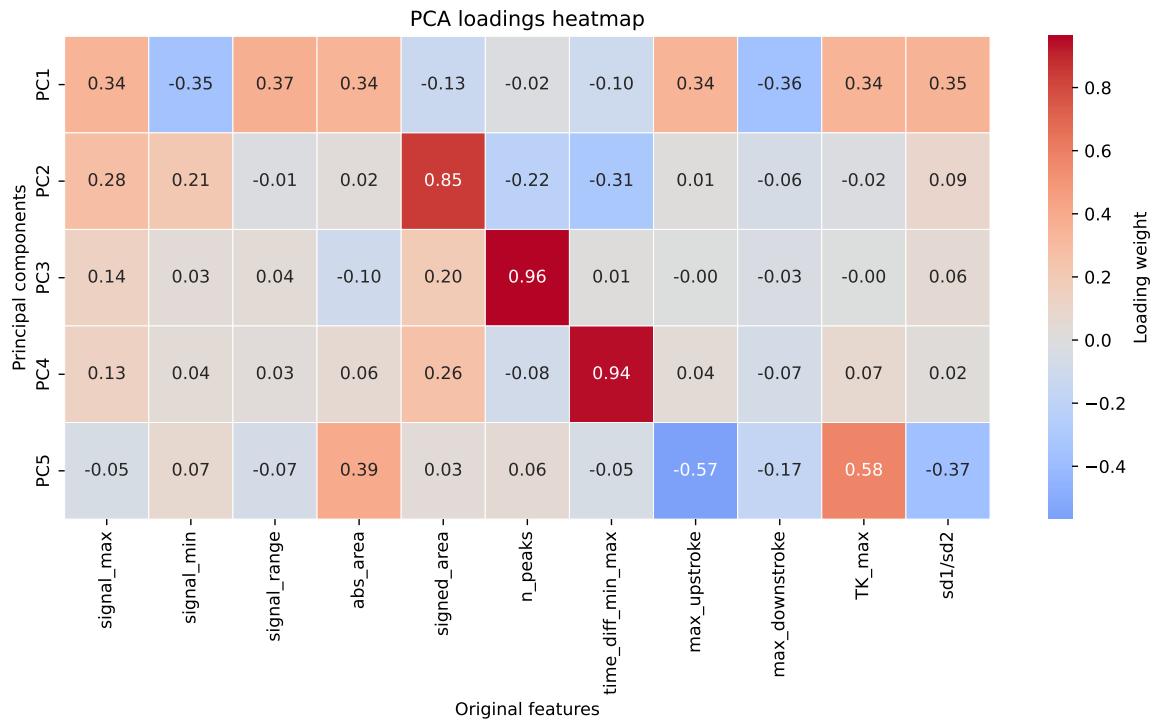


Figure A.8: PCA loadings for the gradient simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.

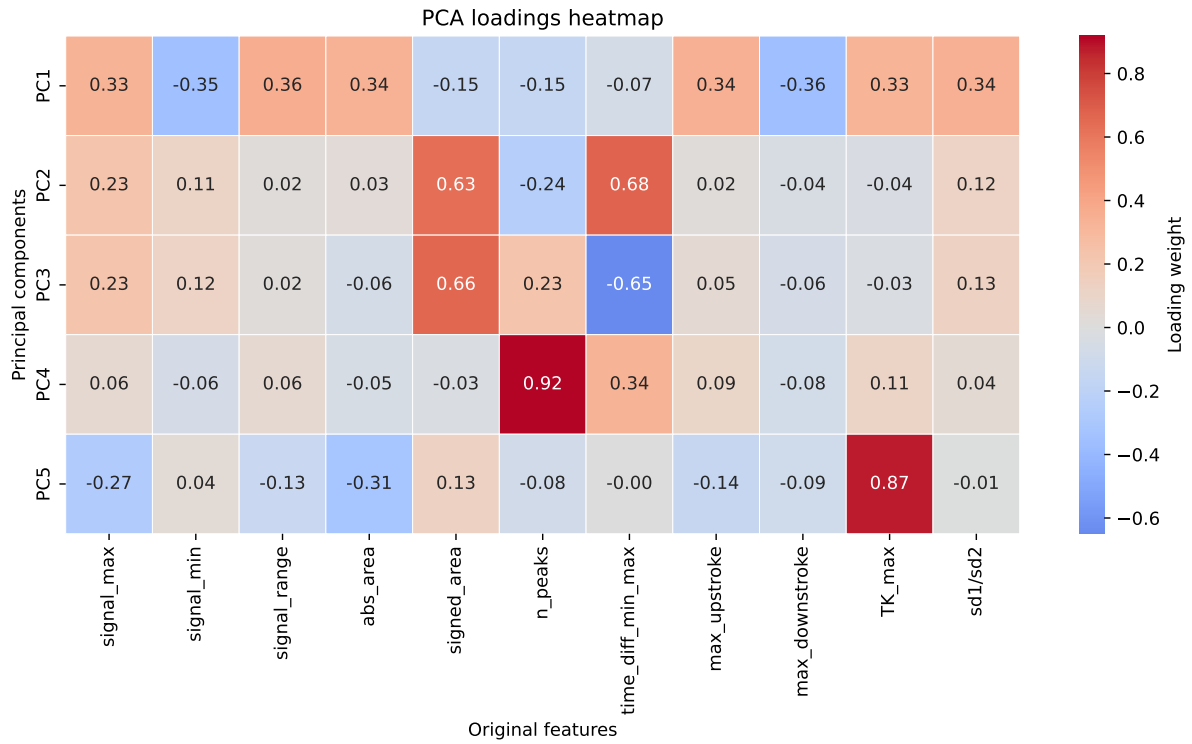


Figure A.9: PCA loadings for the variable stimulus simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.

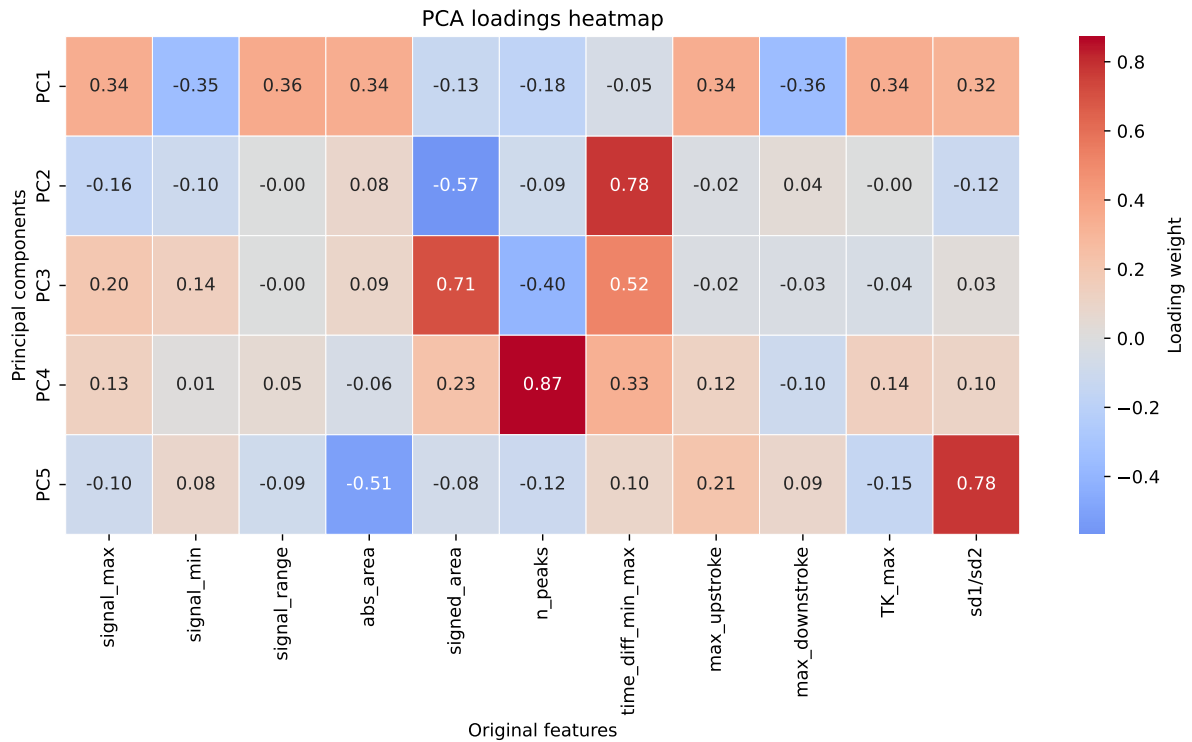


Figure A.10: PCA loadings for the probabilistic boundary simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.

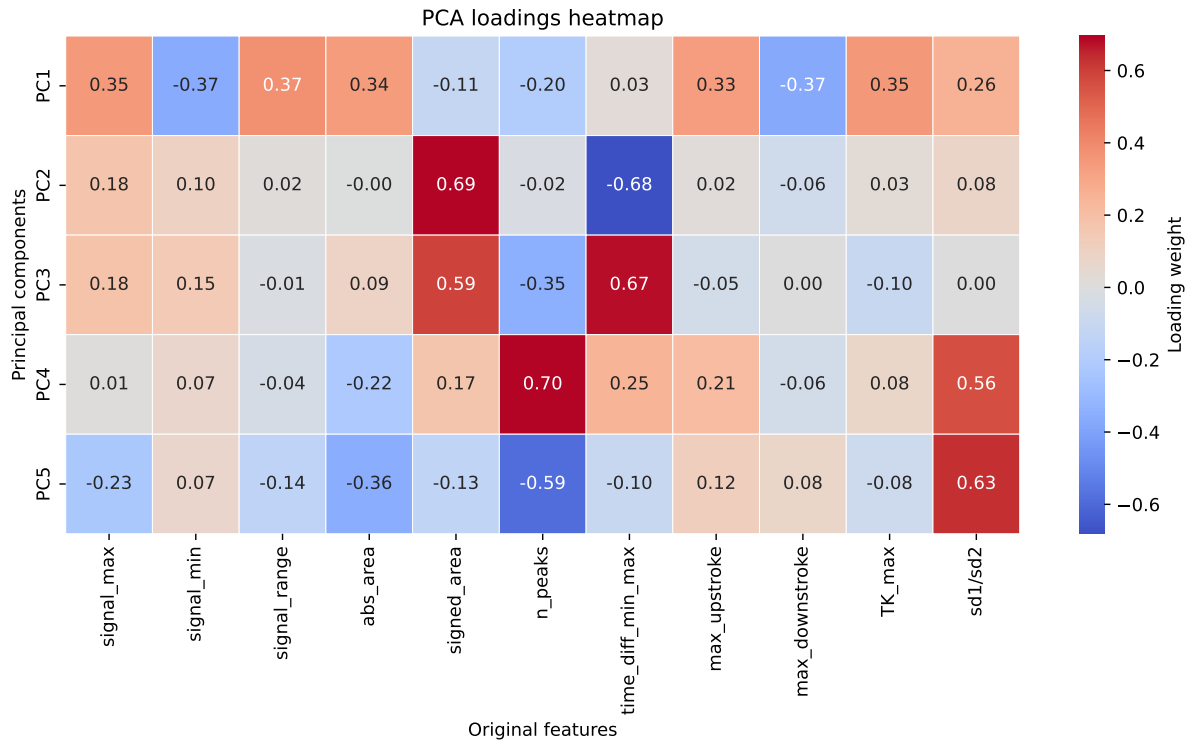


Figure A.11: PCA loadings for the diffuse infarct simulations. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.

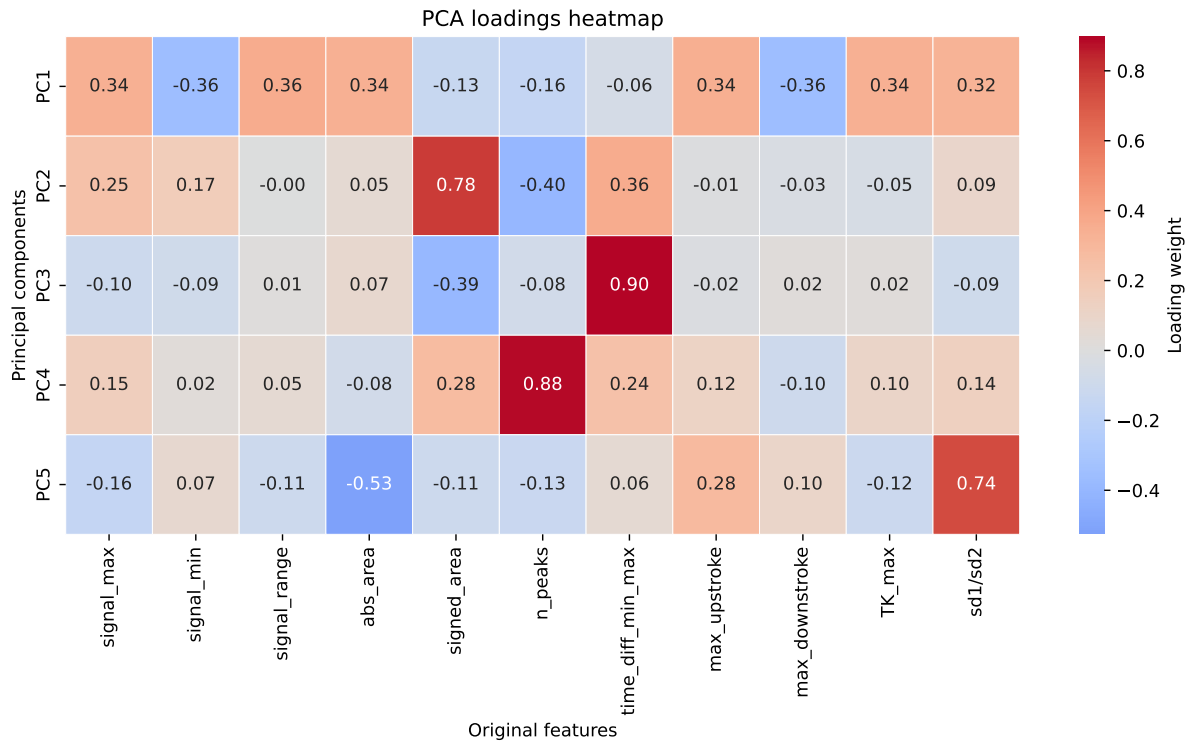
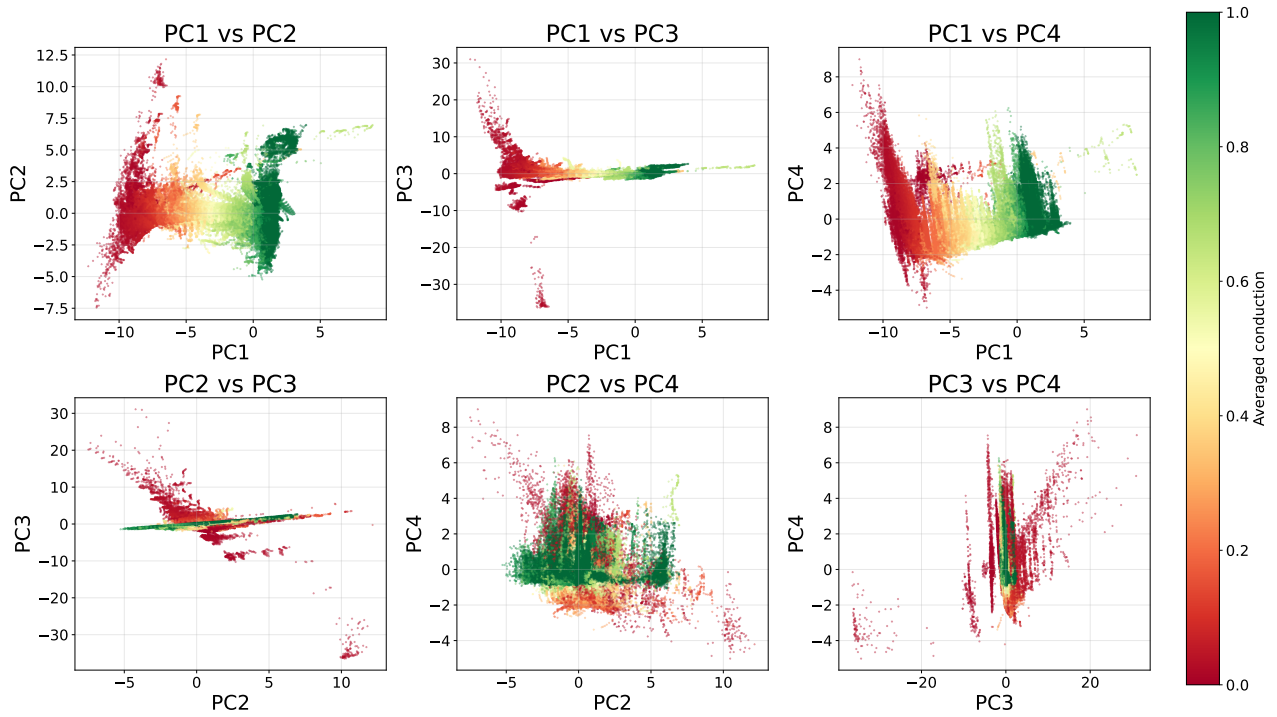


Figure A.12: PCA loadings for all simulations in one dataset. Red colors show positive loadings and blue colors show negative loadings. A more saturated color reflects a higher loading.

A.2.3 Correlations between the principal components

Correlation between PCA components for substrate in function of averaged conduction



Correlation between PCA components for substrate in function of distance to affected tissue

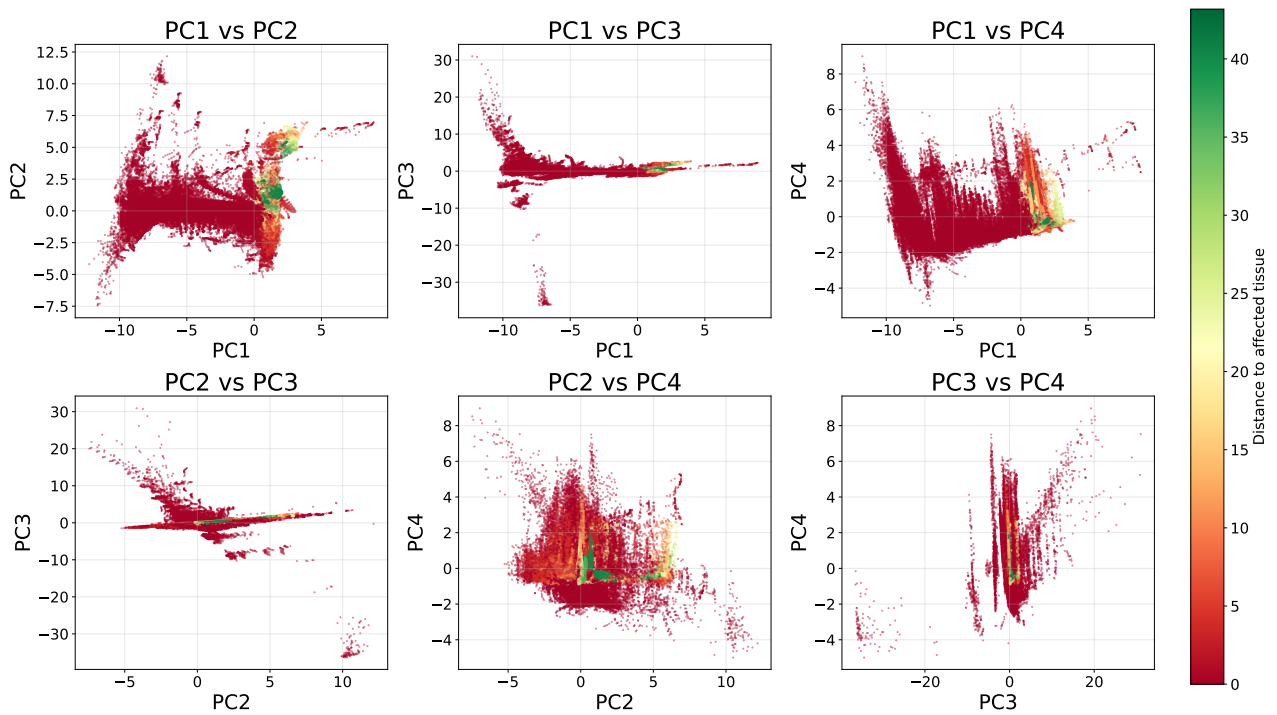
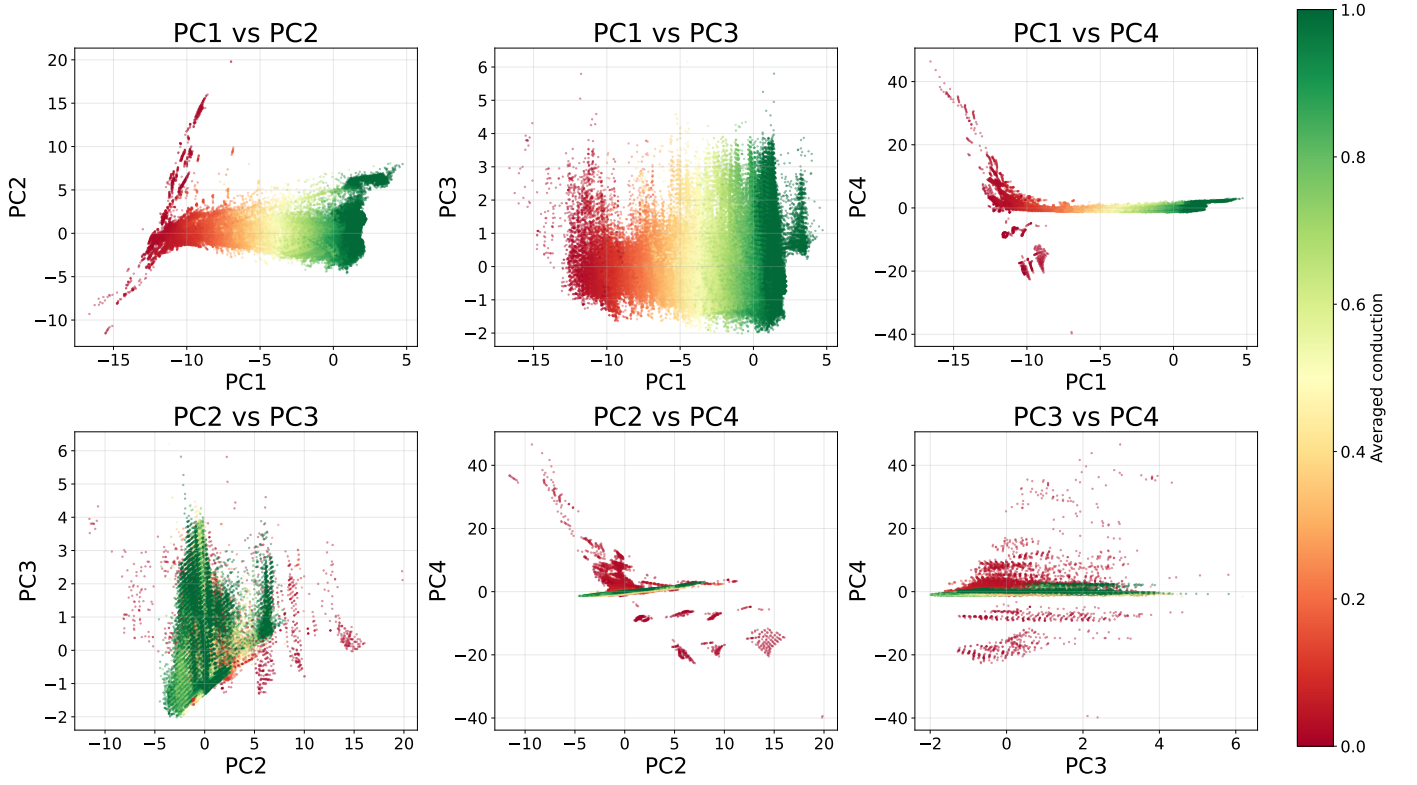


Figure A.13: The correlation between all principal components as a function of the averaged conduction (top panel) and distance to affected tissue (bottom panel) for the substrate simulations.

Correlation between PCA components for gradient in function of averaged conduction



Correlation between PCA components for gradient in function of distance to affected tissue

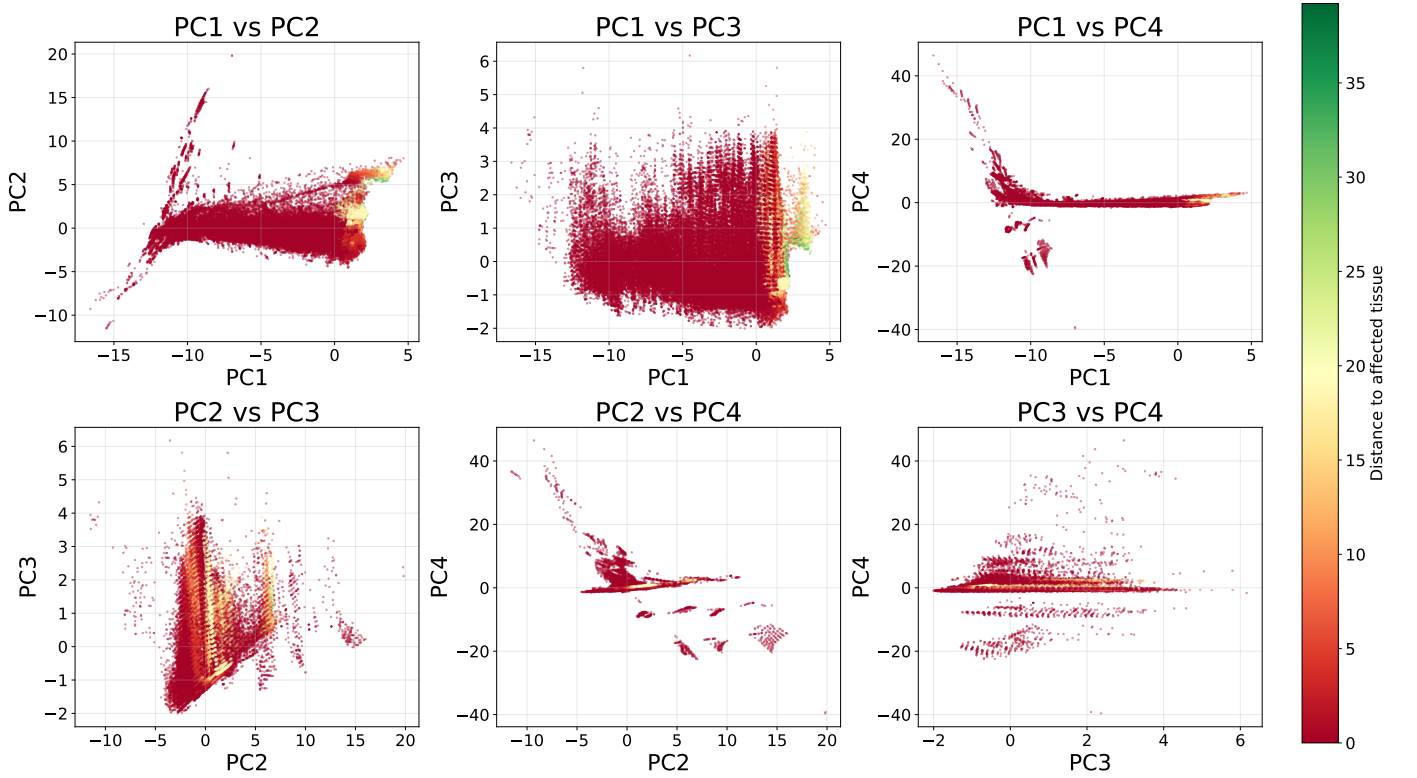
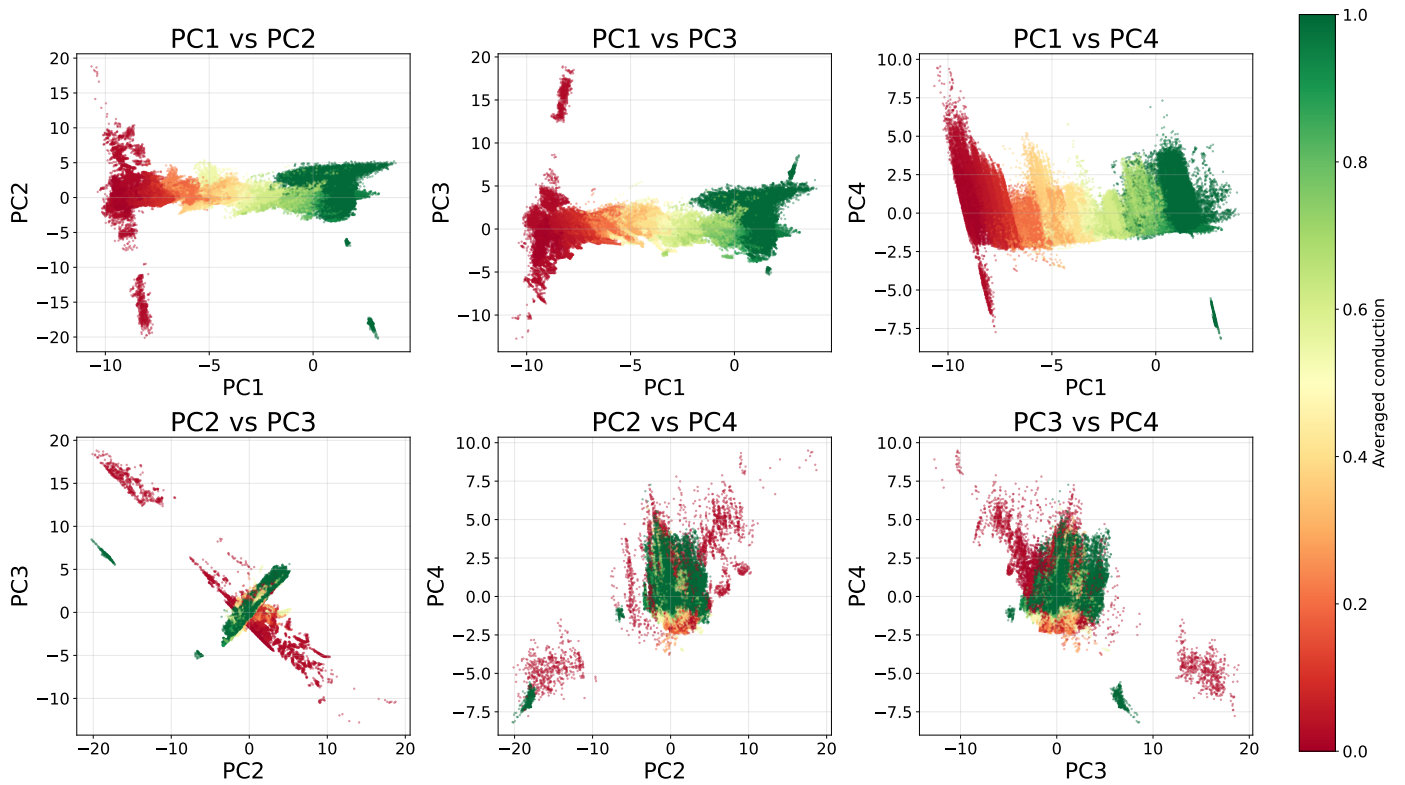


Figure A.14: The correlation between all principal components as a function of the averaged conduction (top panel) and distance to affected tissue (bottom panel) for the conduction gradient simulations.

Correlation between PCA components for variable_stim in function of averaged conduction



Correlation between PCA components for variable_stim in function of distance to affected tissue

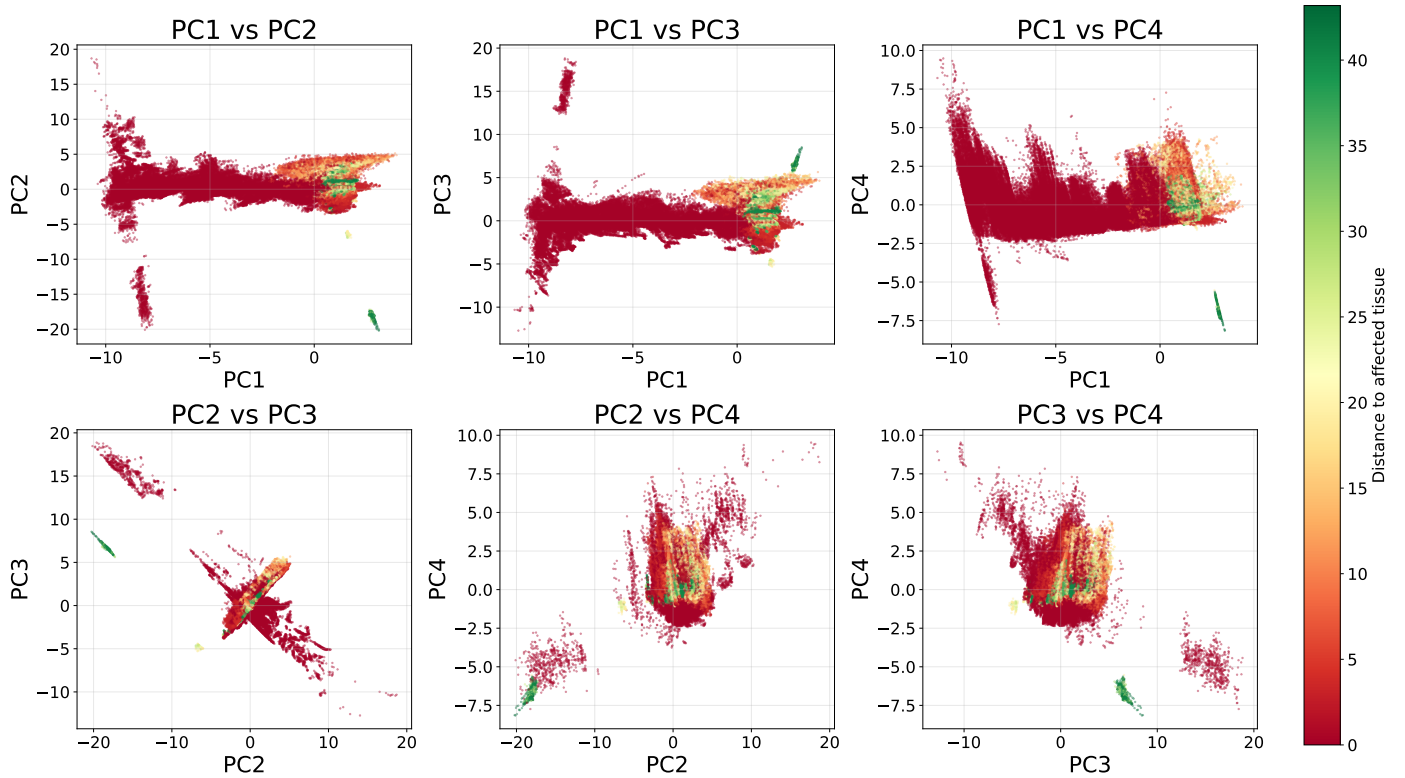
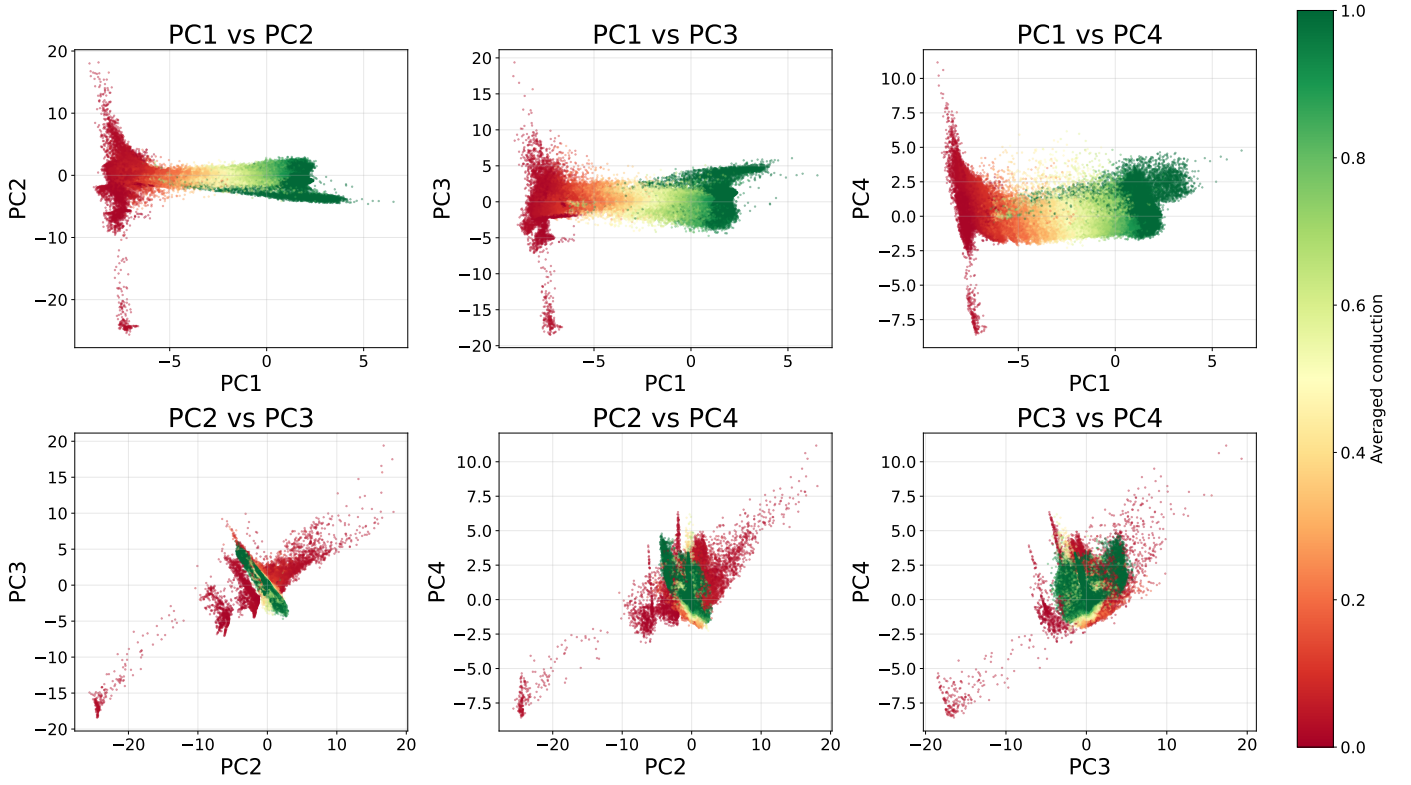


Figure A.15: The correlation between all principal components as a function of the averaged conduction (top panel) and distance to affected tissue (bottom panel) for the variable stimulus simulations.

Correlation between PCA components for probabilistic_boundary in function of averaged conduction



Correlation between PCA components for probabilistic_boundary in function of distance to affected tissue

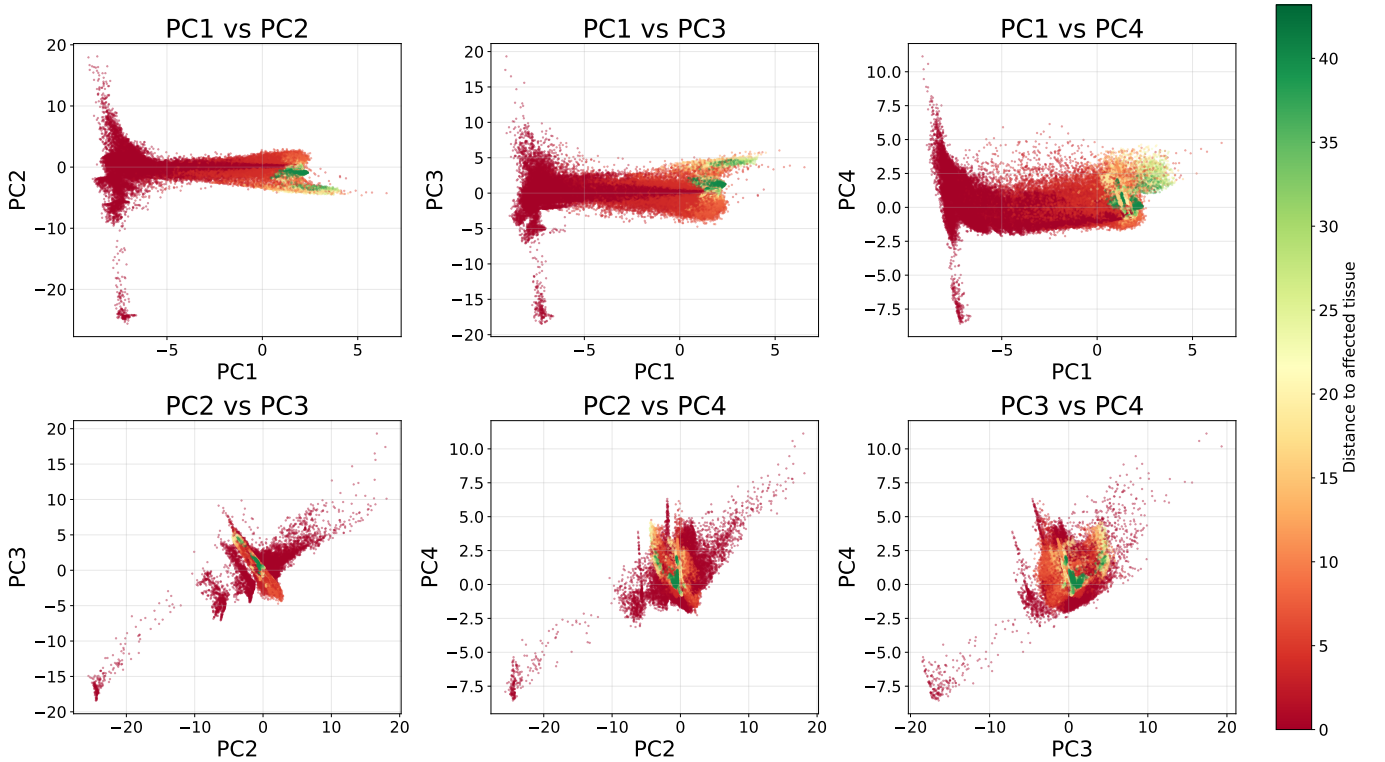
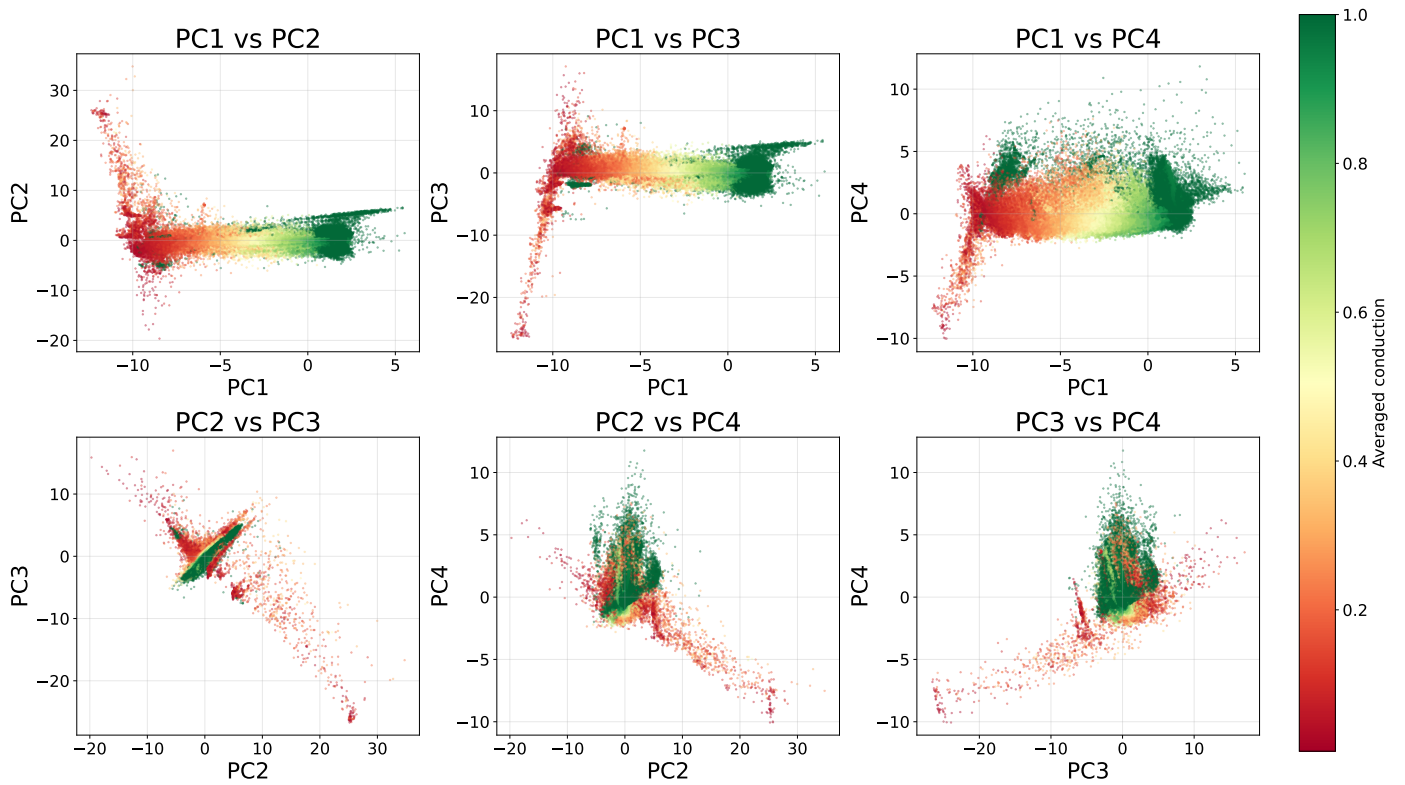


Figure A.16: The correlation between all principal components as a function of the averaged conduction (top panel) and distance to affected tissue (bottom panel) for the probabilistic boundary simulations.

Correlation between PCA components for diffuse_infarct in function of averaged conduction



Correlation between PCA components for diffuse_infarct in function of distance to affected tissue

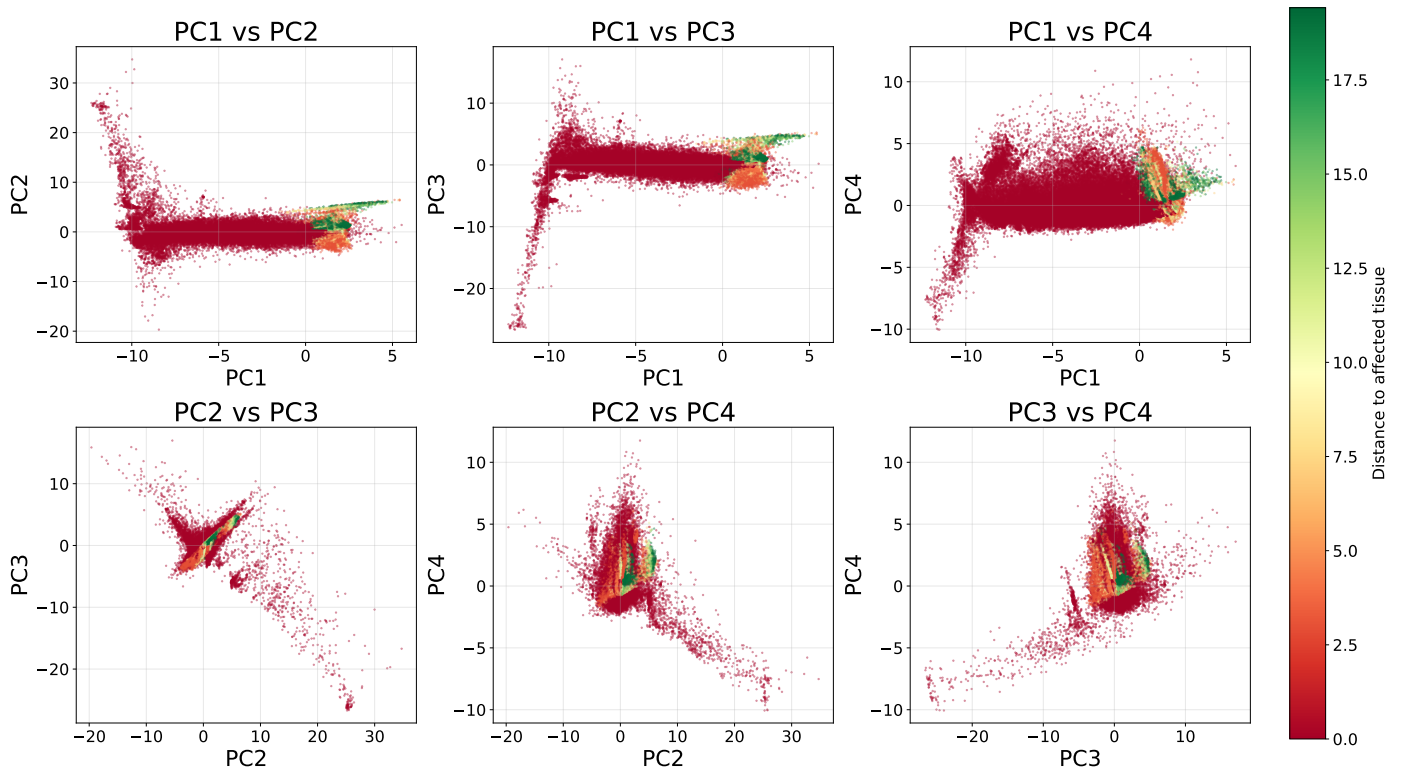


Figure A.17: The correlation between all principal components as a function of the averaged conduction (top panel) and distance to affected tissue (bottom panel) for the diffuse infarct simulations.